Order Number 9429663

The history of the *uhu* transposable element in the Hawaiian *Drosophila*

Wisotzkey, Robert Grier, Ph.D.

University of Hawaii, 1994

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

THE HISTORY OF THE *UHU* TRANSPOSABLE ELEMENT IN THE 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)

MAY 1994

By

Robert Grier Wisotzkey

Dissertation Committee:

John A. Hunt, Chairperson
Kenneth Y. Kaneshiro
Hampton L. Carson
David S. Haymer
Terrence W. Lyttle

ACKNOWLEDGMENTS

I would like to thank Dr. John A. Hunt for his constant support and encouragement, his discussions and his friendship throughout this project. I thank Dr. Hampton L. Carson for his patience, his insight and his love of science, Dr. David S. Haymer for his support and integrity, Dr. Kenneth Y. Kaneshiro for his support and knowledge and Dr. Terrence W. Lyttle for his support and friendship. I am grateful to Dr. L. Ingrid Felger for her help, her knowledge and the training she gave me. Linden T. Doescher, Bill Perriera and Kelvin Kanegawa were always willing to provide technical help. I also thank them for their friendship. I thank Shane Gilmore for making a genomic library for me.

I am grateful to my lab mates, Natasha Y. Patrick for her camaraderie and for her sense of humor, and Atsuko Z. Haruyama for her encouragement. They helped keep everything sane. There are several people who were there for me when I needed them. Of these I would really like to thank Bailey D. Kessing, Jayma L. Martin, Geoffrey E. Dalander, Eric L. Olson, David E. Whiteside, Carol A. Reeb, Christopher G. Prince, C. Christopher Dudley and Kristen L. Anderson.

Finally, I thank my parents, Daniel G. and Dolores M. Wisotzkey for their love and their belief in me.

ABSTRACT

The *uhu* transposable element belongs to the class of elements that have short inverted repeats. It was originally isolated from *Drosophila heteroneura,* a Hawaiian picture-winged *Drosophila* endemic to the Island of Hawaii. Biogeographic and DNA sequence divergence data suggest an ancient origin for the *uhu* element in the Hawaiian *Drosophila.* Biogeographic data suggests that *uhu* arose more than 7 million years ago. Sequence divergence data and phylogenetic analysis suggests that *uhu* was present in a common ancestor of the species. The maximum distance between two isolates suggests that *uhu* has been in the Hawaiian *Drosophila* for 20 million years. Using *in situ* hybridization to polytene chromosomes, the copy number of *uhu* in the planitibia subgroup and the adiastola subgroups of the Hawaiian *Drosophila* is found to be higher in the species endemic to the younger islands than in the species endemic to the older islands. This trend is also seen for the *loa* transposable element in the planitibia subgroup. No complete *loa* elements are found in *D. picticornis* from the island of Kauai, while there are 10 to 20 potentially complete copies of *loa* in the other species. For the *uhu* element, the percentage of sites that are variable for the presence or absence of *uhu* is high in the species on the younger islands, while nearly all the sites in *D. picticornis* are fixed. This would indicate that *uhu* has more recently been active in the species on the younger islands. Since all of the species are single island endemics, and believed to have evolved on the island, the increase in copy number and evidence for transpositional activity is consistent with the idea that there has been increase in the activity of transposable element associated with a speciation event.

TABLE OF CONTENTS

TABLE OF CONTENTS

(CONTINUED)

TABLE OF CONTENTS

(CONTINUED)

LIST OF TABLES

LIST OF FIGURES

LIST OF FIGURES

(CONTINUED)

LIST OF ABBREVIATIONS

#	number
Ala	alanine
Amp	ampicillian
Arg	arginine
Asn	asparagine
Asp	aspartic acid
BCIP	5-bromo-4-chloro-3-indolyl phosphate
bp	base pair
BSA	bovine serium albumine
C-terminus	Carboxyl terminus
cm	centimeter
cpm	counts per minute
Cys	cysteine
dATP	deoxy-adenosine triphosphate
dCTP	deoxy-cytosine triphosphate
dGTP	deoxy-guanine triphosphate

LIST OF ABBREVIATIONS

(CONTINUED)

Dig-dUTP	Digoxygenine labeled deoxy-uriacile triphosphate
dITP	deoxy-indocine triphosphate
DNA	deoxyribose nucleic acid
dTTP	deoxy-thimine triphosphate
dUTP	deoxy-uriacile triphosphate
EtOH	ethanol
g	gram
Gln	Glutamine
Glu	Glutamic acid
h.r.	highly repetitive DNA
His	histidine
hr	hour
IAA	isoamyl alcohol
Ile	isoleucine
IPTG	isopropylthio-β-D-galcatoside

LIST OF ABBREVIATIONS

(CONTINUED)

KAc	Potassium acetate
kb	kilobase pair
Leu	leucine
LINE(s)	Long Intersperrsed Sequences
Lys	lysine
m.r.	middle repetitive DNA
ml	milliliter
Mya	Million years ago
NBT	nitro blue tetrazolium chloride
ng	nanogram
o/n	overnight
OD_{600}	optical density at 600 nanometers
ORF	Open Reading Frame
PCR	Polymerase Chain Reaction
ppt	precipitation

LIST OF ABBREVIATIONS

(CONTINUED)

Pro	proline
r.t.	room temperture
$Rh(DIP)_3^{3+}$	Tris (4,7- diphenyl- 1,10- phenanatroline) rodium (III)
RNA	ribose nucleic acid
rpm	rotations per minute
s.c.	single copy DNA
sdd	sterile double distilled
Ser	serine
SN	supernatant
SV40	Simain Virus 40
Thr	threonine
Tyr	tyrosine
vol.	volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactoside
μg	microgram
μl	microliter

INTRODUCTION

There has been considerable discussion about the role of transposable elements in the evolutionary process. A generally accepted theory is that transposable elements are invisible to selection or parasites on the genome of the host (selfish DNA) (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Other theories suggest that transposable elements have some genetic function: genome organization (Manuelidies, 1982; Bennett, 1982) increasing variability by causing mutation (Chao *et al.*, 1983; MacKay, 1985) or control of gene expression (Britten and Davidson, 1969, 1971; Schwarz-Sommer and Saedler, 1987). Other analysis suggests that transposable elements originated as selfish DNA but have later assumed a cellular or genetic function (von Sternberg *et al.*, 1993). Because of the high copy number of elements and their distribution in the genome even if only a small percentage develop a cellular function, the result could be a large evolutionary effect.

Another aspect of the evolution of transposable elements is their behavior during speciation. The most extensive studies have been on the *L1* element in rodents (Vanlerberghe *et al.*, 1993) and the *Alu* element in primates (Perna *et al.*, 1992). The *L1* element shows different patterns in different species, in *Mus musculus* the elements fall into three clades defined by different levels of divergence. This suggests that there has periods of expansion of the element. Conversely, in two vole species the elements do not fall into clusters, suggesting a more constant level of activity of the element. The *Alu* element shows a pattern in primates very similar to *L1* in

M. musculus, with copies of the element falling into clades (Perna *et al.*, 1992) The location of copies of *Alu* is also conserved between species.

There are problems extending these studies to other systems. It is not possible to use these systems to tests hypothesis derived from the data. The Hawaiian *Drosophila* offer several advantages in looking at the behavior of transposable elements during evolutionary changes. There are two genera of Hawaiian *Drosophila*; *Drosophila* and *Scaptomyza*. The *Scaptomyza*, though a separate genus are believed to have evolved on the Hawaiian Islands from a Drosophila ancestor (Throckmorton, 1975). The genus *Drosophila* is divided into several groups based on morphology; the fungus feeders, antopocerus, modified mouthparts, picture-winged. The picture-winged, the most extensively study group, are divided into four major subgroups (Carson, 1983). These groups and subgroups are reflective of the evolutionary relationship of the species (Thomas and Hunt, 1991; DeSalle and Giddings, 1986). Thus it is possible to pair up, between different subgroups, species that have a similar biogeographic distribution and evolutionary relationship to other members within their group. This is similar to running a replicate experiment, and the repeatability of an observation can be tested. An example of two subgroups that provide this kind of comparison are the planitibia and adiastola subgroups (Figure 1). They are believed to have diverged from each other 5 million years ago (Thomas and Hunt, 1991). There is a representative species on each of the main islands, expect O'ahu. Based on chromosomal inversion data (Carson, 1987) the relationship of the species to other species with in the subgroups, is similar between the subgroups (Figure 2). Because

of this, the distribution of the *uhu* transposable element was studied in these two groups.

Clearly transposable elements are potentially disruptive to the genome they occupy. The movement of a *P* element (transposition) was found to be associated with a 1% decrease in fitness in *Drosophila melanogaster* (Eanes et al., 1988). The loss of fitness was attributed to deletions and rearrangements resulting from the mobilization of *P* elements and not necessarily to mutations caused by insertion into a new site. If transposition of an element causes a reduction in fitness, mechanisms in the host genome that suppress transposition would be favored. Indeed, transposition is considered to be a rare event and movement of elements have been found to occur in bursts (Junakovic and Angelucci, 1986). Heat-shock has been shown to induce transposition of several elements (*mdg-1, 297, 412, B104* and *copia*) in *D. melanogaster* (Junakovic *et al.*, 1986). In addition, transcription of *copia*, a prerequisite for transposition, was found to be induced by both heat-shock and chemical shock in *D. melanogaster* (Strand and McDonald, 1985). Temperature was also found to be associated with the degree of the *P* element induced hybrid dysgenesis in *D. melanogaster* (Kidwell et *al.*, 1977). Some exceptional responses to increased temperature in maize may also be associated with movement of transposable elements (Cullis, 1988). The detectable amount of transposition of the *Ty* element in *Saccharomyces cerevisiae* grown at the sub-optimal temperatures of 15°C and 20°C was 100 fold that of the *Ty* element in yeast grown at the optimal temperature of 30°C (Paquin and Williamson, 1984). Interestingly, no transposition was found in yeast grown at 37°C .

A copia-like element in *D. melanogaster, 1731*, has a sequence showing homology to a steroid activated promoter, allowing for the possibility of hormonal control of transposition (Montchamp-Moreau *et al*, 1993). These studies suggests that stress may relax the repression of, or even induce, transposition of elements. Movement of transposable elements may contribute to a period of genomic instability often seen in founding and peripheral populations, populations which may be considered stressed (McDonald, 1989). The distribution of transposable elements in the Hawaiian *Drosophila*, a group characterized by frequent founder-flush events (Carson *et al.* 1970), may provide evidence for a high amount of recent transposition.

There are two explanations which predict movement of transposable elements in association with the founding of a population or with a population bottleneck. There is a fair amount of evidence for some amount of control over transposition by the host genome. The hybrid dysgenic phenomenon observed for several transposable elements is suggestive of a cytoplasmic control, and evidence of cytoplasmic control exists in the case of P elements (Misra and Rio, 1990). The stress associated with inbreeding during a population bottleneck may inhibit this repression of transposition and allow for an increase in transposition. Inbreeding could increase or decease the copy number of an element in an individual. This could disturb any balance between the host genome and the copy number of the element. The reduction in selection proposed to occur during a population flush following a founding event or bottleneck, including a reduction in the selection against the transposition of an element, would have the similar affect of allowing for a net increase in the rate of transposition, simply because individuals with a

high level of transposition would survive when they normally wouldn't. Alternatively, because environmental stress has been shown to induce transposition of several elements, the novel environmental conditions that can be associated with a bottleneck or with the founding of a population may induce transposition of elements. Because transposition is generally considered to be replicative, resulting in a new copy of the element at a novel location in the genome without removing the parent copy and because excision of an element is a rare event (Charlesworth and Langley, 1989), an increase in transposition associated with a founding event speciation should result in a higher copy number of the element in the new population.

The Hawaiian *Drosophila* offer several advantages for the study of evolutionary processes. The radiation is impressive; at least 25% of all described species of *Drosophila* are endemic to the Hawaiian islands (Hardy and Kaneshiro, 1981). Most species are restricted to a single island (Carson *et al.*, 1970). Extensive phylogenies based on the karyotype are available (Carson, 1983). An estimate of the divergence time between species can be made based on the formation of the Hawaiian Islands. The Hawaiian Islands are volcanic islands in the middle of the Pacific Ocean. They are believed to have formed as the Pacific plate moves over a mantle plume in a northeast direction (Wilson, 1963). Based on potassium-argon and magnetic declination data, Kauai, the northern most high island, is about 5 million years old. Hawaii (the Big Island), the southern most island, is about 500,000 years old. The other main Hawaiian Islands are O'ahu (3.5 million years old), Molokai (1.8 million), Maui (1.3 million) and Lanai. During the Pleistocene, the islands of Molokai, Maui and Lanai were joined because of decrease in the sea level.

Because of this these islands are considered a single island biogeographically and are referred to as the Maui complex or Maui Nui.

Species arising from a population colonizing a new island as it becomes habitable has been argued as the major mode of speciation in the group (Carson *et al.*, 1970). The relative divergence time of the species is inferred from the age of the islands. Data from allozymes (Carson, 1976), DNA-DNA hybridization (Hunt *et al.*, 1981) and sequence divergence of the Adh region (Rowan and Hunt, 1991) and mitochondrial DNA (DeSalle and Giddings, 1986) are in agreement with karyotype-based phylogenies and the relative divergence times between the species from these data are consistent with divergence inferred from the age of the islands. It must be noted that the native habitat of the island of O'ahu has suffered extensive destruction because of human habitation. Representative species from this island are numerous but rare.

A repetitive DNA sequence, *uhu*, was found in *D. heteroneura* on the 3' end of the *Adh* locus (Hunt *et al.*, 1989). This element is found in several other members of the planitibia subgroup. The subgroup is several closely related species that are believed to have arisen during founder-flush events on each of the islands. Several members are homosequential in the banding pattern of their polytene chromosomes with respect to each other. Other members differ from the others by several fixed chromosomal inversions. There is some evidence that the element is transposable (Hunt, *et al.*, 1984, Brezinsky *et al.*, 1990, Brezinsky *et al.*, 1993). *D. picticornis*, which is endemic to the montane rainforests of Kauai, has about 10 copies of the sequence.

D. silvestris, endemic to the Big Island, contains about 150 copies (Hunt *et al.*, 1984). Felger (1988) has isolated several clones of dispersed middle repetitive DNA from *D. silvestris* that have a higher copy number in *D. silvestris* than in *D. picticornis*. This difference in copy number of the elements between the species may reflect a difference in the amount of recent replicative transposition of the element within the species.

Other lines of evidence also suggest a higher level of recent transposable element activity in *D. silvestris*. Cytologically, *D. silvestris* and *D. picticornis* have thirteen fixed inversion differences with respect to each other. *D. silvestris*, in addition, is polymorphic for twelve (12) inversions unique to the species (Carson, 1983). Since transposable elements have been found in association with inversion breakpoints (Perlman, 1983; Lim 1988; Lyttle and Haymer, 1993), the new inversions in *D. silvestris* may be further evidence of an increased amount of transposition occurring in that species. A survey of more species is needed to see if the trend, an increase in a middle-repetitive element's copy number in species endemic to the newer islands, is repeated in other evolutionary lineages and with other elements.

If this increase in copy number of the element is seen in other groups of the Hawaiian *Drosophila*, it would suggest that an increase in the rate of transposition of some elements is associated with the genomic instability during a speciation event, resulting in an increase in the copy number of the element. Drift and selection against the detrimental effects of the element would be expected to decrease copy number over time. The net result would

be a higher copy number in the species subjected to a more recent founding event (i.e. those found on the younger islands).

This study takes two approaches to looking at the relationship of transposable elements to speciation. First, the distribution of the *uhu* transposable element will examined in two groups of closely related species. The variability of the insertion sites will be examined in one group. Secondly, the overall sequence divergence of a specific region of the open reading frame of isolates of the *uhu* element from 10 different species will be compared.

Transposable elements are middle repetitive sequences of DNA, usually having between 10 and 1000+ copies in the genome. In *D. melanogaster*, twelve to eighteen percent of the genome is made up by middle repetitive sequences. Centric heterochromatin makes up about 10% of the middle repetitive DNA and tandemly repeated sequences (rRNA, 5SRNA and histones) account for about 25% of the middle repetitive class (Finnegan, 1985). As much as a quarter of the middle repetitive sequences of *D. melanogaster* are the scrambled, clustered arrays described by Wensink *et al.* (1979). These sequences (300 to 1000 bp long) are found in clusters, several kilobases long, and are dispersed in the euchromatin. Different clusters share some, but not all of the sequences. The arrangement of the sequences is not conserved between clusters. The remaining middle repetitive DNA appears to belong to different sequence families (Spradling and Rubin, 1981). In *D. melanogaster* the majority of these dispersed repetitive sequences range in size from 0.5 to 13 kb, averaging about 5.6 kb (Spradling and Rubin, 1981).

These sequences appear to have no fixed location, which has lead to speculation that they are nomadic (Young and Schwartz, 1980). A portion of these nomadic sequences are able to control their own movement and are referred to as transposable elements.

D. picticornis has an estimated minimum haploid genome size of 8.5 x 10^{10} daltons, with the highly repetitive fraction (Cot<0.05) being about 20% of the genome, the middle repetitive fraction (0.05<Cot<10) being about 19% of the genome and the remaining 61% is single copy (Triantaphyllidis and Richardson, 1980). This is comparable with other Drosophila species and is within the range for the Hawaiian *Drosophila* (Table 1).

Transposable elements can be divided into four classes by their structure. The first of these are elements with Long Terminal Direct Repeats. These include the proviruses in mammals and the *copia-like* elements in *Drosophila*. This whole group is referred to as the retroviral-like elements and are considered the most abundant of the transposable elements (McDonald, 1989). They are retrotransposons, having an RNA intermediate during transposition. The genomic copies of the elements of a family are very similar in structure to one another (Rubin, 1983). Structurally, they are 5 to 8 kb in length, occurring between 20 to 100 times in the genome (Finnegan, 1985). The direct repeat accounts for about 5% of the length of the element. At the edges of these repeats is a short, imperfect inverted repeat. Usually a few bases of genomic target DNA sequences is duplicated at the insertion site (Finnegan and Fawcett, 1986).

The second class are elements with Long Terminal Inverted Repeats. These included the foldback elements (FB) and TE element of *Drosophila*. These elements can be several hundred to several thousand base pairs long. There are between 20 to 30 copies per genome of *D. melanogaster* (Finnegan, 1985). The structure of the element varies, the entire element can consists of the inverted repeats, or a central sequence may be located between the inverted repeats. The repeats generally have a substructure made up of 31 bp tandem repeats. The number of the repeats varying between elements and also between the termini of a single FB element. The are believed to transpose through a DNA intermediate.

A third class is elements without terminal repeats, which includes the I factor and F family in *D. melanogaster*. The F family of inserts show some structural similarity to the Long Interspersed Sequences (LINEs) of mammals. The majority of LINEs are considered to require an RNA intermediate for transposition and are called retroposons (Rogers, 1985). Other retroposons of mammals, like the *Alu* sequences in humans, do not contain terminal repeats (Rogers, 1985).

A final category is elements with Short Inverted Terminal Repeats. This includes the *P* element in *Drosophila melanogaster*, controlling elements in maize, *Tam* in *Antirrhinum majus*, and *Tc1* in *Caenorhabditis elegans*. These are typically about 3kb in length (Finnegan, 1985). The termini of *P* elements are perfect inverted repeats of 31 bp. Genomic copies of *P* elements tend to be very conserved. They are 2.9 kb in length and sequence analysis has revealed three long open reading frames. Variations on the basic

structure can usually be explained by one or more deletions (Rubin, 1983). *Uhu*, the element isolated from *D. heteroneura*, is a member of this category. The 1.6 kb sequence has 46-50 bp imperfect inverted terminal repeats. It contains a large open reading frame, potentially encoding a 192 amino acid protein, which shows a degree of amino acid homology with *Tc1* from *Caenorhabditis elegans*, *Tcb1* (*Barney*) from *C. briggsae*, *HB1* from *D. melanogaster* (Brezinsky *et al.*, 1990), *Bari* from *D. melanogaster* (Ciazzi *et al.*, 1993) and *Minos* from *D. hydei* (Franz and Savakis, 1990). Members of this class are believed to have a DNA intermediate for transposition.

The putative open reading frame of the transposase has been expressed in bacterial systems for both the *Tc1* element of *C. elegans* (Schukkink and Plasterk, 1990) and the *Ac* element of maize (Kunze and Starlinger, 1989). Both of these elements are members of the same class of elements as *uhu*, those with short inverted repeats. Both TcA (the putative transposase from the open reading frame in Tc1) and ORFa (the transposase from the first open reading frame of the Ac element) have DNA binding capabilities. The DNA binding site has been localized to the N-terminal region of the protein. In TcA, the first 39 amino acids are necessary for DNA binding. For ORFa, which is a much larger protein, the first 136 amino acids are not necessary for DNA binding, but amino acids between 180 and 200 are necessary (Kunze and Starlnger, 1989). ORFa also shows amino acid homology in two regions towards the C-terminal end with the putative transposases of the *hobo* element of *Drosophila* and the *Tam* element of snapdragons. In both the TcA and ORFa, the DNA binding regions do not have an obvious secondary structure, but do contain a high proportion of basic amino acids. The binding

activity of the TcA protein appears to be non-specific. The ORFa protein recognizes the sequence AAACGG. This sequence occurs several times in both the 5′ and 3′ ends of the *Ac* element. Preliminary evidence suggests that there is one transposase molecule binding to the sequence, but that there may be a stabilizing or cooperative interaction between the transposase and the DNA, i.e. the binding of one molecule to the target sequence facilitates the binding of other transposase molecules to nearby target sequences.

The phylogenetic distribution of a transposable element may provide clues to it's evolution. Martin *et al.* (1983) looked at the distribution of 5 middle repetitive elements originally isolated from *D. melanogaster*. Sequences homologous to *copia* and *412* (a *copia-like* element) were found to have a broad distribution in the genus. They suggested the elements were present in the genome before the major radiation of *Drosophila* and groups, like the Hawaiian *Drosophila,* that do not have these elements are assumed to have lost them. Other elements (*297*, TIP56 and 77E4), are restricted in distribution to either the melanogaster group (*297*) or to the sibling species of *D. melanogaster* (TIP56, 77E4), suggesting a more recent origin. The *copia-like* element *1731* (Montchamp-Moreau *et al.*, 1993) and the *FB* element (Silber *et al.*, 1989) also have widespread distributions similar to *copia*. These broad distributions may be better explained by horizontal transmission of the element as the sequence divergence would be expected to be too great to be detected by standard techniques if the elements had been inherited vertically (Hunt, pers. comm) The distribution of the *P* element in the genus *Drosophila* is suggestive of a recent invasion of *P* into the genome of *D. melanogaster*. Functional *P* elements have only been isolated from

D. melanogaster. Sequences related to *P* elements are not found in the sibling species of *D. melanogaster.* However, homologous sequences have been found in the distantly related *D. willistoni* group (Lansman *et al.,* 1987). Portions of the P element sequences from *D. melanogaster* and *D. nebulosa* show only 6% difference (Lansman *et al.,* 1987). Though the isolates from *D. nebulosa* do not appear to be functional, the degree of difference suggests a much more recent divergence between *D. nebulosa* and *D. melanogaster* than other lines of evidence, supporting horizontal transmission of the element.

The evolutionary relationship at the DNA sequence level of an element in several different species has been studied extensively for two elements; *mariner* in the insects (Capy *et al.,* 1993) and *L1* in mice and their allies (Vanlerberghe *et al.,* 1993). The *mariner* element was first isolated from *D. mauritiana* of the melanogaster group. It is found in several, but not all, members of the group. This group is divided in to the *melanogaster* cluster and *yakuba* cluster of species. Sequence analysis of *mariner* suggests that it is active in all the species in which it is found. Isolates of *mariner* from species in the *yakuba* cluster are clearly different from isolates from species in the *melanogaster* cluster (Capy *et al.,,* 1993). Within the clusters, the isolates do not fall neatly within the species. Isolates from *D. mauritiana,* for example, will show greater sequence homology to an isolate from *D. simulans* than to another isolate from *D. mauritiana.* The relative divergence of the isolates in the melanogaster group is similar to the divergence of single copy genes between the species and supports the idea that *mariner* was present in an ancestor of the group and has been transmitted vertically to the extant species. *Mariner* like sequences have been found in

the moth *Hyalophora cecropia*, the nematode *Caenorhabditis elegans* and several members of the genus *Zaprionus* (Capy *et al.*, 1993). Using degenerative PCR primers, Robertson (1993) has found sequences homologous to *mariner* in ten non-drosophilid species from six different orders of insects. The phylogenic relations between isolates from all these species show differing patterns of horizontal and vertical transfer, suggesting that *mariner* is able to maintain itself in several different genetic backgrounds.

The *L1* element is a LINE element originally isolated from *Mus musculus*. It is found in several other murine species. Sequence analysis of isolates of *L1* from *Mus musculus* places the isolates into three major clades. The A clade elements have <5% sequence divergence, the F clade showing ~10% divergence and the V clade showing 20-25% divergence (Vanlerberghe *et al.*, 1993). This suggests that there have been major bursts of expansion of the element within the species, followed by extended periods of inactivity. This is similar to what is seen for the *Alu* element in primates (Perna *et al.*, 1992). Isolates from other species of *Mus* show about 5% sequence divergence. Conversely, isolates from two vole species, *Microtus epiroticus* and *Arvicola terrestris*, have no apparent sub-families of the element. Relative sequence divergence of *L1* from these voles species suggest a divergence of 13 million years ago (Vanlerberghe *et al.*, 1993). It is estimated that the vole species diverge ~3.5 Mya. Thus, *L1* appears to have been passed vertically from an ancestral species. The interspecific and intraspecific divergence of the elements are very similar, suggesting that there has been no concerted evolution of the element within either of the species. The difference in the

sequence divergence patterns between the voles and the mice suggests there are several active copies of the *L1* element in the vole species, while only a few, maybe only one, in the *Mus* species (Vanlerberghe *et al.*, 1993).

The copy number of a transposable element in a species is generally consistent between individuals of a species and through time. This suggests that there is an equilibrium between the rate of transposition (μ) -- which should act to increase copy number -- and excision, random drift and selection against the detrimental effects of transposition -- all of which should decrease copy number. Theoretical studies show that given a rate of transposition of 10^{-4}, a small amount of selection, on the order of 10^{-5} against the effects of an individual insertion, is sufficient to establish a selection/ transposition equilibrium of approximately 50 copies of an element per individuals (Charlesworth and Langley, 1989).

It is generally assumed that the number of occupiable sites in the haploid genome (m) is much greater than the number of elements in that genome (n). Assuming that the copy number is at equilibrium, the probability that a site is occupied is x_i, for the entire array of sites (i =1,2,3,...,m), with a mean copy number of elements per individual is

$$n=2\Sigma_i x_i$$

with a variance

$$V=n(1-x)-2ms^2+4D_{ij}$$

s^2 is the variance in the rate occupancy among sites, D_{ij} is the linkage disequilibrium between sites. If s^2 and D_{ij} are small and x, the probability of

occupancy, is much less than 1,

$$V \approx n.$$

Thus the copy number of elements in individuals is expected to follow a Poisson distribution. Data from several elements in *D. melanogaster* shows that this is indeed the case (Charlesworth *et al.*, 1990). Departures from this expected Poisson distribution could suggest either a linkage disequilibrium between sites or a large variation in the probability of the occupancy of a site (f). A positive linkage disequilibrium would increase the variance in copy number between individuals, a negative disequilibrium would decrease the variance.

The probability density of an element frequency x at a site (f) is given by the formula

$$f(x) \approx G(a+b)/G(a)G(b)\ x^{a-1}\ (1-x)^{b-1}$$

$a=4N_e n/(2m-nu)$ approximating the effects of drift and transposition (u), $b=4N_e(s+v)$ approximating the effects of drift, excision (v) and selection against the detrimental effects of transposition (s). If $4N_e$ and m are sufficiently large, a can be ignored and

$$f(x) \approx 1/2\ nx^{-1}(1-x)^{b-1}.$$

Thus, in a large stable population, the occupancy of a site is primarily controlled by the ratio of the copy number of the element to the number of occupiable sites (n/m) and the selection against the affects of individual insertions. A large variation in the occupancy between sites would suggest either unequal selective pressures between sites, or that the historical N_e is small and drift has a stronger affect.

There are several attempts to model the rate of sequence divergence within a class of transposable elements (Ohta, 1985; Charlesworth, 1986; Brookfield, 1986). They are in agreement that rate of sequence divergence will be the result of an equilibrium between mutation -- which increases divergence -- and random genetic drift, transposition and gene conversion -- which increase homogeneity. These models show that the effect of gene conversion is very small and can generally be ignored (Slatkin, 1985; Charlesworth, 1986). Brookfield (1986), in the most general version of the models, shows that the expected divergence between two randomly chosen copies of an element (D) is

$D=2Tv,$

T is the average number of generations since the most recent common ancestor, v is the mutation rate. T can be further defined by

$T=n(1+q)/2\mu$

μ is the rate of transposition and $q = 4\ Ne\mu$. Ne is the effective population of haploid genomes. As q gets large, $T \approx 2Nen$, the copies of the element are behaving like alleles at a single locus. As q gets small, $T \approx n/2\mu$. The copies are essentially independent loci. D would be $4Nenv$ or nv/μ, respectively. If transposition is high (the first case), the expected divergence is equal to the expected divergence for a single copy gene in a population of size Nen. The expected average divergence between two random isolates should be greater than that of a single gene because the homogenizing effects of drift would be smaller because of the larger effective "population". If transposition is low (the second case), the expected divergence approaches infinity, i.e. there should be no relationship between random isolates. Important assumptions

of this model are that all copies of the element are equally likely to transpose and that transposition rates are constant.

Sequence data obtained from *uhu* do not meet these expectation (Brezinsky *et al.*, 1993), showing a smaller divergence time then predicted. This may indicate a violation of the assumptions, either the effective copy number of *uhu* is smaller than the total copy number or that the rates of transposition have not been constant through time or equal in all species.

A portion of *uhu* elements has been sequenced in 5 species of the planitibia sub-group (Brezinsky *et al*, 1993). The isolates from *D. heteroneura* and *D. silvestris* cluster together, as do the isolates from *D. differens* and *D. planitibia*. The isolates from *D. picticornis* are equally distant from each other as they are from the isolates from the other species. The divergence between the isolates from *D. heteroneura* and *D. silvestris* is slightly greater than the divergence between the Adh regions of the species. The sequences of the isolates from *D. differens* and *D planitibia* are virtually identical, showing a much smaller divergence than the Adh region between the two species. This could indicate a recent hybridization event between the two species, resulting in an expansion of a single copy of the element throughout the genome, replacing the other copies of the element. More isolates will need to be sequenced to address this possibility. An alternative, though unlikely, explanation is that the selective pressures on the Maui Nui group of islands work to homogenize the *uhu* elements in species on these islands. Comparing the sequences from isolates of *D. adiastola* and *D. peniculipedis*

should address this possibility. A final possibility would be a horizontal transfer through some unknown vector.

This work looked at the distribution of *uhu* in select species of two evolutionary lineages of the picture-winged group, the *planitibia* and *adiastola* sub-groups and from *D. mimica* of the modified mouth-parts group of the Hawaiian *Drosophila.* The two picture-winged subgroups were chosen because they can provide a parallel test of the relationship of transposition and speciation. Both groups have representative species on each of the major islands (expect O'ahu) whose relationship can be inferred from chromosomal inversions.

D. picticornis (endemic to the rainforests of Kauai), *D. planitibia* (Maui), *D. differens* (Molokai), *D. heteroneura* and *D. silvestris* (The Big Island) were chosen from the planitibia sub-group. This group is the best studied of the Hawaiian *Drosophila.* The evolutionary relationships of these species have been studied at the cytological, morphological, behavioral and molecular level. *D. planitibia, D. differens, D. heteroneura and D. silvestris* are homosequential in their polytene chromosomal banding pattern, but *D. silvestris* has 12 polymorphic inversions not found in the other species These species have 13 fixed inversions with respect to *D. picticornis* (Carson, 1983). Previous work suggests that the *uhu* elements in *D. planitibia, D. differens, D. heteroneura* and *D. silvestris* are active while in *D. picticornis* the copies appear to be degenerative (Brezinsky *et al.,* 1993).

The species *D. ornata, D. adiastola, D. peniculipedis* and *D. setosimentum,* of the adiastola sub-group, were chosen because they are

similar in their geographic distribution to the planitibia group species. Like *D. picticornis, D. ornata* is found in the montane rainforests of Kauai. *D. adiastola* is endemic to Maui and Lanai, *D. peniculipedis* is endemic to Maui and *D. setosimentum* is endemic to the Big Island. In addition, *D. adiastola* and *D. peniculipedis* have 13 fixed inversions when compared the *D. ornata* . *D. setosimentum* has ten fixed inversions with respect to *D. adiastola* and *D. peniculipedis* as well as 13 unique polymorphic inversions (Carson, 1983). As in the planitibia sub-group, this may indicate an increase in the activity of transposable elements in the species found on the younger islands.

D. mimica, a modified mouth-part species endemic to the Big Island, is equally diverged from the nine picture-winged species. Cytologically, *D. mimica* has over 45 fixed inversion differences from the planitibia and adiastola subgroups and is considered distant from the picture-winged (Carson, 1983). Sequence divergence of the *Adh* region suggests that the lineage leading to *D. mimica* diverged from the picture-winged around 7 million year ago (Thomas and Hunt, 1993). Southern blot analysis and *in situ* hybridization to polytene chromosomes shows *D. mimica* to have a high copy number of *uhu*. Stocks of all of the species are available except the *D. mimica* and *D. ornata* . Collection sites are known for these and wild-caught samples were used.

MATERIALS AND METHODS

ANIMALS

Ten different species of Hawaiian *Drosophila* were used for this study (Table 2). Laboratory stocks are available for seven of these species: *D. adiastola* W79B3 *D. peniculipedis* Y18P8, *D setosimentum* Y36, *D. picticornis* U71J1, *D. heteroneura* W33B3, Q71G2, *D. planitibia* U84Y, *D. differens* U43V1, *D. silvestris* U26B9, U28T2, U34B4, W12B7, Y46R9. The collection sites of the *D. silvestris* stocks is given in Table 3 and Figure 3. The U28T2 stock of *D. silvestris* and the Q71G2 stock of *D. heteroneura* was used for the molecular work. The other stocks of *D. silvestris* and the W33B3 stock of *D. heteroneura* were used for population analysis. Unless otherwise noted these stocks were used as the source of all the materials used in the following procedures. Stocks that are now extinct are: W33B3 and Q71G2 for *D. heteroneura,* U84Y for *D. planitibia* and U43V1 for *D. differens*. Wild-caught or laboratory F_1 were used for *D. mimica* and *D. ornata.* All laboratory stocks are isofemale lines except for *D. picticornis,* which is an isofemale line derived from a mass reared stock. F_1's from *D. mimica* and *D. ornata* are from mass cultures. Wild-caught *D. silvestris* from one population and wild-caught *D. picticornis* from two populations were used in examining the population distribution of *uhu* .

All stocks are kept under standard conditions (Kaneshiro, 1976).

COLLECTION OF WILD FLIES

Wild *D. mimica* were collected from a population at Bird Park in Hawaii Volcanos National Park on the Big Island. The flies were kept in mass culture

under standard laboratory conditions. Upon the production of larvae the adults were stored at -70°C until needed for DNA extraction.

Four wild female *D. ornata* were collected a from a population in the Alakai Swamp on April 8, 1991. The females were kept in mass culture under standard laboratory conditions. When the females started producing eggs, they were separated into individual vials. No further larvae were produced, and the females died shortly afterwards. They were stored at -70°C. The larvae were smeared.

Several wild-caught females of *D. silvestris* were collected from a population near the Kulani cone by Hampton Carson. One of these females produced larvae that were used in this study.

ISOLATION OF DNA FROM *DROSOPHILA*

DNA from the following species was already isolated: *D. silvestris, D. heteroneura, D. planitibia, D. differens* and *D. picticornis* , *D. adiastola* and *D. mimica* (Bishop and Hunt, 1988). The method used for this DNA isolation is given in Appendix C. DNA from *D. peniculipedis* was isolated using this method by Shane Gilmore. Additional DNA from *D. mimica* was needed during the course of this study and was isolated using the Lifton method (Appendix D). DNA to make genomic libraries for *D. setosimentum* and *D. ornata* was isolated using the Lifton method modified for small numbers of flies (Appendix E). Ten individuals from the laboratory stock of *D. setosimentum* were used, and two wild caught females from *D. ornata* were used for this procedure.

GENOMIC LIBRARIES

Genomic libraries were already available for the following species: *D. silvestris, D. heteroneura, D. planitibia, D. differens, D. picticornis, D. mimica* and *D. adiastola*. The making of these libraries has already been described (Thomas and Hunt, 1991). They were made by the method described in Promega protocol guide, (Titus, 1991) using a modified lambda bacteriophage EMBL 3. Genomic libraries were made from about 50 *D. peniculipedis* by Shane Gilmore, 10 *D. setosimentum* and 2 *D. ornata*. The libraries were made using the method in Appendix F using EMBL 4 as the vector, another modified lambda bacteriophage. Estimates of the concentration of DNA for the *D. setosimentum* and *D. ornata* genomic DNA preparations were made using the commercially available DNA dipstick. This method was chosen because of the small volumes needed to perform the assay. The timing of the partial digests of *D. setosimentum* and *D. ornata* and the concentration of MboI used was the same as was determined for *D. peniculipedis*. This assumed that the number and frequency of MboI sites was similar between the three species.

ISOLATION AND PLAQUE PURIFICATION OF PHAGE CONTAINING SEQUENCES HOMOLOGOUS TO *UHU*.

E. coli strain K802 were grown overnight, in NCZYM media to a optical density 600 of ~6. 100 µl bacteria culture, 100 µl dilution of phage library and 100 µl SM buffer were mixed and incubated for 20 minutes at 37°C. 3 ml of 0.8% agarose/NCZYM media at 42°C was added to the phage/bacteria cocktail and poured immediately and smoothly onto a 100 mm diameter petrie dish with 1.5% agar/NCZYM. After the top agarose had hardened the plates were incubated,

upside down at 37°C. The dilution of the phage library was determined to produce 1000-2000 phage plaques per plate.

82 millimeter diameter nitrocellulose of nylon filters (0.45 micron) where laid on top of the plates, removed, placed in 0.5 M NaOH 1.5 M NaCl to denature the DNA. The pH of the filters was neutralized by placing the filters in 100 mM Tris pH 7.5 150 mM NaCl for at least 2 minutes followed by a final wash in 2X SSC. Nitrocellulose filters were air dried and then baked, under vacuum for 2 hours at 80°C to bind the DNA to the filters. The DNA was cross-linked to the nylon filters by exposure to ultraviolet light using a Stratagene cross linker. Replicated lifts were done for each plate.

The filters were probed with a plasmid containing a portion of the *uhu* sequences that had be labeled with digoxygenin using a random priming method (Appendix F). Hybridization conditions were using the method of Church and Gilbert (1984) (Appendix G). Detection of hybridization is also given in Appendix F. Hybridization temperatures for species within the plantibia subgroup was at 60°C, for species outside the planitibia subgroup was 50°C. All washes were in 4X SSC at hybridization temperature.

Plaques that gave a positive signal were isolated from the plate using a sterile Pasteur pipette and suspended in SM buffer over a drop of chloroform. The procedure was repeated, but at a lower titer of phage, expecting only 100-200 plaques per plate. A single, well isolated colony was isolated as before. The procedure was repeated until all colonies gave a positive signal when probed with *uhu*.

The individual clones were then plated at a concentration to give confluent plaques. Phage were harvested from confluent plates by placing 5 ml SM buffer on the plate and storing it at 4°C overnight. The SM was removed and was stored at 4°C over chloroform.

DNA was isolated either directly from the lysates from the confluent plates by the Aloha method (Appendix H) or by the phage mini-prep method (Appendix I).

SUBCLONING OF *UHU* FROM PHAGE.

Phage clones were digested using either Sal I for EMBL 3 clones or EcoR1 for EMBL 4 clones. These digests should cut the inserted DNA for the arms of the phage. The digested phage were mixed with the chimeric plasmid pZF18u that was cut with the same enzyme at a ratio of approximately 4:1 of DNA concentrations, and ligated using standard techniques. *E. coli*, strain DH5a, were transformed using standard techniques (Appendix J). Plasmid containing inserts were identified using the blue-white selection method. Plasmids were isolated using a rapid boiling mini-prep (Appendix K). The fragment that contain the sequence homologous to *uhu* was detected for both the phage and plasmid clones using standard Southern blot techniques. Detection was using non-radioactive digoxygenin probes (Appendix F).

PCR OF WHOLE UHU AND A 400 BP REGION OF THE OPEN READING FRAME

Polymerase chain reactions using two different sets of primers were done on plasmid subclones and phage clones containing *uhu* sequences and on genomic DNA. The first set of primers amplified a 400 bp region in the coding sequence of the putative transposase. These primers, the profile used and the template concentrations have already been described (Brezinsky *et al.*, 1993). Asymmetric PCR reactions for sequencing was also done with these primers. The concentration of one primer was reduced 100 fold as described in Brezinsky*et al.* (1993).

The second set of primers were designed from the terminal repeats and should amplify the entire *uhu* element. They differed by one base, to reflect the differences between the terminal repeats:

5' TAT ACA GTG TCT TAC AGC 3'

5' TAT ACA GTG TCT CAC AGC 3'

PCR reaction conditions were standard, with following reaction profile;

95°C 1 minute

40°C 30 seconds

72°C 2 minutes

for 30 cycles followed by an extension of 4 minutes.

The concentration of the primers were in great excesses. This was to give the primers a competitive advantage over the molecules being generated in annealing to the target sequence, because of the high similarity between the

terminal repeats. This effectively doubles the concentration of these sequences when compared to a more conventional PCR amplification. Hybridization between the ends of the same molecule, or between the ends of different molecules would inhibit the reaction, hence the need to provide excess primer.

EXONUCLEASE DIGESTION OF SUBCLONES OF UHU.

A series of Nested deletion subclones were generated using the commercially available Erase-a-Base kit (Titus, 1991). The kit uses Exonuclease III, which removes nucleotides from a free 5' end of a DNA helix. At 30°C, Exonuclease III removes ~200 bases per minute. The resulting single stand is removed using S1 nuclease. The plasmid is then blunt-ended and ligated. Series of overlapping deletion subclones were generated using this method for Adia2 (plasmid A2S1), Set5 (S5B-15), Mim1 (M1A), Set1 (S1E2), Silv3.

SEQUENCING

Sequencing was done using the Sequenase Kit (US Biochemical), which employs the Sanger chain termination method (Sanger *etal.*, 1977). The kit uses a modified T7 DNA polymerase. The sequencing reactions were run on a 8% polyacrylamide gel or the commercially available 5% LONG Ranger (Hydrolink) at 75 Watts.

EXOPLASMIDS

A series of overlapping deletion subclones were sequenced for Adia2, Set5, Set1, Mim1 and Silv3. Sequencing reaction was primed using primers that were homologous to the plasmid near the cloning site, the M13 universal primer and the reverse primer. The sequencing reaction proceeding into the cloned

DNA. From the nested deleted clones, a series of overlapping sequences were obtained. The minimum overlap was about 50 bases. The average overlap was greater than 100 bases. The series of overlapping sequences covered the length of the cloned *uhu* element. Discrepancies between the overlapping sequences, compressions, single base deletions or insertions when compared to the reference *uhu* sequence (Het1) were resolved using dITP reaction conditions.

INTERNAL PRIMING OF SUBCLONES

Sequence was obtained for a portion of the open reading frame using a primer with homology to the sequence at bases 651 to 675 in the *uhu* sequence. The sequence of the primer was CAG GTG CAG GAT GAA ATG GGG. This primer was also used in the PCR amplifications described above. This was used to obtain sequences from plasmids that were known from Southern Blot analysis to contain *uhu*.. The following sequences were obtained using this method: Silv13, Plan11, Adia1, Set3, Pen1 and Pen4. Compressions, single base deletions or insertions when compared to the reference *uhu* sequence (Het1) were resolved using dITP reaction conditions.

DIRECT SEQUENCING OF ASYMMETRICAL PCR PRODUCTS.

DNA sequence for Orn1 and Adia5 was obtained by directly sequencing asymmetrical PCR product by the method described by Brezinsky et *al.* (1993).

CLONED PCR PRODUCTS

The sequence from Mim2 was obtained from PCR product that was cloned into the PCR-script plasmid (Stratagene) using manufactures instructions. Sequencing was primed using the plasmid based primers described above.

READING DNA SEQUENCES AND ANALYZING DNA SEQUENCES.

DNA sequences were read directly into the computer using a program for the Graphbar sonic digitizer written by Dr. John Hunt. Sequence homology and sequence manipulation was obtained using the DSPA program by Christian Marck (1986).

SEQUENCE ALIGNMENT

The sequences were aligned using the Clustal V Multiple Sequence Alignment Program (Higgins and Sharp, 1988). The alignment was corrected by comparing the sequences by eye.

PHYLOGENY CONSTRUCTION

Phylogenies were obtained using the Phylip package of programs (Felseinstein, 1993). DNA distances were obtained using Kimura's Two Parameter method (Kimura, 1980). Neighbor joining and maximum likelihood methods were used to construct phylogenies based of the DNA distances. A bootstrap analysis was done for the neighbor joining tree. The significance of the shortest parsimony trees was tested using the method of Templeton (1983).

SYNONYMOUS-NONSYNONYMOUS COMPARISON OF READING FRAME.

The synonymous and nonsynonymous rates of change for the open reading frame were obtained using a the LWL85 program by Li, Wu and Luo (1985), which is available on the Med School Vax.

IN SITU HYBRIDIZATION OF ELEMENTS TO POLYTENE CHROMOSOMES

In situ hybridization of *uhu* to polytene chromosomes of third instar larvae will be done for all ten species. A procedure has been developed that has given reliable results for representatives from the three subgroups (Appendix F).

The probe was constructed by labeling a plasmid containing a *uhu* by random priming, or by incorporating digoxygenin labeled nucleotides during a PCR amplification. PCR conditions were the same as those described above. Commercially available digoxygenin labeled nucleotides mixture was used in a 3:2 ratio with non-labeled nucleotides in this procedure. The copy number and insertion sites for the 5' and 3' ends of the *loa* element (Figure 4) determined for the planitibia group species by *in situ* hybridization using probes labeled using random priming.

RESULTS

THE DISTRIBUTION OF *UHU* AND *LOA* IN THE HAWAIIAN *DROSOPHILA*

The *uhu* transposable element is found in several representatives of the Hawaiian *Drosophila*. Figure 5 shows a HinD III restriction digest of genomic DNA probed with the *uhu* element. Distinct bands are seen in the six representatives of the picture-winged group of Hawaiian *Drosophila* (*D. silvestris* Hilo and Kona side populations, *D. picticornis, D. grimshawi, D. adiastola* and *D. setosimentum*) as well as in the modified mouthparts (*D. mimica*) and the antopocerus subgroups (*D. adunca*). No hybridization is seen in the fungus feeder (*D. nigra*), the Sophophora (*D. melanogaster* and *D. mauritiana*) or in the non-Drosophilid Dipterin (*Ceratitus capitiata*). Hybridization is seen with *Scaptomyza albovittata,* but there are no distinct bands. The condition of the *Scaptomyza* DNA was very poor and degraded. The hybridization may be explained by non-specific binding to the degraded DNA. The banding pattern is different between all of the species, as well as between the two populations of *D. silvestris*.

Approximately 2 μg of DNA was loaded in each lane, so the amount of hybridization should be an indication of relative copy number between the species. Using this criterion *D. mimica* should have the highest copy number of *uhu, D. silvestris, D. grimshawi, D. setosimentum* and *D. adiastola* having similar copy numbers, *D. picticornis* having slightly less and *D. adunca* having the lowest copy number.

Table 4 shows the copy number of *uhu* and of *loa* in 5 species of the planitibia subgroup, 4 species of the adiastola subgroup and *D. mimica* of the modified mouth parts based on *in situ* hybridization to polytene chromosomes. For the picture-winged species, there is general agreement between the relative copy numbers obtained by Southern blot analysis (Figure 5) and *in situ* hybridization. There is a disagreement between the relative copy numbers for *D. mimica,* which shows a relatively high copy number based on Southern blot analysis, but a relative low copy number by *in situ* hybridization. This could indicate that a large proportion of the uhu elements are in the heterochormatin, which does not polytenize, in *D. mimica.*

In both the planitibia and adiastola subgroups, the copy number of *uhu* is inversely proportional to the age of the islands on which the species is found. In both cases the species found on Kauai have the lowest number of *uhu*, while the species on the Big Island have the highest copy number. This pattern is repeated again in the planitibia subgroup by another transposable element, the *loa* element (Table 4). The *loa* element is not found in the adiastola subgroup. The *loa* element is a LINE element without terminal repeats (Felger and Hunt, 1993) and is believed to integrate into the genome in a 3' to 5' direction. By probing with both the 3' and 5' ends of the element in the same individual (Figure 6), *D. picticornis* has no complete copies of the element, so the element is not active in this species. In the other species, a third of the elements are potentially complete, so there is still a possibility that *loa* is active in these species.

The percentage of insertion sites that are variable for the presence or absence of an element was determined by comparing the distribution of *uhu*

along the arms of polytene chromosomes (Figure 7). Two individuals from the same laboratory population was examined for *D. differens* and *D. planitibia.* Two laboratory populations are compared (two individuals from each population) for *D. silvestris* and *D. heteroneura.* Two individuals from a laboratory population and an F_1 from a wild-caught female were compared for *D. picticornis.* All the laboratory populations are isofemale lines. In the species from Maui Nui and the Big Island, about 50% of the sites are variable for the presence of an *uhu* element (Table 5). These estimates are from an inbred line and are probably underestimates of the variability in occupancy of a site by *uhu* in these species. When populations are compared, the percentage of variable sites is higher, supporting this idea. Less than 15% of the sites occupied by an *uhu* element in *D. picticornis* are variable. This comparison is between a wild-caught F_1 and a laboratory population that was established from a female that was collected at a different site which is at a higher elevation, this value is a good estimate of the variability in the *uhu* site occupancy for *D. picticornis.*

The copy number of the *uhu* element on individual chromosomes in 14 individual *D. silvestris* from five old laboratory populations and an F_1 from a wild-caught female from a new population is shown in Table 6. The laboratory populations are isofemale lines that were collected at different locations on the Big Island. The original populations have different frequencies of several polymorphic chromosomal inversions (Craddock and Carson, 1989). A majority of the polymorphic inversions have been maintained in the laboratory stock. Because there is evidence that transposable elements accumulate in areas around polymorphic inversions because of reduced recombination, an increase in the copy number of *uhu* might be expected on the chromosomes with a high number

of polymorphic inversions. Nevertheless, no heterogeneity in the copy number was found on any of the chromosomes (G_i) nor in the total number of elements per individual (G_T). ΣG_i, which tests whether there is a trend to deviate from random, are also not significant. ΣG_i-G_T, which tests whether any deviations are homogeneous, is also not significant.

SEQUENCE COMPARISON OF *UHU* FROM ISOLATES FROM PLANITIBIA SUBGROUP SPECIES, ADIASTOLA SUBGROUP SPECIES AND *D. MIMICA*, A MODIFIED MOUTHPART SPECIES.

The sequence from isolates of *uhu* from *D. silvestris, D. adiastola, D. setosimentum* and *D. mimica* is shown in Figure 8 compared to the three complete *uhu* sequences from *D. heteroneura* already reported by Brezinsky *et al.* (1993). The sequence from the silvestris 3 (Silv3) isolate is complete. The setosimentum 5 (Set5) and mimica 1 (Mim1) sequences are degenerative, containing deletions, including the loss of one of the terminal repeats. The adiastola 2 (Adia2) and setosimentum 1 (Set1) were truncated during the original cloning. The AUG start codon reported by Brezinsky *et al,* (1989) is located at base 394. Mim1 is deleted in this region and Set5 has a point mutation in the third position, going from AUG to AUA. The stop codon of the *Tc1* element of *Ceanorhabditis elegans* is located at 1430. Mim1 again is deleted in this area and Set5 has a point mutation in the first position, changing to an AAA from TAA. An alternative stop occurs in frame 11 amino acids downstream. In the initial comparison to the open reading frame of Tc1, the *uhu* open reading frame had a 10 bp imperfect duplication at base 1136 to 1145. This produced a three amino acid duplication and a frame shift that resulted in a premature termination of the

reading frame in comparison to Tc1 (Figure 9). This duplication occurs in a region of high sequence homology amongst all of the Tc1- like elements, suggesting a functional importance. The Mim1 sequence stops at base 1131 and Adia2 has a deletion from 1133 to 1308. Neither Silv3 and Set5 have the 10 base pair duplication. The 10 bp duplication has been found in isolates from *D. planitibia* and *D. silvestris*. (Figure 9). The duplication is not found in an isolate from *D. peniculipedis*. The three heteroneura *uhu* sequences also had a 12 bp deletion, further downstream from the duplication, at base 1284, when compared to the Tc1 open reading frame. Again, neither Silv3 or Set5 have this deletion, resulting in an open reading frame that is similar to that of Tc1.

Adia2 has several small deletions at the start of the open reading frame that results in a frame shift and early termination of the reading frame. Set5 has several nonsense mutations throughout the reading frame as well as a single base deletion, and frame shift, near the end of the reading. Silv3 has a 4 bp insertion resulting in a frame shift (Figure 10).

Figure 11 shows the synonymous versus nonsynonymous distance matrix for the open reading frame of the seven isolates of *uhu*. The sequences were first corrected (inserts and duplications removed, dashes put in for deletions) to give the alignment of amino acids. A synonymous substitution does not result in an amino acid change, a nonsynonymous substitution changes the amino acid. The nonsynonymous distance should be smaller than the synonymous in regions under selective pressure. The synonymous/nonsynonymous ratio is highest for the pairwise comparison of Het4, Silv3 and Mim1, indicating these were the most recently active elements. The synonymous/nonsynonymous ratio is nearly 1 for

all of the comparisons with Adia2 and Set5, indicating that the sequence has been free to diverge for a longer period of time.

The pairwise distances for the 5' noncoding region are given in Figure 12 in comparison to the synonymous and nonsynonymous distances The distance for this region tends to be closer to the synonymous distance of the open reading frame. This would indicate that the DNA sequence in the 5' non-translated region is not under selective pressure.

There are varying degrees of match between the inverted terminal repeats within an element (Figure 13). Brezinsky *et al.* (1990) suggested that the terminal repeat is 47 base pairs, but the actual ends of the element are uncertain. The terminal repeats are presented with 51 base pairs and the degree of match from the additional 4 bases is very low. Using the 47 bp terminal repeats, the terminal repeats of Silv3 have only one difference. The terminal repeats of Het3 and Het4 have 4 mismatches. The reading frame of Silv3 and Het4 imply they were the most recently active of the elements. This suggests that the degree of homology between the terminal repeats may also be important in determining the activity of an element. Of the elements where only one terminal repeat is available, Adia2 and Mim1 conserve the size of the repeat, Set5 has a deletion in the 3' terminal repeat.

Two features in the terminal repeats appear to be conserved between *uhu, Tc1, Tcb1* and *Hb1* (Brezinsky *et al.*, 1990). The first is the sequence CAGTG or CAGTA near the beginning of the repeat. The second is an A-T rich region near the end of the repeat. These features also appear in the terminal repeats of the *Bari* element and the *Minos* element, two other Tc1-like elements. The sequence

CAGTA also occurs in the 5' splice sight of the small t-antigen of Simian Virus 40 (SV40). Lee and Barton (1993) have demonstrated that a 17 base pair region containing the CAGTA motif at the SV40 intron splice sight forms a three-dimensional structure that maybe important in the splicing of the intron. The *uhu* terminal repeat shows 59% sequence homology to this 17-mer, and 46% homology when the CAGT sequence is removed. The CAGTG or CAGTA motif, or one with only a single base change, is also found in several other transposable elements (Figure 14). Of the 14 transposable elements with short inverted repeats examined, only the P element of *D. melanogaster* and an unnamed element from *Salmonella typhimurium* do not contain the CAGTG or CAGTA motif. A CAGTA motif does occur in the 3' terminal repeat of the *S. typhimurium* element. Only 44 out of 90 random sequences 30 bp long contained this motif. Using Fisher's Exact test, the CAGTG motif occurs in terminal repeats of transposable elements more frequently than would be expect by chance ($p = 0.009$, one-tailed test).

POLYMERASE CHAIN REACTION AMPLIFICATION OF UHU ELEMENTS USING PRIMERS TO THE TERMINAL REPEATS.

It is possible to amplify *uhu* elements using primers to the terminal repeats from both plasmids (Figure 15) and genomic DNA (Figure 16). A fragment of DNA of around 1.7 kb was amplified from plasmid sub-clones P1D and H3K, from the species *D. planitibia* and *D. heteroneura* respectively. The presence of sequence homologous to *uhu* in these plasmids was confirmed by sequencing from the original plasmid. A 1.7 kb fragment was also amplified from plasmid O6E3, from *D. ornata*. It was not possible to confirm the presence of *uhu* in this

plasmid. A 400 bp fragment was amplified from plasmid S5B, from *D. setosimentum*. This fragment was cloned and sequenced. It corresponded to a 3' *uhu* end from base 1298 to the terminal repeat. Further sequencing showed only weak homology at base 1298 to the primer, primarily to the 3' end, and that the 5' terminal repeat was not present in the plasmid.

A 1.7 kb fragment was amplified from *D. silvestris* genomic DNA. A weak 500 bp band may also be present. No 1.7 kb band was amplified from *D. setosimentum* and *D. picticornis*, though a band of about 700 bp was seen in *D. setosimentum*. Sequencing was not done to confirm that it was *uhu* that was amplified.

SEQUENCE ANALYSIS OF A 400 BP REGION IN THE OPEN READING FRAME

A 400 bp region of the putative open reading frame was analyzed in 12 isolates of *uhu* from 5 species in the planitibia subgroup (Brezinsky *et al.*, 1993). This work adds sequence from this region for twelve more isolates; one each from *D. silvestris* and *D. planitibia* of the planitibia subgroup, 8 from species in the adiastola subgroup, 3 from *D. adiastola*, 2 each from *D. setosimentum* and *D. peniculipedis* and one from *D. ornata*, and two from *D. mimica*. (Figure 17). The synonymous/nonsynonymous rate of change for these isolates is presented in Figure 18. Synonymous rates of change that are 2.5 times greater than the nonsynonymous rates are underlined and shaded. Synonymous rates that are twice that of the nonsynonymous rates are underlined. A higher rate of synonymous substitutions indicates selective pressure to maintain the amino acid sequence. Four isolates; Pen, Het4, Silv3 and Mim1, consistently show a synonymous rate twice that of the nonsynonymous rate. These are probably the

most recently active elements. There is great variability in the synonymous/nonsynonymous ratio between the isolates. This indicates that selection has not been constant between the elements. This has probably been the result of the elements losing their functional transposase at varying times in the past.

A phylogenetic analysis for the 24 isolates is presented in Figures 19, 20 and 21. These phylogenic trees were made using three different algorithms, each using different assumptions. In this analysis, the Neighbor Joining (Figure 19) and the Maximum Likelihood (Figure 20) phylogenies are based on Kimura's Two Parameter distance measure, which weights transversions and transitions; deletions are ignored. The branch lengths along the trees are proportional to the estimated genetic distance. The Neighbor Joining tree presented is a consensus tree of a hundred trees generated by randomly sampling portions of the data in a bootstrap analysis. The values at the nodes indicate the number of times that node occurred with all of the species to the right of it, out of 100 trees. Values greater than 50 are considered significant. The parsimony analysis (Figure 21) counts all changes, including deletions. No distances are calculated. The trees produced are different between the three methods. Two clusters are consistent among the three trees; [Het4, Het3, Het1, Silv2, Silv13] [Diff1, Diff2, Plan2, Plan3, Plan4] as is the sister group [Mim2, Plan11]. These clusters have the highest bootstrap values in the Neighbor Joining tree, and the confidence limits on these nodes do not overlap with other nodes in the Maximum Likelihood tree.

There are eight other trees of equal length to the parsimony tree presented (Figure 22). They differ in the branching order within the Het/Silv and

Diff/Plan clusters and the placement of Silv3. In both the Neighbor Joining tree and the Maximum Likelihood tree, which use genetic distances, the branch lengths to the nodes are much shorter than the branch lengths leading to the individual taxa. All of trees are unrooted, but are presented with Mim1 as an arbitrary outgroup for ease of comparison.

DISCUSSION

ANCIENT ORIGIN OF THE *UHU* TRANSPOSABLE ELEMENT IN THE HAWAIIAN *DROSOPHILA*

The *uhu* transposable element has a broad distribution in the Hawaiian *Drosophila.* It has been found using Southern blot analysis in every member of the picture-winged group examined. This includes members of the planitibia, adiastola and grimshaw subgroups. The major groups of the picture-winged species are believed to have diverged from each other about 5 million years ago (Mya). This distribution of *uhu* is suggestive that it was present in the genome of an ancestral species of the group, and should be present in every species in the group. The *uhu* element is also present in *D. mimica* of the modified mouth-parts group and the antopocerus *D. adunca,* but not in *D. nigra* of the fungus feeder group or *Scaptomyza albovittata,* a Hawaiian scaptomyzoid. The modified mouth-parts diverged from the picture-winged group 5 to 7 Mya, the fungus feeders diverged ~10 Mya and the scaptomyzoids diverged 25-30 Mya. Mitochondrial DNA analysis groups the antopocerus with the modified mouthparts (DeSalle and Giddings, 1986)

If this distribution of the element is maintained as more members of the modified mouth-parts, fungus feeders and Scaptomyzoid groups are examined, it would suggest that *uhu* either arose, or was somehow transferred into the lineage leading to the modified mouth-parts and picture winged sometime after the divergence of the fungus-feeders, but before the divergence of the modified mouthparts. This would place the origin of the *uhu* element in the Hawaiian

Drosophila between 7 and 10 Mya. An alternative explanation is that *uhu* is in the fungus-feeders and the scaptomyzoids and is too diverged to be detected with the techniques used or that it has been lost in *D. nigra* and *S. albovittata.* Robertson (1993) has used degenerative PCR primers to find sequences homologous to the *mariner* element in six orders of insects. Such a method may useful in looking for *uhu* in the fungus-feeders and *Scaptomyza.*

The phylogenetic analysis of 24 isolates of *uhu* from 10 species also suggests an ancient origin for the element in these species. In all three of the trees produced, there is a lack of clustering of the copies of *uhu* into species groups. Some copies of *uhu* are more closely related to copies in another species than they are to copies within their own species. This implies that the copies of *uhu* diverged from each other before the species themselves diverged, and more than one copy of *uhu* was passed vertically from the ancestral species to the daughter species. For the two trees based on distance measures (Figures 19 and 20), the branches leading to the nodes are short when compared to the length of the branches leading to the individual copies of *uhu*. This is consistent with an ancient divergence of the isolates from each other. It also argues against a recent horizontal transfer of the element between species explaining the distribution of the element. If that was the case the branch lengths leading to the isolates would be relatively short.

The distance matrix for 24 isolates of *uhu* from 10 species of Hawaiian *Drosophila* for a 400 bp region of the putative open reading frame is presented in Figure 18. The synonymous distances are above the diagonal, nonsynonymous changes are below. If the sequence is under selective constraints, the

synonymous substitutions should be retained more frequently and the ratio of synonymous to nonsynonymous should be greater than one. There is great variability in this ratio between all the isolates ranging from 0 to 4.8. The synonymous/nonsynonymous ratio for the Adh coding region for several of these species are all around 7 (Thomas and Hunt, 1991), showing similar selective pressures on Adh in the species. The varying ratios for the different isolates of *uhu* suggests there has been different selective pressures on the isolates. This is probably the result of the elements losing the ability to autonomously transpose at different times in the past. Four isolates consistently show high synonymous/ nonsynonymous ratios; Pen1, Het4, Silv3 and Mim1. This is indicative that these elements, of the isolates under study, were the most recently capable of autonomous transposition. Three of these four isolates are from Big Island species and the third is from a Maui species (Pen1). This is what would be expected if the *uhu* has been more active in the species on these islands than in the species on Kauai. The maximum distance between two isolates, Mim1 and Pen4, is 0.778. This is greater than the divergence of the Adh region between *D. heteroneura* and *D. nigra*. If we assume that copies of *uhu* are diverging at the same rate as the synonymous rate for Adh, these two isolates diverged from each other ~20 Mya (Thomas and Hunt, 1991). This places *uhu* in the Hawaiian *Drosophila* before the divergence of the fungus-feeders. This is in disagreement with the current biogeographic data, which places the arrival of *uhu* between the divergence of the fungus-feeders and the modified mouthparts. This implies that *uhu* was lost in the lineage leading to *D. nigra*. Other species of fungus-feeders may still have *uhu*, and a broader survey of species in this group needs to be done.

The cluster analysis of the copies of the *uhu* element was done using three methods; Neighbor Joining, Maximum Likelihood and Parsimony. The Neighbor Joining and Maximum Likelihood build trees based on genetic distances. Kimura's Two Parameter measure was used in both cases. The Parsimony analysis minimizes the total number of changes. The three methods gave three different trees. The trees have three species clusters in common. The first is the Het/ Silv cluster grouping of [Het1, Het3, Het 4, Silv2, and Silv13], the second is the clustering of [Plan2, Plan3, Plan4, Diff2, and Diff3] and the final is the pairing of [Mim2 and Plan11]. In the bootstrap analysis of the Neighbor Joining tree, these are the only nodes with a bootstrap value greater than 50. These are also the only nodes in the Maximum Likelihood analysis where the 95% confidence intervals do not overlap with neighboring nodes. Nine trees of equal length are produced by parsimony analysis. They differ primarily in the branching order of the isolates within the first two clusters. The branching of the isolates outside these clusters varies greatly between the trees produced by the three methods. In none of the trees produced do all the isolates from one species or subgroup cluster together.

There are several factors that may explain the differences between the trees. Several of the sequences have deletions or unresolved sequences. The deletions are probably analogous, not homologous. Parsimony analysis does not distinguish between the two, thus degenerative sequences will cluster together in the parsimony analysis. The deletions may also cause problems in the distance based phylogenies. Rates of change may not be equal along the length of the sequence analyzed. Distances measured between isolates will vary depending on the deletions within the sequences. Another concern is that some isolates

have been free to diverge, while others have been constrained to maintain an active transposase, the constraint on the elements may have changed (been removed) at varying times. Thus there is a violation of the assumption of an equal rate of sequence divergence in all lineages and the rate of sequence divergence has not been equal in all lineages. Two isolates that have been under selective pressure may appear more similar to each other than to an isolate that has been free to diverge, even though this may not be the true phylogeny. Using distance estimates based on synonymous changes may minimize this affect, but the standard errors on the distance estimate also increases. Because of these concerns, it may not be possible to obtain the true phylogenetic relationships of the isolates.

In both of the distance based phylogenies, the distance to the nodes is much less than total distance to the isolates, indicating that the isolates have been diverging independently from each other for the majority of their history. This and the failure of the isolates to cluster nicely within a species or subgroup suggests that the *uhu* element was present in a common ancestor of all of the species. The average synonymous distance between isolates of *uhu* from *D. adiastola* and *D. heteroneura* is very similar to the synonymous distance for the Adh region between the to species (0.22 and 0.189 respectively), indicating the isolates of *uhu* have been diverging from each other for as long as the Adh region. The synonymous distance for *uhu* between *D. mimica* and *D. heteroneura* is much greater than the synonymous distance for Adh (0.34 and 0.15, respectively) as is the comparison between *D. mimica* and *D. adiastola* (0.46 for *uhu*, 0.18 for Adh). The isolates of *uhu* appear to have diverged from each before

the species diverged. This supports that idea that *uhu* was present in the common ancestor of these species.

The synonymous/nonsynonymous ratio of Plan2, Plan3, Plan4, Diff2 and Diff3 against the other isolates are consistently near one, suggesting that these elements are inactive and the open reading frame has been free to diverge for a relatively long period of time. The overall genetic distances between these isolates is very small, indicating that they are recently diverged from one another. Together, this leads to the conclusion that these isolates represent the replicative transposition of an "inactive" copy of the element, a copy of the element that is not capable of making its own transposase. This implies that in the genome of *D. planitibia* and *D. differens* there is a copy of *uhu* with an intact reading frame. It also demonstrates that the transposase of *uhu* is trans acting, and is capable of mobilizing some crippled elements. This is consistent with what is known from other transposable elements. The activity of the Tc1 element, for example, has been shown to be variable depending on the genetic background, which indicates that the activity is dependent on other copies of Tc1 (Collins, 1987). The sequence divergence between the these five isolates of *uhu* is less than the divergence between the Adh locus of *D. planitibia* and *D. differens* (Brezinsky, *et al.*, 1993). The synonymous/ nonsynonymous ratio is not indicative of these isolates being under selective pressure, so selection probably does not explain the high sequence homology. Another possibility is that these isolates are allelic, occupying the same locus. This is unlikely because the restriction map of the flanking regions are different (Hunt, unpublished data).

The sequence of another isolate from *D. planitibia* (Plan 11) shows a much greater divergence. The isolates from the two adiastola groups species on Maui, *D. adiastola* and *D. peniculipedis*, also show a high sequence divergence, so if selection pressures are maintaining a high sequence homology in *uhu*, the selection is not equal for all copies of the element in species on Maui and Molokai. The high degree of sequence homology is probably indicative of a recent divergence between the copies. Thus copies of *uhu* in *D. planitibia* and *D. differens* show divergence that is less than a single copy gene for these species. This suggests a horizontal transfer of *uhu* between these two species. *D. planitibia* and *D. differens* are currently allopatric, being endemic to different islands. The two islands, Maui and Molokai, were most recently connected 15,000 years ago during the Pleistocene (Carson, 1983). The degree of sequence divergence is consistent with the transfer and amplification of a copy of *uhu* during the joining of Maui and Molokai. It is not possible to determine the direction (from *D. planitibia* to *D. differens*, or vice versa) or the mechanism of transfer (hybridization, viral transfer, etc.) from the current data.

SEQUENCE ANALYSIS OF *UHU* FROM ISOLATES FROM *D. SILVESTRIS, D. SETOSIMENTUM, D. ADIASTOLA* AND *D. MIMICA*

From the sequence data available, none of the current isolates of *uhu* are believed to be active. Brezinsky *et al.* (1993) reported the complete sequence from three isolates of *uhu* from *D. heteroneura*. These elements have a 10 bp duplication at base 1136 and 12 bp deletion at base 1284 in the open reading when their sequence is compared to the reading frame of Tc1. The 10 bp duplication resulted in a premature termination of the reading frame. The

present work adds the complete sequence of an isolate from *D. silvestris* (Silv3) as well as the incomplete sequence of isolates from *D. adiastola* (Adia2), *D. setosimentum* (Set5) and *D mimica* (Mim1). All of the new sequences have either insertions, deletions or point mutations that disrupt the open reading frame. The 10 bp duplication and the 12 bp deletion are not found in Silv 3 or Set 5. The 10 bp duplication is also missing in an isolate from *D. peniculipedis* (Pen 4). Thus a putative transposase of *uhu* is much more similar to the *Tc1* transposase then originally suspected. The ratio of synonymous to nonsynonymous distances is consistently highest for isolates Het4, Silv3 and Mim1. This suggests that these are the most recently active of the isolates.

The phylogenetic distribution of the 10 base pair duplication suggests that it is fairly ancient and has been maintained in several lineages. Sequence data confirms its presence in isolates from *D. heteroneura* (Het1, Het3, Het4), *D. silvestris* (Silv13) and *D. planitibia* (Plan11). Phylogenetic analysis suggests that it could be present in an isolate from *D. mimica* (Mim2). It is not present in isolates from *D. silvestris* (Silv3), *D. setosimentum* (Set5) and *D. peniculipedis* (Pen4) Thus, at least in *D. silvestris*, there are two populations of *uhu*, one with the duplication and one without. The divergence of the sequence between Plan11 and the isolates from *D. heteroneura* and *D. silvestris* suggests an ancient divergence between the elements.

In some phylogenies the duplication appears to be polyphyletic. One possible explanation for the observed phylogenetic relationship of the copies containing the duplication is a gene conversion or recombination event. To explain this data, the recombination most likely occurred in *D. planitibia*, where a

more divergent sequence was combined with the duplicated sequence. This scenario would be supported if Mim2 is found not to contain the duplication. Even under this scenario, the duplication was present before the divergence of *D. planitibia* from *D. heteroneura* and *D. silvestris* and it has been maintained in the lineages leading to the extant species. Evidence from the other isolates from *D. planitibia* and *D. differens* suggests that it is possible for copies of *uhu* which do not code for a functional transposase to increase in copy number.

The fact that the duplication has been maintained in at least three different lineages is not surprising. Another possibility is that the truncated transposase maintains some function. The high synonymous/ nonsynonymous ratio for Het4 suggests that it has been under recent selective pressure, supporting this idea. This would imply that the C-terminus of the transposase is not necessary for function. Another possibility is that the transposase is a dimer or multimer in its active conformation. This would imply only one functional C-terminus would be necessary for protein function. Since several proteins that interact with DNA form multimers, including RecA (Lewin, 1990), and there is some evidence of protein-protein interactions in the function of the transposase of the Ac element (Kunze and Starlinger, 1989), this may be the case. The region where the duplication occurs is highly conserved amongst all the *Tc1-like* elements, suggesting a functional importance. The region is rich in polar and negatively charge amino acids, so it is probably on the surface of the protein and may be involved in protein-protein interactions. The duplication of the 10 bp results in an imperfect duplication of three amino acids. The additional amino acids are polar or negatively charge. This may have the affect of increasing the protein-protein interaction, giving the truncated transposase a competitive advantage in

dimer formation or it may increase the stability of the dimer. This would help explain the maintenance and apparent selective pressures observed on the truncated transposase.

COMPARISON OF THE TERMINAL REPEATS OF ISOLATES OF *UHU* FROM *D. HETERONEURA, D. SILVESTRIS, D. SETOSIMENTUM, D. ADIASTOLA* AND *D. MIMICA* WITH THE TERMINAL REPEATS FROM OTHER *TC1-LIKE* ELEMENTS.

The *uhu* element has a 47 bp imperfect inverted terminal repeat. In three isolates from *D. heteroneura*, the repeats show vary degrees of similarity between the 5′ and 3′ repeats within an isolate. This work adds the 5′ and 3′ terminal repeats from Silv 3 and the 3' terminal repeat of Set 5 and the 5′ terminal repeats for Adia 2 and Mim 1. For the isolates where both terminal repeats are available, there is a weak correlation between the identity of the terminal repeats to each other and the fidelity of the open reading frame and the synonymous/ nonsynonymous ratio. This suggests that the degree of the match between the elements affects their relative mobility. The related element from *C. elegans, Tc1*, which has a proven mobility, the terminal repeats are perfect, supporting this idea (Harris *et al.* 1988). However, the *Bari* element from *D. melanogaster*, though mobile, has imperfect inverted repeats (Ciazzi *et al.*, 1993), so the relationship is not exact.

The terminal repeats of the *Tc1-like* elements, though differing greatly in size, ranging from 23 for the *Bari* element to 255 for *Minos*, have two features in common (Brezinsky *et. al*, 1993). They all have a AT-rich region on the inside of the terminal repeat, and they have the consensus sequence CAGTG or CAGTA. Though the mode of transposition is not known, if the terminal repeats are

necessary for transposition, these features are probably involved in the transposition process. Indeed, certain features suggest that they may be involved in a recombination like process. The AT-rich region would have a lower melting temperature and could serve as a site to separate the two strands of DNA. The CAGTA motif is found in the terminal repeats of a majority of elements with short inverted terminal repeats. As these elements would be expected to have similar methods of excision, it is possible that the area of the CAGTA consensus sequence serves as a site for nicking the DNA strand which would be necessary prerequisite for excision. Lee and Barton (1993), using Tris(4,7-diphenyl-1,10-phenanthroline)rodium(III) [$Rh(DIP)_3^{3+}$], identified sequences used in intron splicing. $Rh(DIP)_3^{3+}$ recognizes three dimensional shape of nucleic acid sequences (Pyle and Barton, 1990; Chow and Barton, 1992). It has been shown to recognize cruciform structures (Kirschenbaum *et al.*, 1988) and Holliday junctions (Lee and Barton, 1993), both of which are involved in the recombination. They conclude that the tertiary structure formed in supercoiled DNA by these sequences may be recognized by enzymes involved in recombination. The 5' intron of the small t-antigen of simian virus 40 (SV40) has the sequence CAGTA (Figure 7). When the 5′ terminal repeat of *uhu* is compared with this 17 bp CAGTA containing sequence, 59% sequence homology is observed. This suggests that the terminal repeat of *uhu* adopts a three dimensional structure that is similar to the structure at intron splice sights. Excision of element may involve a process similar to recombination or DNA repair. This is consistent with models of transposition. Current models of excision for the *Tc1* (Plasterk and Groenen, 1992) and *P* elements (Engels *et al*, 1990) suggest that the excision of the element

results in a double stranded break, that is then repaired off the homologous chromosome.

The fact that the *Tc1* element for *Caenorhabditis elegans* and the four *Drosophila* elements (*uhu, Hb1, Bari-1 and Minos*) have several similarities raises the possibility of horizontal transfer between the species (Kidwell, 1993). Beyond the conserved motives in the terminal repeats, what homology exists is in the amino acid sequence of the putative transposase (Brezinsky *et al.*, 1993). The nucleotide distances are very great between the elements. The putative transposases are equally divergent from each other (Ciazzi *et al.*, 1993). It is possible that the *Tc1-like* are related and spread by means of horizontal transmission. If this is the case, the degree of sequence divergence suggests that it happened a long time ago. It is equally possible that the similarity between the elements has been conserved from an ancestral element that was present in an ancestral species.

COPY NUMBER AND RELATIVE MOBILITY OF THE *UHU* AND *LOA* ELEMENTS IN THE HAWAIIAN *DROSOPHILA*.

The copy number of the *uhu* element was previously examined using *in situ* hybridization to polytene chromosome with a tritium labeled probe in five species of the planitibia subgroup; *D. picticornis, D differens, D. planitibia,* and *D. heteroneura.* An increase in the copy number of the *uhu* element was found that correlated with the age of the island on which the species is endemic (Hunt *et al,* 1984).

The present work examined a second species group, the adiastola group using the digoxygenin probes. It also examined several individuals from *D. picticornis* and *D. silvestris*. The copy number of the *uhu* obtained between the two techniques is similar for *D. silvestris*, but not for *D. picticornis*. The difference between the techniques may be a reflection of the low sequence homology of the *D. picticornis* element to the *D. heteroneura* element used as a probe. The tritium labeled probes were RNA, so only the areas that hybridized would have been detected. The digoxygenenin labeled probes were DNA and every molecule that hybridized would be detected, so the bands should be more intense. Some of the bands may be been missed, discounted as background using the tritium labeled RNA probe. The copy number of *uhu* shows a similar increase in the species found on the younger islands in the adiastola group species. Thus in both species groups, the species endemic to Kauai, the oldest island, have the fewest number of copies of the *uhu* element. While the species endemic to the Big Island, the youngest island, have the highest copy number. This trend is also repeated with another transposable element, the *loa* element. Within the planitibia group species, *D. picticornis* has the lowest copy number (14) and *D. silvestris* has the highest copy number (62). The other species have an intermediate copy number. The *loa* element belongs to the LINE family of elements, which are characterized by having the 5′ end of the sequence deleted in the presumably inactive element. When *in situ* hybridization is done with probes from both the 5′ and 3′ end of *loa* in these species, *D. picticornis* has no 5′ end sequences of the element indicating that *loa* is not capable of autonomous transposition. All the other species have 10 to 20 copies containing both the 5′ and 3′ ends of the element, so there are potentially active *loa* elements these

species. For both the *uhu* element in the planitibia and adiastola subgroups and the *loa* element in the planitibia subgroup there is a negative correlation between the copy number of the element and the age of the island on which the species is endemic.

Hunt *et al.*, (1984) suggested that in the planitibia subgroup, the increase in copy number of *uhu* in the species on younger islands represented a higher level of activity of the element in these species. This higher level of activity may be associated with the founding of a new population on a new island as it became habitable. The biogeography of the planitibia and adiastola subgroup, as well the relationship of the species based on chromosomal inversions, suggests that the subgroups have a similar evolutionary history. The similar increase in the copy number of the *uhu* element in both this subgroups suggests a similar cause, and is consistent with the idea of a correlated increase in the activity of an element associated with the founding of a new population. The relative copy number of the *loa* element in planitibia subgroup is also in agreement with this idea.

There are two ways to explain the higher copy number of these elements in the species found on the younger islands. The first is that there is an actual increase in the copy number resulting from the founding of a new population. The other is that these elements generally maintain a high copy number, but drift and selection against the effects of transposition have reduced the copy number in the populations in more stable environments (i.e. those on the older islands). In the first scenario, the copy number of the elements are in a transposition/ selection equilibrium. This would imply that the common ancestor of *D. plantibia*

and *D. silvestris* living on Maui had a copy number of *uhu* very similar to the current 110 seen in *D. planitibia*. When the population that gave rise to *D. silvestris* was founded on the Big Island, the change in selective pressures either allowed for, or induced, a net increase in the transposition of *uhu*. As the population of proto-*silvestris* became established, a new and higher equilibrium copy number was established, which has been maintained in the extant populations of *D. silvestris*. *D. planitibia*, which continued to evolve on Maui, maintained the equilibrium copy number of the ancestor.

In the second scenario, the copy number of *uhu* in proto-*picticornis* was high when the population was founded on Kauai. Selection against the affects of transposition have decrease the copy number of *uhu* to the current copy number in *D. picticornis*, while the changes in the selection pressure associated with the founding of a population have allowed *uhu* to maintain a higher copy number in the resulting species.

It is not possible to differentiate between these two scenarios with the current data. Under the first, a higher copy number is expected in the species on each progressively younger island, but not necessarily a predictable number. If, under the second scenario, we assume a constant rate of loss of the element, we would expect a linear relationship between the copy number and the time of divergence. In this case the copy number of *D. differens* becomes problematic, being too low, favoring the first case. The high occupancy rate for insertion sites in *D. picticornis* favors the hypothesis that *uhu* is being lost. If *uhu* is in a transposition/ excision equilibrium in *D. picticornis*, every site would be expected to have a low rate of occupancy because of the low copy number of the element.

The high site occupancy suggests that *uhu* has drifted to fixation or loss at most sites is this species. Conversely, the differences in the copy number between *D. silvestris* and *D. heteroneura* suggest that a new equilibrium is established during the founding of a new species. If *uhu* and *loa* are just maintaining a copy number, this number would be expected to be much more similar between these two species. The copy number of *uhu* and *loa* in another member of the plantibia subgroup, *D. setosifrons*, could provide more information about the evolution of transposable elements. Chromosomal and biogeographic data suggest that *D. setosifrons* diverged from *D. picticornis* within the last 0.5 million years. Current evidence suggests that both *uhu* and *loa* are inactive in *D. picticornis*. If these elements were inactive 0.5 Mya when *D. setosifrons* diverged, they should also be inactive in *D. setosifrons*. Because of the bottle neck that was probably associated with the founding of *D. setosifrons*, nearly every copy of *uhu* and *loa* would be in fixed in location. Sampling error could have increased or decreased the copy number of the elements. If these elements were active at the time *D. setosifrons* was founded, the variability in site occupancy should be high and the copy number would be expected to be higher than *D. picticornis*. Assessing the status of the elements in *D. setosifrons* will give a minimum estimate of when the elements became inactive, and how fast elements are lost from the genome once they are inactive.

The copy number of *uhu* from fourteen individual *D. silvestris* from six populations were examined. No significant difference was found between the individuals or between the populations. Because the number of polymorphic inversions is different between the populations, the reduced recombination between the heterozygotes might be expected to increase the number of *uhu* in

the areas around the inversions. Both the Maulua and Kilauea populations have several polymorphic inversions on the fourth chromosome, however, the fourth chromosome in these stocks do not have a significantly higher number of *uhu* than the fourth chromosomes of the other stocks. The consistent copy number of *uhu* between the populations suggests that the element is in a transposition/ excision equilibrium within the species. Considering that some of the stocks were established from wild populations that are less than 2100 years old (Carson *et al.*, 1990) and that the majority of the flies examined are from isofemale lines that have been in the lab for several years, this result is surprising. It suggests that the equilibrium has been maintained during the founding of the populations or that there is an "optimum" equilibrium number of *uhu* for *D. silvestris*. The copy number of a wild caught F_1 is in the middle of the range of the lab flies, arguing against the possibility of a homogenizing effect of laboratory conditions. This argues that the copy number of the *uhu* is maintained in a species and that there is not a increase in the relative transposition of the element associated with a founding event, and supports the idea that elements are being selected against and are being lost in the species on the older islands.

The variability in the occupancy of an insertion site of an *uhu* element is near 50% within a population of all species of the planitibia subgroup examined expect *D. picticornis*, where the variability in occupancy is 7.4%. All of the populations examined are isofemale lines. Because of the bottleneck associated with the founding of the laboratory stock, the degree of shared sites may be higher than in the original population, though there is some evidence that the level of variability is maintained, not reduced, in laboratory stocks (Di Franco *et al.*, 1993). In the three cases where more than one population was examined, the

level of insertion site variability within the species increases. In *D. silvestris* and *D. heteroneura* this increase is dramatic. In *D. picticornis*, though the addition of another population doubles the number and percentage of variable sites, the percentage of sites varying in the presence or absence between populations is still lower than the number of variable sites within a population of the other species. This suggests that the majority of sites occupied by *uhu* are homozygous for the presence of the element in *D. picticornis*. These elements are probably inactive and have drifted to fixation. Other inactive elements would have been lost. This implies that the copy number of *uhu* was higher, and there has been a decrease in copy number due to drift and selection against the detrimental affects of transposition in *D. picticornis* (i.e. individuals with active elements would have a lower fitness).

The large number of sites variable for the presence of an element in the other species suggests that *uhu* is active, or has been recently active. The present work shows that the majority of sites are variable for the presence of an element in the species from Maui Nui and the Big Island. The actual level of site occupancy, the number of times a site is occupied by an element in several different individuals, was not addressed. Population models predict that the majority of elements should be inserted into a unique location (Charlesworth and Langley, 1989). Deviations from this expected result could indicate that the insertions of *uhu* are not selectively neutral, and that selection is not equal between sites. Two individuals could have an element at the same site if they inherited from a common ancestor. A high percentage of insertion sites shared between two individuals could indicate that the population has gone through a recent bottleneck and expansion. This is what is seen when the number of shared

sites is compared between individuals from the same isofemale line, which is higher than the number of shared sites between individual from different lines.

The majority of the evidence suggests that the *uhu* element is being selected out of the genome and is being lost in the species on the older islands. The *uhu* element appears to be able to maintain a higher level of transposition in the species on the younger islands. It has been suggested that the changes associated with the founding of a new population allow for a higher level of transposition. One possibility is that in a population flush following a bottleneck or founding population, individuals that normally would have a lower fitness are able to reach sexual maturity and reproduce. There is evidence that the transposition of an element causes a reduction in fitness (Eanes *et al.*, 1988). In a stable population, individuals whose elements are transposing at a higher rate would be at a disadvantage and be selected out of the population. In a growing population, they would survive, and so would the higher rate of transposition. Another possibility is that inbreeding associated with founding of the population increases or decreases the copy number of the element within individuals, increasing the variance in the copy number between individuals. A mating between individuals that have a large difference in copy may create a situation similar to hybrid dysgenesis, causing an increase in transposition. A final possibility is that the increased homozygosity dampens the genomic control of transposition, allowing for an increase in transposition.

There is a discrepancy between Southern Blot analysis and *in situ* hybridization in the estimated copy number of *uhu* in *D. mimica.* Southern blot analysis suggests that *D. mimica* has a very high copy number of *uhu*, while *in*

situ hybridization shows *D. mimica* to have 60 copies of *uhu* in the euchromatin. The hybridization conditions of the *in situ* hybridizations are more stringent than the Southern blots, thus this difference may be a result of the majority of copies in *D. mimica* being to divergent to be detected in *in situs* but are detectable using Southern blots. Another possibility is that *uhu* is arranged tandemly in *D. mimica, in situ* hybridization cannot differentiate between one or several copies inserted into a location, so the copy number would be underestimated by this technique. Another possibility is that the majority of *uhu-like* sequences in *D. mimica* are in the heterochromatin and would not be detected using *in situ* hybridization because the heterochromatin does not polytenize. This would be similar to the *Tc1-like* element *Bari* which has a high copy number in the heterochromatin of *D. melanogaster*. Metaphase chromosomes from *D. mimica* and a related species from O'ahu, shows an increase in the heterochromatin in *D. mimica* (Yoon *et al.*, 1972). Thus there is an apparent correlated increase in the copy number of *uhu* in the heterochromatin and the amount of heterochromatin in *D. mimica*. The copy number and distribution of uhu in the mimica-like species from O'ahu and Kauai and from other modified mouthparts, as well as confirming the high copy number of *uhu* in *D. mimica* heterochromatin by *in situ hybridization* to metaphase chromosomes is necessary to further elucidate the relation between the gain of *uhu* and heterochromatin.

SUMMARY

The biogeographic distribution of the *uhu* element and the sequence divergence between isolates of *uhu* from different species is consistent with a ancient origin of *uhu* in the Hawaiian *Drosophila*. The biogeographic data

indicates that *uhu* has been in the Hawaiian *Drosophila* for at least 7 million years. The sequence divergence between copies of *uhu* is consistent with the idea that *uhu* was present in an ancestral species and has been vertically inherited by the extant species. The maximum divergence between two isolates suggests that *uhu* has been in the Hawaiian *Drosophila* for 20 million years. This implies that members of the fungus feeder group may have *uhu* elements, suggesting the need for a broader survey.

The amino acid sequence of the putative transposase has homology to the transposase of the Tc1 element from *Caenorhabditis elegans* (Brezinsky *et al.*, 1990). Some isolates of *uhu* have a 10 bp duplication in the open reading frame with respect to the Tc1 reading frame. The duplication results in a premature termination of the open reading frame. This duplication appears to be ancient and has been maintained in several lineages. Copies of *uhu* with and without the duplication have been found in *D. silvestris*. The ratio of synonymous/ nonsynonymous rates of change for one isolate with the duplication suggests that there has been pressure to maintain the amino acid sequence, suggesting that the truncated transposase is at least partially functional. The 10 base pair duplication is in a region that is highly conserved between *Tc1-like* elements. The amino acids in this region are negatively charge or neutrally polar and may be involved in protein-protein interactions. The duplication adds 3 amino acids that are negatively charged or polar, and may stabilize the protein interactions.

The terminal repeat shows 59 % sequence homology to a sequence that has been shown to adopt a three dimensional structure in supercoiled DNA, suggesting that the terminal repeat adopts a similar structure. There is evidence

that the structure is recognized by proteins involved in recombination (Lee and Barton, 1993) and could serve as a site of DNA breakage during the excision of the element.

The copy number of the *uhu* element is higher in species endemic to the younger islands for both the adiastola and planitibia subgroups. The *loa* element shows a similar trend in the planitibia subgroup. The variability in insertion site occupancy of *uhu* is high in the planitibia subgroups species endemic to Maui Nui and the Big Island, but very low in *D. picticornis* of Kauai. No full length *loa* elements are found in *D. picticornis*. There are potentially complete *loa* elements in the species endemic to Maui and the Big Island. This is consistent with the idea that there has been increase in transposition of these elements that is associated with the founding of a new species.

APPENDIX A

Table 1: The organization of the genome of several species of *Drosophila.* sc represents the single copy fraction, mr is the middle repetitive fraction, hr is the highly repetitive fraction. (Felger, 1988).

Species	sc	mr	hr	Reference
D. melanogaster	72%	18%	10%	Schachat and Hogness 1974
D. melanogaster	70%	12%	12%	Manning *et al.*, 1975
D. melanogaster	72%	12%	16%	Spradling and Rubin 1981
D. arizonensis	69%	18%	13%	Schulze and Lee 1986
D. subobscura	72.5%	18%	9.5%	Felger 1988
Hawaiian Drosophila				
D. crucigera	76%	15%	9%	Triantaphyllidis and Richardson 1980
D. pilimana	64%	12%	24%	Triantaphyllidis and Richardson 1980
D. engyochracea	63%	17%	20%	Triantaphyllidis and Richardson 1980
D. silvarentis	60%	18%	22%	Triantaphyllidis and Richardson 1980
D. mimica	60%	25%	15%	Triantaphyllidis and Richardson 1980
D. picticornis	61%	19%	20%	Triantaphyllidis and Richardson 1980

Table 2: The species used in this study, their endemic island, the collection site and year, and morphological group and subgroup.

Species	Stock	Collection site
Picture-wings:		
plantibia subgroup		
D. silvestris	U28T2	Kilaeua Forest, 1977
D. heteroneura	W33B3	Wailuku, Hawaii, 1983
D. heteroneura	Q71G2	Ola'a, Hawaii, 1972
D. planitibia	U84Y	Waikamoi, Maui, 1979
D. differens	U43V1	Waikou Stream, Molokai, 1977
D. picticornis	U71J1	Kokee, Kauai, 1978
adiastola subgroup		
D. adiastola	W79B3	Waikamoi, Maui, 1986
D. peniculapedis	Y18P8	Hanaulu, Maui, 1987
D. setosimentum	Y36	Ola'a, Hawaii, 1988
E. ornata	NONE	Alakai, Kauai, 1991
Modified Mouthparts		
D.mimica	NONE	Bird Park, Hawaii, 1990

Table 3: The stocks of *D. silvestris* used and their collection sites.

Stock	Collection site
U28T2	Kilaeua Forest, 1977
U26B9	Kahuku Ranch, 1977
U34B4	Kohala, 1977
W12B7	Maulua, 1980
Y46R9	Ola'a, 1989
Z1G3	Kulani, 1993

Table 4: Copy number of the *uhu* and *loa* elements in the different species based on counts from *in situ* hybridization. Number of individuals examined in parenthesis. Two numbers are presented for the *loa* element, the first is the copy number when a probe from the 3' end of the element is used, the second is the copy number when a probe from the 5' end of the element is used. + indicates the the presence of *uhu* based on Southern blot analysis to genomic DNA

Species	Island	*uhu*	*loa* (3'/5')
Picture-wings			
planitibia subgroup			
D. picticornis	Kauai	25.6 ± 1.2 (3)	14 / 0
D. differens	Molokai	58.5 ± 2.1 (2)	47 / 12
D. planitibia	Maui	117.5 ± 2.1 (2)	36 / 15
D. silvestris	Big Island	123.6 ± 8.2 (14)	62 / 23
D. heteroneura	Big Island	166.3 ± 6.3 (3)	40 / 12
adiastola subgroup			
D. ornata	Kauai	51.0 ± 2.6 (3)	0
D. adiastola	Maui	96.3 ± 6.0 (3)	0
D. peniculipedis	Maui	122.0 ± 7.1 (2)	0
D. setosimentum	Big Island	153.0 ± 4.2 (2)	0
grimshaw subgroup			
D. grimshawi	Maui	+	

Table 4: (continued)

Modified Mouthparts		
mimica subgroup		
D. mimica	Big Island	60 (1)
Antopocerus		
D. adunca	Maui	+
Fungus Feeders		
D. nigra	Maui	0

Table 5: Number of sites showing variability for the presence or absence of the *uhu* element in the polytene chromosomes of 5 planitibia sub-group species. In each case, two randomly chosedn individuals were compared.

Species	# variable sites/total sites
D. picticornis	
within U71J1 stock	2/27 (7.4%)
Total	4/29 (13.8%)
(between U71J1 and one wild-caught F_1)	
*D. differens**	
within U43V1 stock	32/74 (43.2%)
*D. planitibia**	
within U84Y stock	85/160 (53.1%)
*D. silvestris**	
within U26B9 stock	78/156 (50%)
within U28T2 stock	55/164 (33.5%)
Total	191/228 (83.8%)
(between all individuals)	
*D. heteroneura**	
within W33B9 stock	87/210 (41.4%)
Total	172/254 (67.7%)
(between W33B9 and one individual Q71G2)	

*Unpublished data from John Hunt

Table 6: The copy number of *uhu* on individual chromosomes from 14 *D. silvestris* from 6 populations. G test of heterogeneity within the chromosome between individuals is given at the bottom of each column as is a G test for the total copy number per individual.

CHROMOSOME	X	2	3	4	5	TOTAL
STOCK (Collection site)						
Kahuku						
U26B9-1	24	18	26	29	19	116
U26B9-2	34	19	21	26	18	118
						117 ± 1.1
Kilauea						
U28T2-1	29	22	29	34	22	136
U28T2-2	32	25	30	35	20	142
						139 ± 4.2
Kohala						
U34B4 -1	27	24	32	27	15	125
U34B4 -2	24	17	28	34	20	123
U34B4 -3	26	17	24	32	15	114
						120.7 ± 5.9
Ola'a						
Y46R -1	25	20	23	34	20	122
Y46R -2	24	20	32	36	19	131
Y46R -3	26	21	26	34	19	126
						126.3 ± 4.5

Table 6: (continued)

Maulua						
W12B7 -1	28	21	25	31	20	125
W12B7 -2	26	20	23	32	26	117
W12B7 -3	22	18	27	33	14	114
						118.7 ± 5.7
Kulani						
Z1G3	25	21	26	31	19	122
						123.6 ± 8.2
G_i =	5.32	3.64	7.85	3.56	3.98	G_T = 7.11
df =	13	13	13	13	13	13
p >	0.05	0.05	0.05	0.05	0.05	0.05

$\Sigma G_i = 24.37$

df = 65

$p > 0.05$

$\Sigma G_i - G_T = 17.26$

df = 52

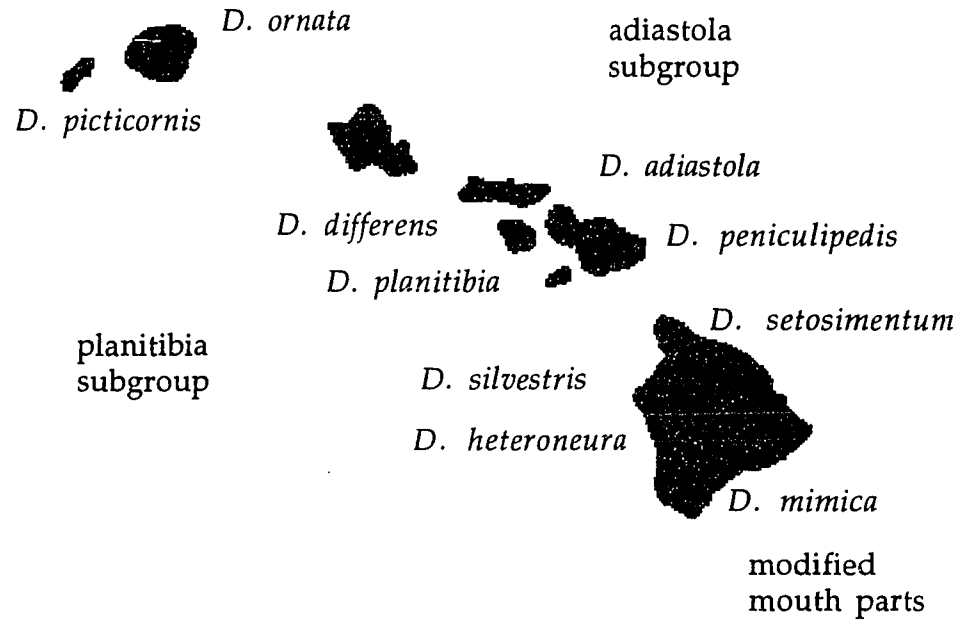
$p > 0.05$

APPENDIX B

Figure 1: A) Species from the planitibia and adiastola subgroups of the picture-wing Hawaiian *Drosophila*. *D. picticornis* and *D. ornata* are endemic to Kauai, *D. differens* is endemic to Molokai, *D. planitibia* and *D. peniculipedis* are endemic to Maui, *D. adiastola* is endemic to Maui and Lanai. All of the other species are endemic to the Big Island of Hawaii. *D. mimica,* a modified mouthpart endemic to the Big Island is also shown.

B) The relative ages of the main Hawaiian Islands.

A)


D. ornata
adiastola subgroup
D. picticornis
D. adiastola
D. differens
D. peniculipedis
D. planitibia
D. setosimentum
planitibia subgroup
D. silvestris
D. heteroneura
D. mimica
modified mouth parts

B)

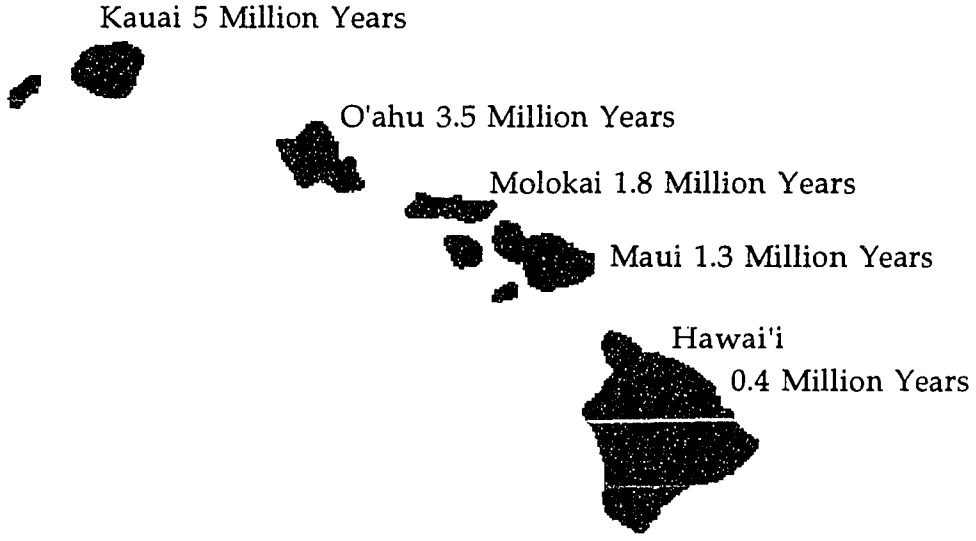

Kauai 5 Million Years
O'ahu 3.5 Million Years
Molokai 1.8 Million Years
Maui 1.3 Million Years
Hawai'i
0.4 Million Years

Figure 2: The phylogenetic relationships of the species based on chromosomal inversions. Each inversion is represented by a lower case letter as a suffix to the chromosome number (i.e. 3m is the m inversion on the third chromosome). Because the alphabet was used more then once, superscripts were added (i.e. $4d^2$). 4p indicates that the inversion is fixed, while 3m/+ indicates a polymorphic condition. Boxes with rounded corners indicate a hypothetical ancestral population. Boxes with square corners represent existing species. (Modified from Carson, 1987).

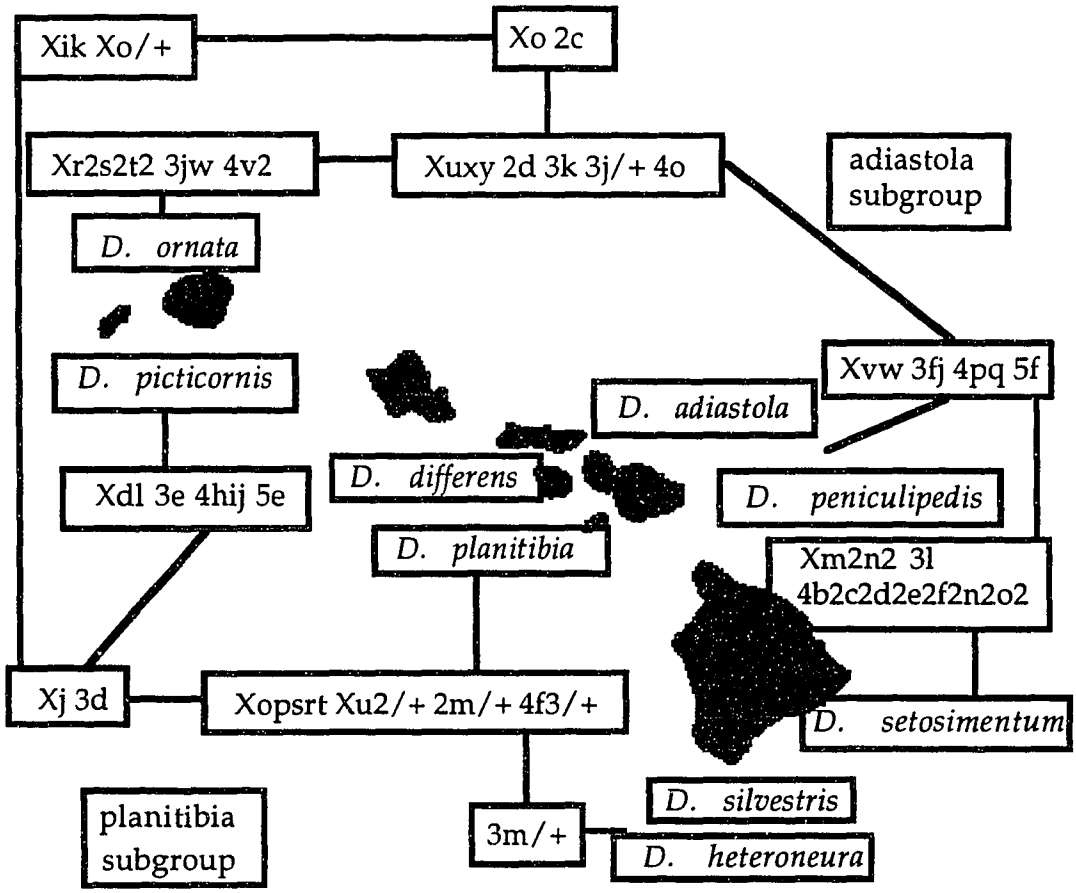

Xik Xo/+
Xo 2c
Xr2s2t2 3jw 4v2
Xuxy 2d 3k 3j/+ 4o
adiastola subgroup
D. ornata
D. picticornis
Xvw 3fj 4pq 5f
D. adiastola
D. differens
Xdl 3e 4hij 5e
D. peniculipedis
D. planitibia
Xm2n2 3l 4b2c2d2e2f2n2o2
Xj 3d
Xopsrt Xu2/+ 2m/+ 4f3/+
D. setosimentum
D. silvestris
planitibia subgroup
3m/+
D. heteroneura

Figure 3: The approximate location of the collection sites of the different stocks of *D. silvestris* on the Big Island of Hawaii.

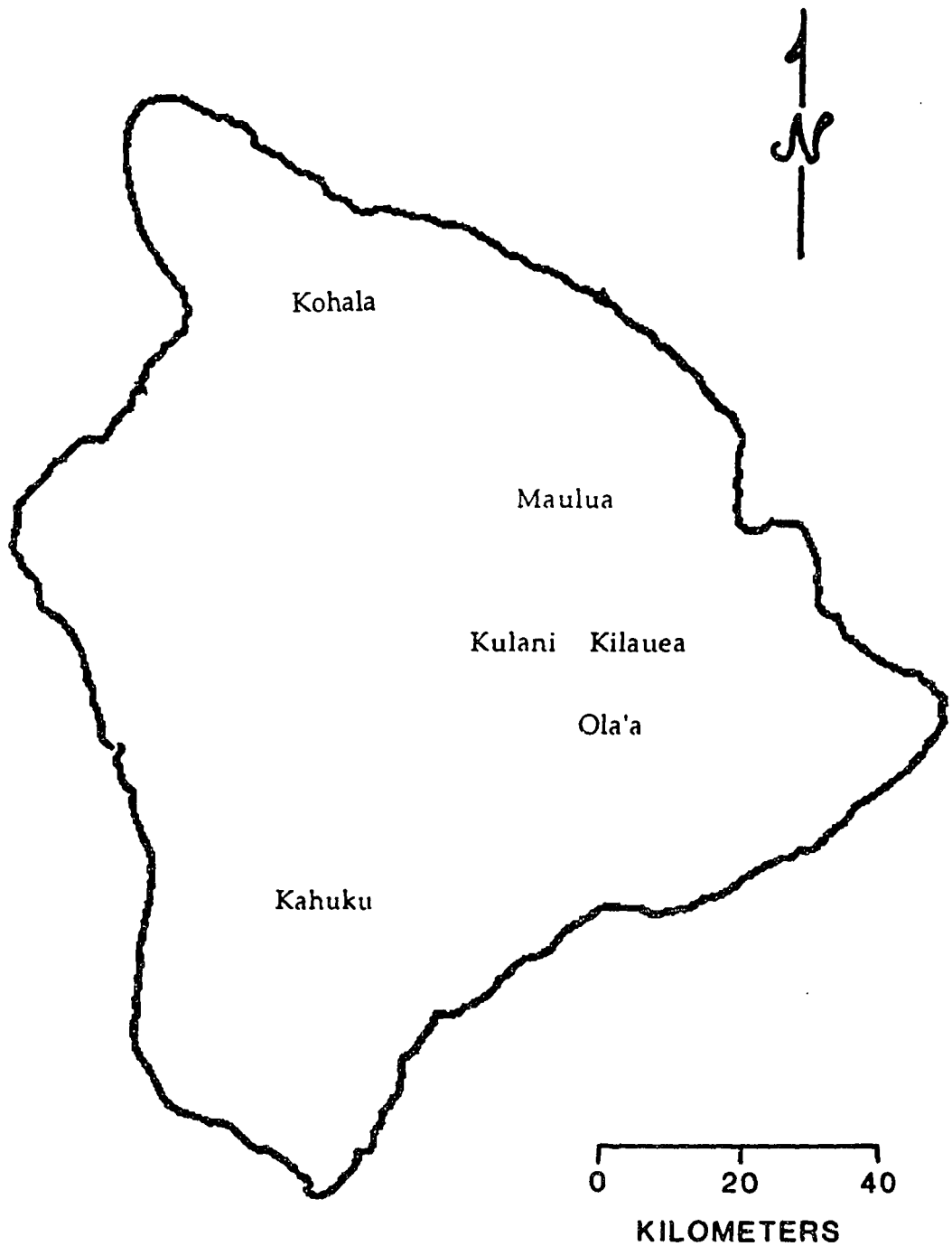

Kohala
Maulua
Kulani
Kilauea
Ola'a
Kahuku
0
20
40
KILOMETERS

Figure 4: Genetic map of the *loa element* (Felger and Hunt, 1993). a) The top indicates features of the *loa* element. ORF-1 and ORF-2 indicate the open reading frames. b) The restriction map of the *loa* element. Boxes below the map indicate the regions used as probes. c) A representation of a *"hapa"* fragment, the 3' end of a *loa* element.

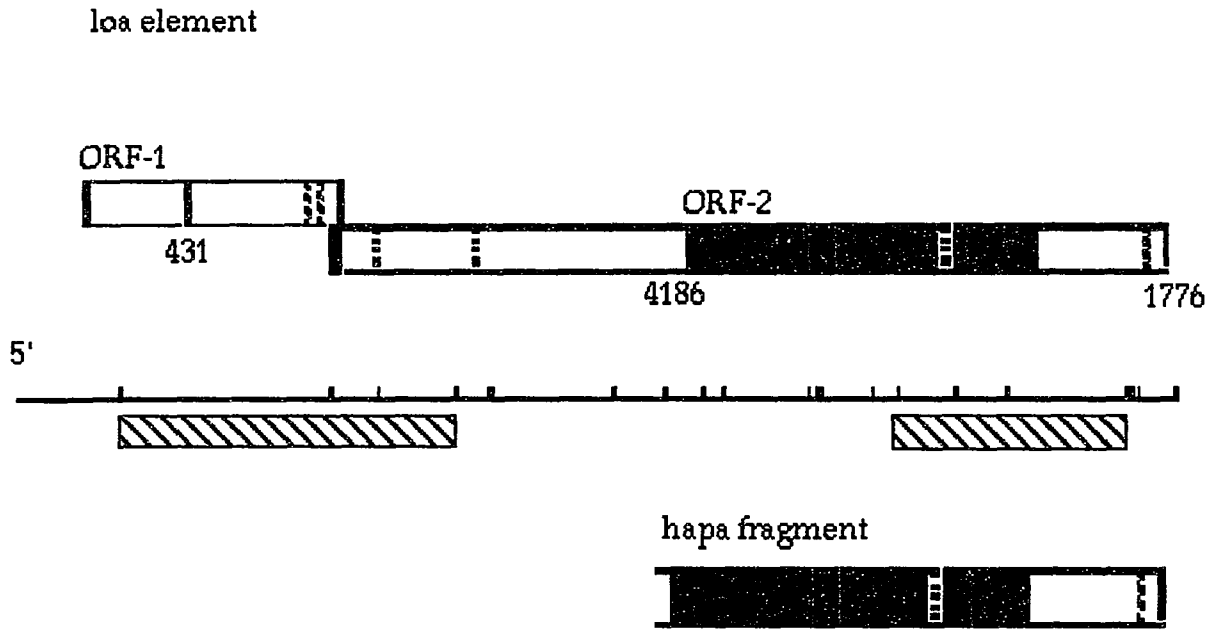
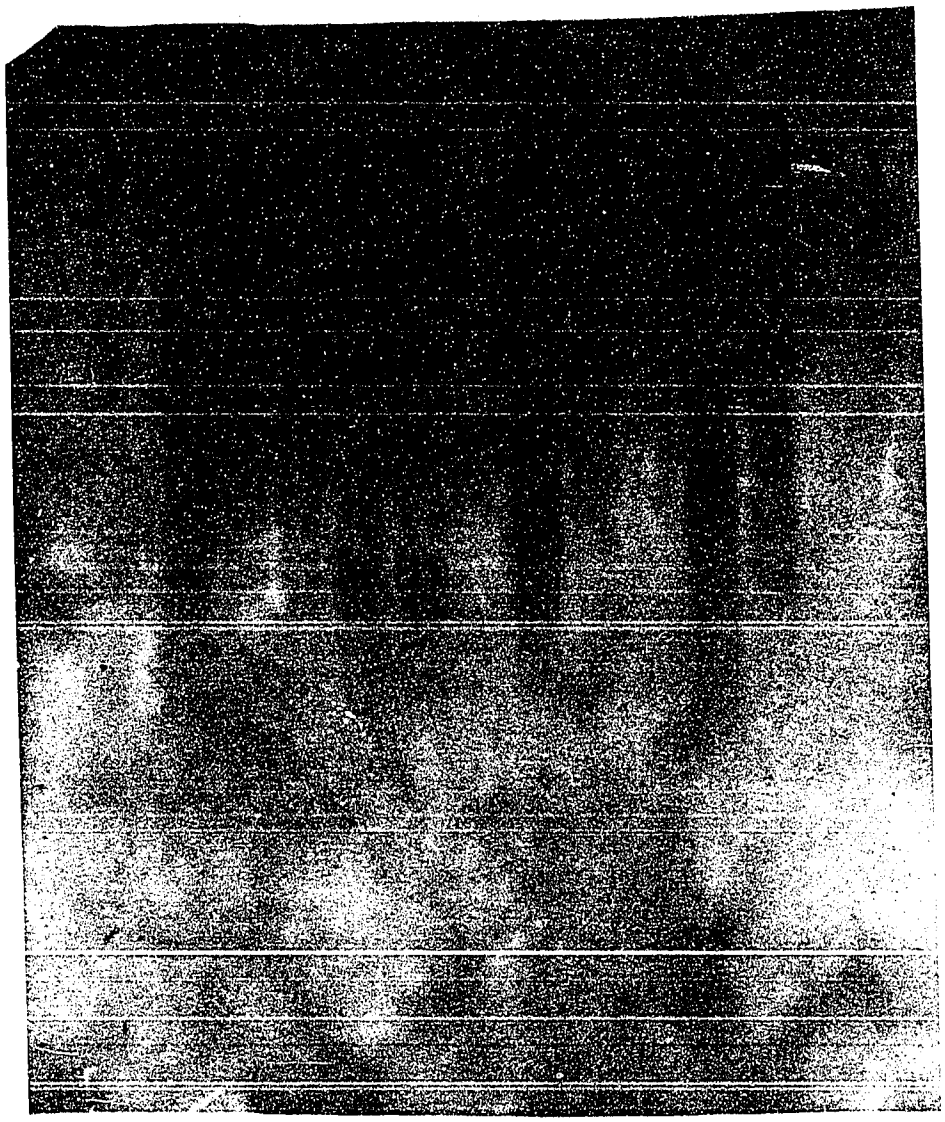

loa element
ORF-1
ORF-2
431
4186
1776
5'
hapa fragment

Figure 5: a) Hind III restriction digest of genomic DNA from several species of Hawaiian *Drosophila,* and probed with digoxygenin labeled *uhu*. b) HinD III restriction digest of genomic DNA from several species of Hawaiian *Drosophila,* two Sophophora (*D. melanogaster, D. mauritiana*) and the Medfly, *Ceritítus capitata* and probed with digoxygenin labeled *uhu*. Approximately 2μg of DNA was loaded in each lane.

A)


Kona D. silvestris
Hilo D. silvestris
D. grimshawi
D. setosimentum
D. adiastola
D. mimica
D. adunca
D. nigra
Scaptomyza albovitatta
D. picticornis

B)

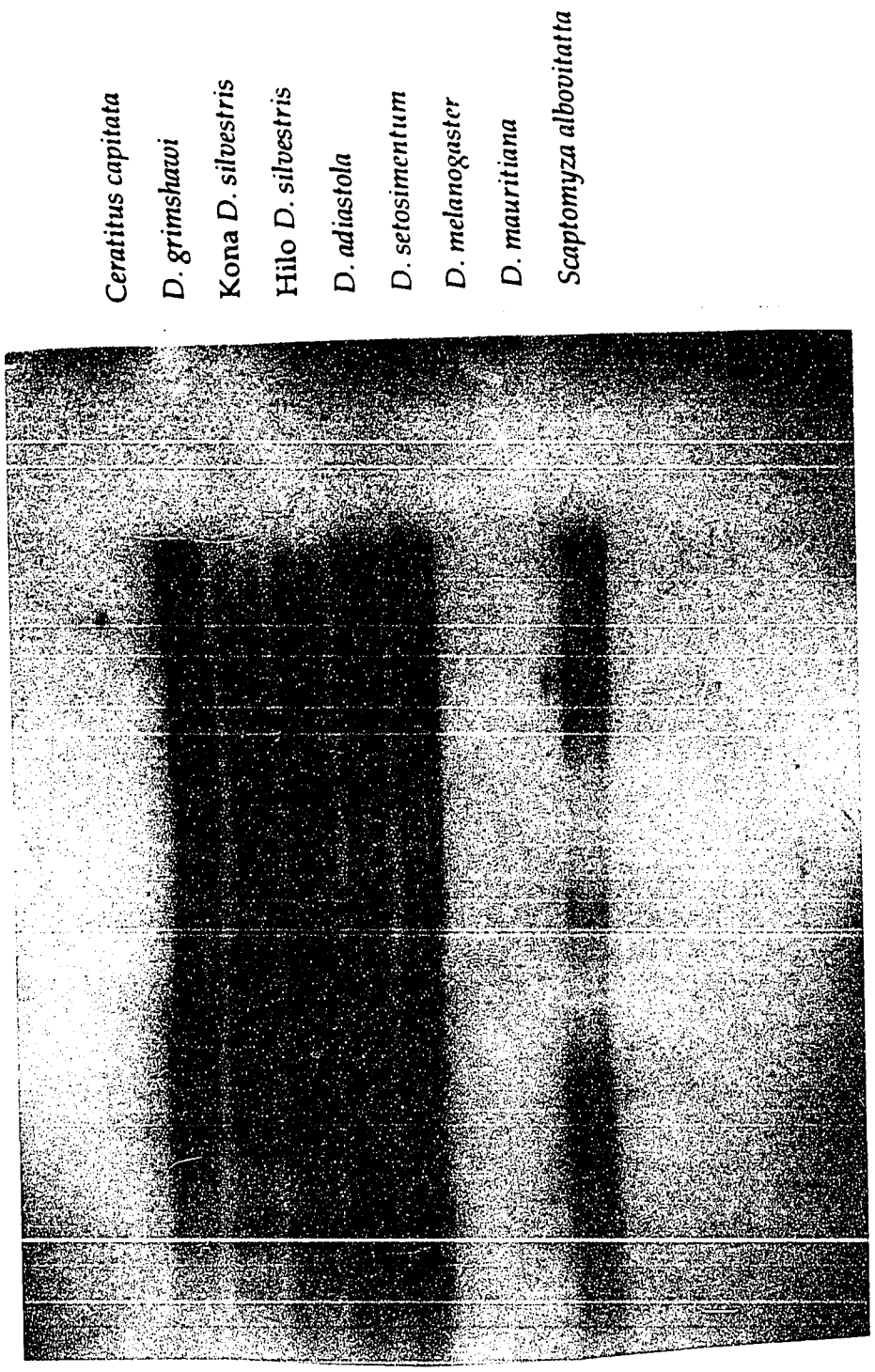
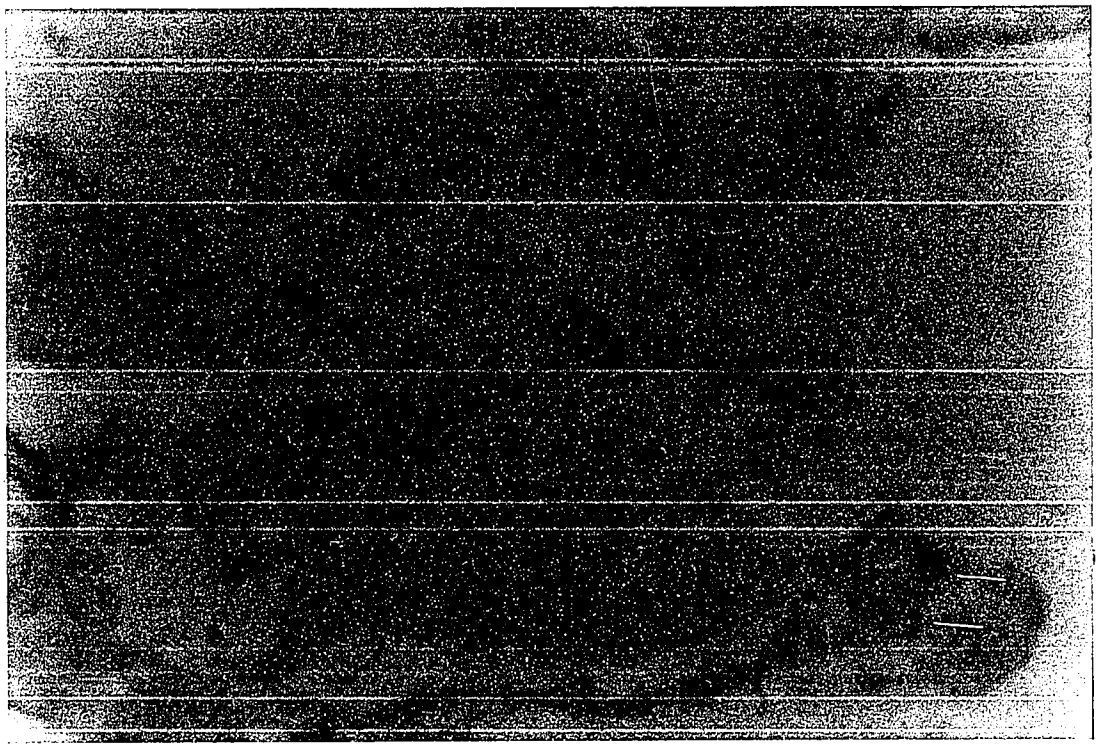

Ceratitus capitata
D. grimshawi
Kona *D. silvestris*
Hilo *D. silvestris*
D. adiastola
D. setosimentum
D. melanogaster
D. mauritiana
Scaptomyza albovitatta

Figure 6: Chromosome 4 from the same individual *D. heteroneura*. a) Probed with the 3' end of the *loa* element. b) Probed with the 5' end of the *loa* element. Arrows indicate an example of a hybridization band.

A)



B)

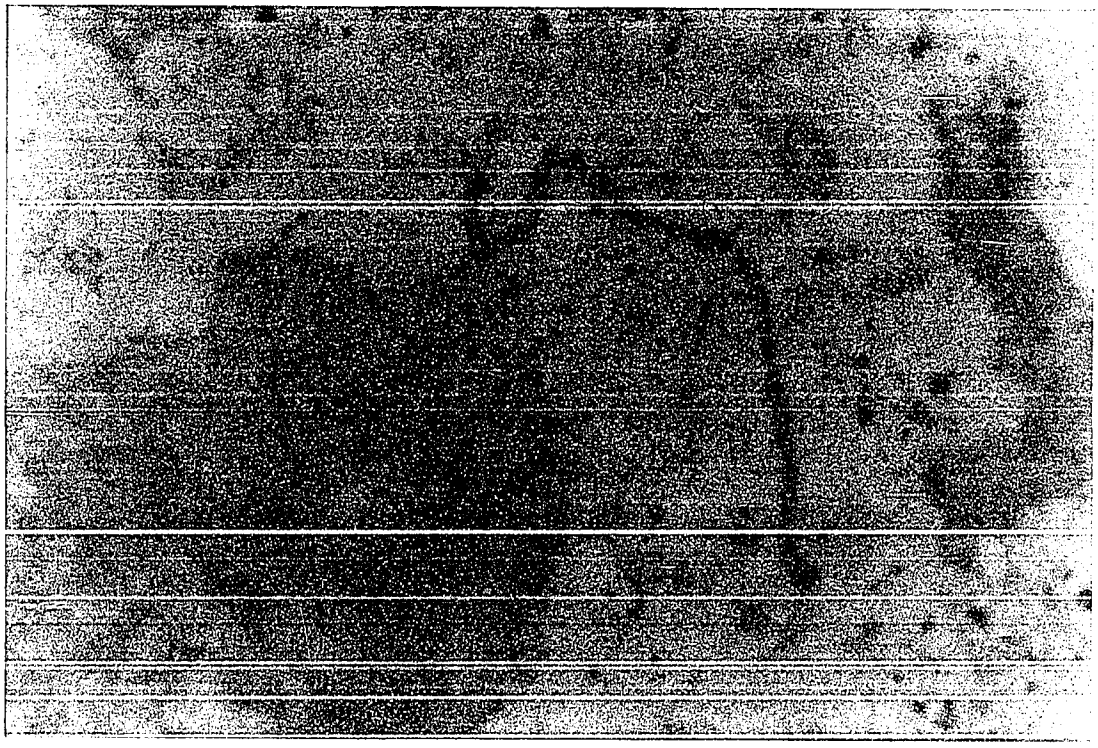
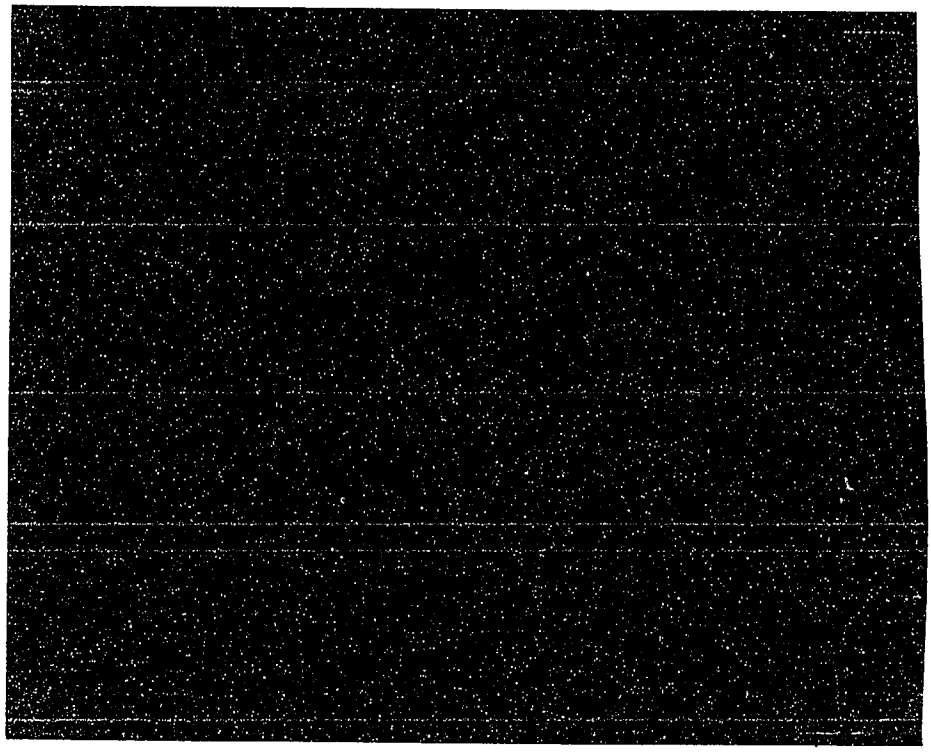


Figure 7: Chromosome 4 from two different individual *D. heteroneura* probed with the *uhu* element. Arrow indictes a site that is variable for the presence of absence of a *uhu* element.



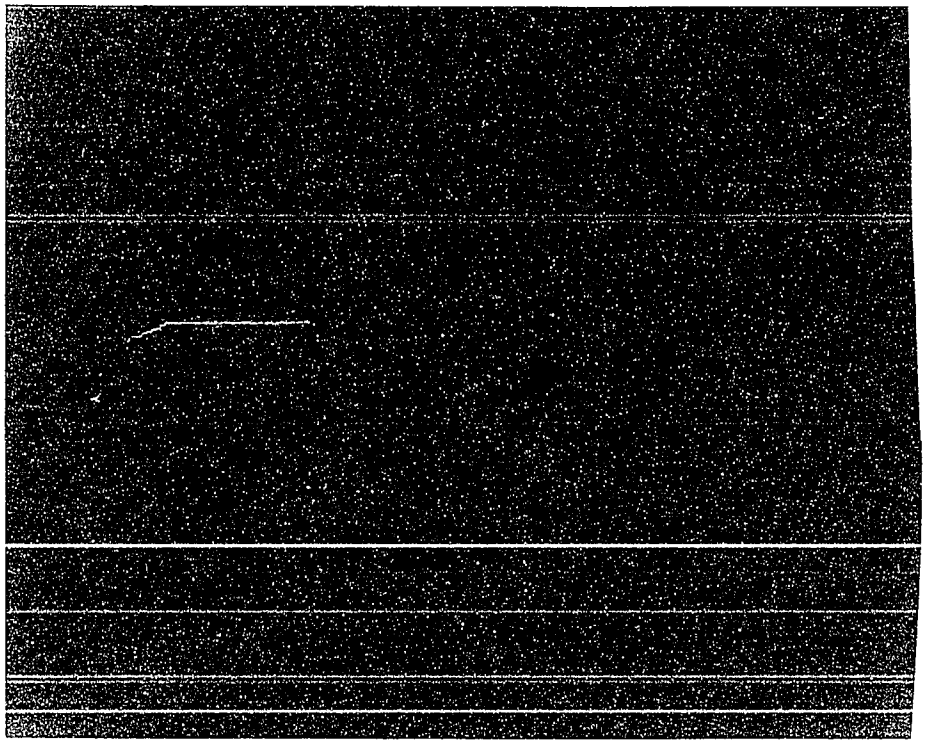


Figure 8: The sequence alignment of seven isolates of *uhu* from five species of Hawaiian *Drosophila*. The putative start and stop codons are in bold type and underlined. Het 1, Het3 and Het 4 sequences are previously published (Brezinsky *et al.*, 1993). (Het indicates isolates from *D. heteroneura*, Silv = *D. silvestris*, Adia = *D. adiastola*, Set = *D. setosimentum*, Mim = *D. mimica*).

```
Het1   TATATATA--   --AATATATA  CAGTGTCTTA  CAGCTCAACT  GGACCAGTGC   50
Het3   AGAATCTA--   --TATATATA  CAGTGTCTCA  CAGCTCAACT  GGAACAGTGC
Het4   TAGTAATA--   --TATATATA  CAGTGTCTCG  CAGCGTATTT  GGACCAGTGT
Silv3  -TATGATA--   --AATATGTA  CAGTGTCTCA  GACCTTATTT  CGACCAGTGT
Adia2  TGTATATAGG   TCTGTATGCA  CAGTGACTCA  GAGCTTATTT  GGACCAGTGC
Set1   ----------  ----------  ----------  ----------  ----------
Set5   ----------  ----------  ----------  ----------  ----------
Mim1   ---TAATC--   --TGCATATA  CAGTGACTCA  GAACGTATCT  GATGCTGTGT

Het1   CTAGCAAAAA   TTTTAATTGC  CTGCCATAAA  CTAATTATCC  ATTATTTTTC  100
Het3   CTAGCAAAAA   ATTTAATTGC  CTGCAGTAAA  CTAATTATCC  AATATTTTTT
Het4   CTAGCAAAAA   ATTTAATTGC  CTGCCATAAA  CTAATTATCC  ATTATTTTTC
Silv3  CTAGCAAAAA   TTTAAATTGC  CTGCCATAAA  CTAATTATAC  ATTGTTTATC
Adia2  CTAGCAAAAT   TTTTAATTGC  CTGCCATAAA  GTAATTATAC  ATTATTTTTC
Set1   ----------  ----------  ----------  ----------  ----------
Set5   ----------  ----------  ----------  ----------  ----------
Mim1   CAAGAAAAAT   TTTTTGTTGA  ATGCCATTAA  TTTCACAGCC  -----TTTTT

Het1   AAAAATTCCA   AAGACCGATG  GCAGGTACAT  ATATTAACCA  CCAAAATGAA  150
Het3   AAAAATTCCA   AAGACCGATG  GCAGGTACAT  ATATTAACTA  CCATAATGAA
Het4   AAAAATTCCA   AAGACCGATG  GCAGGTACAT  ATATTAACCA  CCAAAATGAA
Silv3  AAAATTTTTA   A-GACCGATG  GCAGGTACAT  ATATTAACCA  CCAAAACGAA
Adia2  AAAAATTCCA   AAGACCTATG  GCAGGTACAT  AAA-------  ----------
Set1   --AGCTCCCA   AAGA------  ----------  ----------  ----------
Set5   ----------  ----------  ----------  ----------  ----------
Mim1   CAAAATTCCA   ATGTCCGAAG  GTAGATACAT  ATATTTAGCA  ---GAATGAC

Het1   TATATGATCC   CAATAAACTG  GGGTTTCCCA  CCTGCTAGGT  CGGGTTATGT  200
Het3   TATATGATCC   CAATAAACTG  GGGTTTTCCA  CCGGCTAGGC  CGGGTTATGT
Het4   TATATGATCC   CAATAAACTG  GGGTTTCCCA  CCTGCTAGGT  CGGGTTGTGT
Silv3  TATATGACTC   CAATAAACTG  GAGTTTCCCA  CCTGCTAGGT  CGGGTTATGT
Adia2  ----------  ----------  ----------  -------GGT  CGGGTTATGT
Set1   TATATGAACC   CA-TAAA-TG  TGTTTCCCCA  CCAACTAGGT  CGCCTTATGT
Set5   ----------  ----------  ----------  ----------  ----------
Mim1   ----------  ----------  ----------  ----------  ----------

Het1   AAAAAAGTAC   CTTAATTTAT  GGTTACATAT  TATTTGGACC  AGCGGCGTTA  250
Het3   AACAAAGTAC   CTTAATTTAT  GGTTTCATAT  TATTTGGAAC  AATGGCGTTA
Het4   AAAAAAGTAC   CTTAATTTAT  GGTCACCTCT  TTTTTGGACC  AGCGGCGTTA
Silv3  AAAAAAGTAC   CTTAATTTAT  GGTCACAGCT  TATTTCGAAC  AGTTGCATTA
Adia2  AAAAAAGTAC   CTCAATTTAT  TGTCAAAGCT  TATTTGGACC  AGTAAGATTA
Set1   AAAAAAGTAC   CTCCATTTTT  GGTCAAAGCT  TATT------  ----------
Set5   ----------  -----TTTAA  GGTCAAATCT  TATTTCAAGA  AGTCACATTA
Mim1   ----------  ----------  ----------  ----------  ----------

Het1   TGGACACCTG   GGTGCCATAA  AACCCGG---  -------ATT  TTTTACGTCA  300
Het3   TGGACACCTG   GGTGCCATAA  AACCCGG---  -------ATT  TTTTACGTCA
Het4   TGGACACCTG   GGTGCCATAA  AACCCGG---  -------ATT  TTTTACGTCA
Silv3  TGGACACCTG   GGTGCCATAA  AACCCTG---  -------ATT  TTTTACGTCA
Adia2  TGGACACCTG   GGTGCCATGA  AACCCGGTGA  AACCCGTATT  TTTTAAGCCA
Set1   TGGACACCTG   GGTGCCATAA  GACCCGG---  -------ATT  TTTTACGTCA
Set5   TGGACGCCTA   GTTGCCATAA  AACCCGG---  -------ATA  TTTTACGTCA
Mim1   ----------  ----------  ----------  ----------  ----------
```

```
Het1   GGTTGATTAT  TTTCGGTATA  AATAGACCAA  TCCTTCGTAG  ----TCAGTT  350
Het3   GGTTGATTAT  TTTCGGTATA  AATAGACCAA  TCCTTCGTAG  ----TCAGTT
Het4   GGTTGATTAT  TTTCGGTATA  AATAGACCAA  TCCTTCGTAG  ----TCAGTT
Silv3  AGTTGATTAT  TTTCGGTATA  AATAGACCAA  TCTTTCGTAG  ----TCAG--
Adia2  GGTTGATTAT  TTTCGGTATA  AATTGACCAA  TC-TTCGTAG  AGTGTTAGTT
Set1   GGTTAATTAT  TTTTGGTATA  AATAAACCAA  TACTTCGTAG  ----TTAGTT
Set5   GGTTGATTAA  TTTTGGTATA  AATACACCAA  TACGTCGTAG  ----TT-GTT
Mim1   ----------  ----------  ----------  ----------  ----------

Het1   TAGTTATATC  CTGCATCTCG  GGTGCAACCA  GCCAACAAGG  CATATGGGCA  400
Het3   TAGTTATATC  CTGCATCTCG  GGTGCAACCA  GCCAACAAGG  CATATGGGCA
Het4   TAGTTATATC  CTGCATCTCG  GGTGCAACCA  GCCAACAAGG  CATATGGGCA
Silv3  --ATTACA--  --------AG  AACATAGTCA  GGCAACAAGG  ACTATGGGTA
Adia2  TAGTTATATT  CTGCATCTCG  GGTGCAACCA  GCCAGTCCTT  C-TATGGACA
Set1   TAGTTATATC  CTACATCTCG  GGTGCAACCA  GCCAACAAGG  ACTATGGGCA
Set5   TAGTTATATC  CGGCATCTCG  GGTGCAACCA  GCCAACAGGG  ACTATAGGCA
Mim1   ----------  ----------  ----------  ----------  ---------

Het1   AGCGGACTAC  CATTGAACAA  CGGAAACTGA  TCCTGGAACA  TTTCAAGATT  450
Het3   AGCGGACTAC  CATTGAACAA  CGGAATCTGA  TCCTGGAACA  TTTCAAGATT
Het4   AGCGGACTAC  CATTGAACAA  CGGAATCTCA  TCCTGGAACA  CTTCAAGATC
Silv3  AACGGACTAC  CATTGAACAA  CGGAAACTCA  TCCTGGAACA  CTTCAAAATC
Adia2  A-CGG-CT-C  TATTG--TCA  CG--ATCTCA  TCCTG-AACA  CTTCCAAAT-
Set1   AGCG-ACTAG  TATTGATCAA  CG-AATCTCA  TCCTG--ACA  CTTCCAAAT-
Set5   AACGGACGAC  TATTGATCAC  TGGAATCTCA  TCCTC--ACA  CTTCCAAATG
Mim1   ----------  ----------  ----------  ----------  ----------

Het1   GGATATTCAT  ATCGCCAAAT  AGCTAAAATG  GTAAATCTAA  GTACCACAAC  500
Het3   GGATATTCAT  ATCGCCAAAT  AGCTAAAATG  GTAAATCTAA  GTACCACAAC
Het4   GGATATTCAC  ATCGCCAAAT  AGCTAAAATG  GTAAATCTAA  GTACCACAAC
Silv3  GGATATTCAC  ATCGCCAAAT  AGCTGAAATG  GTCAATCTAA  GCAAGTCAAC
Adia2  GGATATTCAC  ATCGCCAAAT  AGCTGAAATG  GTTTATCTAA  ACTACTTAAC
Set1   GGATATTCAC  ATCGCCAAAT  AGC-----TG  GTTAATCTAA  GCAACTCAAC
Set5   GGATATTCAC  ATCGCCAAAT  AGCTGAAATG  GTCAATCTAA  GTAAGTCAAC
Mim1   ----------  ----------  ----------  ----------  ----------

Het1   TGTATTCAAC  ATCATTCGGC  GCTTCGTCGA  CGAAAATCGG  ATAGAGGACA  550
Het3   TGTATTCAAC  ATCATTCGGC  GCTTCGTCGA  CGAAAATCGG  ATAGAGGACA
Het4   TGTATTCAAC  ATCATTCGGC  GCTTCGTCGA  CGAAAATCGG  ATAGAGGACA
Silv3  TGTATACAAC  ATTATTCGGC  GCTTCGTCGA  CCATAATCAG  ATAGAGGACA
Adia2  TGTATACAAC  ATCATTCGGA  GCTTTGTGGA  CGAAAATCGG  ACAGAGGACA
Set1   TGTATACAAC  ATCATTCGGA  GCTTTGTCCA  CTAAAATCGG  ACAGAGGACA
Set5   TGTATACAAC  ATCATTCGGC  GCTTCGTCGA  CGAAAATCGG  ACAGAGGACA
Mim1   ----------  ----------  ----------  ----------  ----------

Het1   AGGGCAGAAA  GGCACCAAAC  AAGATTTTCA  CCGAACAGGA  GGAGCGGAGG  600
Het3   AGGGCAGAAA  GGCACCAAAC  AAGATTTTCA  CCGAACAGGA  GAAGCGGAGG
Het4   AGGGCAGAAA  GGCACCAAAC  AAGATTTTCA  CCGAACAGGA  GGATCGGAGG
Silv3  CGGGCGGAAT  GGCACCAAAC  A---------  -------GGA  AGAGCGGAGA
Adia2  AGGGCAGAAT  GGCACCAAAC  AAGATTTTCA  CCGAACAGGC  GGAGCGGAGG
Set1   AGG-CAGAAT  GGCACCAAAC  AAGATTTTCA  CCGAACAGGA  GGA-CGGAGG
Set5   AGAGCGGAAT  GGCACCAAAT  AAGAATTTCC  CCGAACAGAA  GGAGCGGATG
Mim1   ----------  --CCCGAAAA  AATATTTTCA  CAGAACTTGA  AGAGCAGAAG
```

```
Het1   ATCATCAGGA  AAA-TAAGGG  AAAATCCCAA  GCTATCGGCT  CCAAAACTGA  650
Het3   ATCATCAGGA  AAA-TAAGGG  AAAATCCCAA  GCTATCGGCT  CCAAAACTGA
Het4   AGCATCAGGA  AAA-TAAGGG  AAAATCCCAA  GCTATCGGCT  CCAAAACTGA
Silv3  ATAATCAGGA  AAA-TACGGG  AAAATCCTAA  GTTATCGGCT  CCAAAACTGA
Adia2  ATCATCAGGA  AAA-TAAGGT  GAAATCGTTA  GCTATCGGCT  CCAAAACTGA
Set1   ATC-------  ----------  ----------  ----------  ----------
Set5   ATCTTCAGGA  AAA-TAAGGG  GAAACCGTAA  GCTATCGGCT  CTAAAACTGA
Mim1   ATCATCAGAA  AAAATTAAGG  AAAATCTTAG  GCTA------  ---AAACTGA

Het1   CTCAACAGGT  GCAGGATGAA  ATGGGGAAAA  AGTGCAGTGT  GCAAACTGTG  700
Het3   CTCAACAGGT  GCAGGATGAA  ATGGGGAAAA  AGTGCAGTGT  GCAAACTGTG
Het4   CTCAACAGGT  GCAGGATGAA  ATGGGGAAAA  AGTGCAGTGT  GCAAACTGTG
Silv3  GTCAACAGGT  GCAGGATGAA  ATGGGGAAAA  AGTGCAGTGT  AGAAACTGTG
Adia2  C-CAACACGG  GCAGGATGAA  ATGGGGAAAA  CGTGGAGTGT  AAAAATTGTG
Set1   ----------  ----------  ----------  ----------  ----------
Set5   CCCAACAGGT  GCAGGATGAA  ATGGGGAAAA  AGTGGAGTGT  ATAAATTGCG
Mim1   CACAACCGGT  GCAGGACGAA  ATGGGGAAAT  CGTG------  -----CTGTG

Het1   CGCCGGGTTC  TGCACAACCA  TGACTTTAAT  GCCCGAGTAC  CACGGAAGAA  750
Het3   CGCCGGGTTC  TGCACAACCA  TGACTTTAAT  GCCCAAGTAC  CACGGAAGAA
Het4   CGCCGGGTTC  TGCACAACCA  TGACTTTAAT  GCCCGAGTAC  CACGGAAGAA
Silv3  CGCCAGGTTC  TGCGCAACCA  TGACTTCAAT  GCCCGAGTAC  CACGGAAGAA
Adia2  CACCTGGTTC  TGCACCACCA  TGACTTTAAT  GCCCGAGAAC  CACGGAAGAA
Set1   ----------  ----------  ----------  ----------  ----------
Set5   CGCCTGGTTC  TAATCAACCA  TGACTTTAAT  GCCCGAGTAC  CACGAATGAA
Mim1   CG--------  ---GTAACTA  T-ACCTTAAT  TCCCGAGTAC  TCCAGAAGAA

Het1   GCCATTTATA  A-GCACAAAA  AATAAAGGGA  CTAGGATGAC  GTTCGCCAAA  800
Het3   GCCATTTATA  A-GCACAAAA  AATAAAGGGA  CTAGGATGAC  GTTCGCCAAA
Het4   GCCATTTATA  A-GCACAAAA  AATAAAGGGA  CTAGGATGAC  GTTCGCCAAA
Silv3  GCCCTTTATA  A-GCGCAAAA  AATAAAGGGA  CTAGGATGAC  GTTCG-----
Adia2  GCGTTTTATA  A-GCGCATAA  AATAAATGGA  CCAGAATGAC  GTTCTTCAAA
Set1   ----------  ----------  ----------  ----------  ----------
Set5   GCGTTTTATA  A-GCGCAACA  AATAAAGGGA  CCAGGATGAC  GTTCTCCAAA
Mim1   GCCCTTCACA  AAGAATAAAA  AACTAACGGA  CTAGGTTGAC  GTTTGTCAAC

Het1   ACCCACTTGG  ACAAGGATTT  GGAGTTCTGG  AACACAATCA  TATTTGAAGA  850
Het3   ACCCACTTGG  ACAAGGATTT  GGAGTTCTGG  AACACAGTCA  TATTTGAAGA
Het4   ACCCACTTGG  ACAAGGATTT  GGAGTTCTGG  AACACAATCA  TATTTGAAGA
Silv3  -CCCACTTGG  ACAAGGATTT  GGAGTTCTGG  AACACAGTCA  TATTTGAAGA
Adia2  TCCCACTTGG  ACAAGGATTT  GAAGTTCTGG  AACACACTCG  TATTTGAAGA
Set1   ----------  ----------  ----------  ----------  ----------
Set5   ATCCACTTGG  ACAAGAATTT  AAAGTTCTGG  AACACAGTCA  TAGTTGAAGA
Mim1   ACCCACATGG  ACAAGGATTT  GA---TTT--  -----ATCTA  TTTTTGAAGA

Het1   TGAGTCCAAA  TTCATAATTT  TTGGCTCGGA  CGGACGGAAT  TATGTGCGGC  900
Het3   TGAGTCCAAA  TTCAACATTT  TTGACTCGGA  CGGACGGAAT  TATGTGTGGC
Het4   TGAGTCCAAA  TTCATCATTT  TTGGCTCGGA  CGGACGGAAT  TATGTGTGGC
Silv3  TGAGTCCAAA  TTTAACATTT  TTGGATCGGA  CGGACTGAAT  TATGGGTGGC
Adia2  TGAGTCCAAA  TTAAAAATTT  TTGGCTCGGA  GAGAGGGAAT  TGTGTGTGGC
Set1   ----------  ----------  ----------  ----------  ----------
Set5   TAAGTCCAAA  TTCAAAATTT  TTGGCTCGGA  CCGACGGCAT  TATGTGTGGC
Mim1   CTAGTCCAAA  TTCAAAATTT  GTGGCTCGAA  CGGAAGGAAG  TTTGTGTGGC
```

```
Het1   GACAGTCCAA  TACTGAGCTG  AATCCCAAAA  ACCTAAAGGC  AACAGTGAAG  950
Het3   GACAGTCCAA  TACTGAGCTG  AATCCGAAAC  ACCTAAAGGC  AACAGTGAAG
Het4   GACAGTCCAA  TACTGAGCTG  AATCCGAAAC  ACCTAAAGGC  AACAGTGAAG
Silv3  GACAATCCAA  TACTGAGCCG  AATCCCAAAG  GCTTAAAGCC  ACCGGTCAAG
Adia2  GACAGTCCAA  TAC-GAGCTG  GATCCCAAAA  ACGTAAAGGC  AACAGTCAAG
Set1   ----------  ----------  ----------  ----------  ----------
Set5   GACAGTCCAA  TACTGAGCTG  GATCCCAAAA  ACGTGAAGGC  AACAGTCAAG
Mim1   GACAGTCCAA  C---------  ----------  ----------  ----------

Het1   CACGGCGGAG  GAAGTGTCAT  GGTATGGGC-  ---ATGTATC  TCCGCAGCCA  1000
Het3   CACGGCGGAG  GAAGTGTCAT  GGTATGGGC-  ---ATGTATC  TCCGCAGCCG
Het4   CACGGCGGAG  GAAGTGTCAT  GGTATGGGC-  ---ATGTATC  TCCGCAGCCG
Silv3  CACGGAGGAG  GAAGTGTCAT  GGTATGGGCG  GGCATGTATC  TCGGCAGCCG
Adia2  CACGGCGGAG  GAAGTGACAT  GGTATGGGC-  ---ATCTATC  TCGGCAGCCA
Set1   ----------  ----------  ----------  ----------  ----------
Set5   CACGGCGGAG  GAAGTGACAT  GGTATGGGC-  ---ATCTTTC  TCGGCAGCCA
Mim1   --CGGT--AG  GAAATG----  ----------  ----------  -CGGCAGCCA

Het1   GCGTCGGAAA  TTTGGTGTGT  ATTGAAACAA  CAACGGACAG  GAATGTGGAC  1050
Het3   GCGTCGGAAA  TTTGGTGTGT  ATTGAAACAA  CAACGGACAG  GAATGTGGAC
Het4   GCGTAGGGAA  TTTGGTGTTC  ATTGAACCAA  CAATGGACAA  GAATGTATAT
Silv3  GCGTCGGAAA  TTTGGTGTTT  ATTGAAACAA  CAATCGACAA  GAATGTGTAC
Adia2  GCATCGGGAA  TTTGGTGTTC  ATTGAAACAA  CAATGGACAA  GAATGTGTTC
Set1   ----------  ----------  ----------  ----------  ----------
Set5   GAGTCGGTAA  TTTGATGTTT  ATTGGAACAA  AAATGAACAA  GAATGTGTAC
Mim1   GTGTCGGAAA  CTTGGCGATT  ATTGAGACAA  CAATGGATAG  GAATATGTAT

Het1   CTCAGTATAT  TAAAGGAAAA  TTTACTCCAA  AGTGCCGAGA  AGCTAGGAAT  1100
Het3   CTCAGAATAT  TAAAGGAAAA  TTTACTCCAA  AGTGCCGAGA  AGCTAGGAAT
Het4   CTCAATATAC  TAAAGGAAAA  TTTACTCCAA  AGTGCCGAGA  AGCTAGGAAT
Silv3  CTCAATATAT  TAAAGGAAAA  TGTACTCCAA  AGTGCCGAGA  AGCTAGGAAT
Adia2  TTCAATATAT  TAAAGGGAAA  TTTACTCCAA  AGTGCCGAGT  AGCTTA-AAT
Set1   ----------  ----------  ----------  ----------  ----------
Set5   TTCAATATAT  AAAAGAAAAA  TTTACTCCAA  AGAGCCGAGA  AGCTAG-AAT
Mim1   CTTAATATAC  TTTA---AAA  TTAATACAAA  AGCGCCGATA  TGGTAGAAAT

Het1   CCGACGTACT  TTCCGGTTCT  ACCAGGACAA  CGACCAGGAC  AACAACCAAG  1150
Het3   CCGACGTACT  TTCCGGTTCT  ACCAGGACAA  CGACCAGGAC  AACAACCAAG
Het4   CCGACGTACT  TTCCGGTTCT  ACCAGGACAA  CGACCAGGAC  AACAACCAAG
Silv3  CCTACGTACT  TTTCGGTTGT  ACCAGGAAAA  CGACC-----  -----CCAAG
Adia2  TCGATGTACT  TTCCGGTTCT  ACCACGACAA  TG--------  ----------
Set1   ----------  ----------  ----------  ----------  ----------
Set5   ACGACGTACT  TTACGGTTCT  ACCACGACAA  TGACC-----  -----CCAAG
Mim1   CCGACGTATT  TTCCGGTTAT  ACCAAGAGAA  T---------  ----------

Het1   CATAAGTCCG  GATTAGTACC  GTCCTGGCTT  ATCTGGAACT  GCCCCCACAT  1200
Het3   CATAAGTCCG  GATTAGTACA  GTCCTGGCTT  ATCTGGAACT  GCCCCCACAT
Het4   CATAAGTCCG  GATTAGTACA  GTCCTGGCTT  ATCTGGAACT  GCCCCCACAT
Silv3  CATAAGTCCG  GATTTCTACA  GTCCTGGCTT  ATCTGGAACT  GCCCCCACAT
Adia2  ----------  ----------  ----------  ----------  ----------
Set1   ----------  ----------  ----------  ----------  ----------
Set5   CATAAGTACG  GATTATTACA  GTCCTGGCTT  ATTTGGAACT  GCCCCCACAT
Mim1   ----------  ----------  ----------  ----------  ----------
```

```
Het1  GATAATTTAA  CCGGCCCAGT  CTCCAGATGT  AAATGTTATT  TAAAATTTGT  1250
Het3  GATAATTCCA  CCGGCCCAGT  CTCCAGATGT  AAATGTTATT  TAAAATTTGT
Het4  GATAATTCCA  CCGGCCCAGT  CTCCAGATGT  AAATGTTATT  TAAAATTTGT
Silv3 GATAATTCCA  CCGGCCCAGT  CTCCAGGTTT  AAATGTTATT  GAAAATTTGT
Adia2 ----------  ----------  ----------  ----------  ----------
Set1  ----------  ----------  ----------  ----------  ----------
Set5  GATAATTCCA  CCGGCCCGGT  CTCCAGATTT  AAATGTTATT  GATAATTTCT
Mim1  ----------  ----------  ----------  ----------  ----------

Het1  GGGATCTGCT  GGAAAATAAC  ATCCGGAATC  ACAG---ATC  ---------C  1300
Het3  GGGATCTGCT  GGAAAATAAC  ATCCGGAATC  ACAG---ATC  ---------C
Het4  GGGATCTGCT  GGAAAATAAC  ATCCGGAATC  ACAG---ATC  ---------C
Silv3 GGGCTCTGCT  GGAAAATAAC  ACCCGGAATC  ACAACATATC  A-GCAAACAG
Adia2 ----------  ----------  ----------  ----------  ----------
Set1  ----------  ----------  ----------  ----------  ----------
Set5  GGGCTCTGCT  GGAAAATAAC  ATCCGGAATC  ACAGCGTATC  TTACAAGCAG
Mim1  ----------  ----------  ----------  ----------  ----------

Het1  AATCTCAAAA  AATGCTTTGC  TGGATGAGTG  GAGCAAAATC  AGTCCAGAAA  1350
Het3  AATCTCAAAA  A-TGTTTTGC  TGGATGAGTG  GAGCAAAATC  AGTCCAGAAA
Het4  GATCTCAAAA  A-TGCTTTGC  TGGATGAGTG  GAGCAAAATC  AGTCCAGAAA
Silv3 GATCTCAAAA  A-TGCTCTGC  TGGATGAGTG  GAGCAAAATC  AGTCCAGAAG
Adia2 -------GTA  G-TGGTCTGC  TGGATGAGTG  GTGCAAGATC  CGT-------
Set1  ----------  ----------  ----------  ----------  ----------
Set5  GATCTCAAAA  A-TGATCTGC  TGGATGAGTG  GAACAAGATC  AGTCCAGAAA
Mim1  ----------  ----------  ----------  ----------  ----------

Het1  CTACCCGGAA  GCTGGTATCT  TCGATGAATA  ATAGGTTAAT  GGAAGATATT  1400
Het3  CTACCCGGAA  GCTGGTATCT  TCCATGAATA  ATAGGTTAAG  GGCAGTTATT
Het4  CTACCCGGAA  GCTGGTATCT  TCGATGAATA  ATAGGTTAAG  GGAAGTTATT
Silv3 CTACCCGGAA  GCTGGTATCT  TCCATGAATA  ATAGGTTAA-  GGAAGTTGTT
Adia2 -TACCTGCAG  GTCGA-----  ----------  ----------  ----------
Set1  ----------  ----------  ----------  ----------  ----------
Set5  CTACCCGGAA  GCTGGTATCT  TCCATAAACA  ATAGGTTAAG  GGAAGTCGTT
Mim1  ----------  ----------  ----------  ----------  ----------

Het1  AAGGCTAAAG  GATATCATAC  TAAGTATTAA  CATCCTTATT  TAAGTTTTTA  1450
Het3  AAGGCTAAAG  GATATCATAC  TAAGTGTTAA  CATCCTTATT  TAAGTTTTTA
Het4  AAGGCTAAAG  GATATCATAC  TAAGTGTTAA  CATCCTTATT  TAAGTTTTTA
Silv3 AAGACCAA--  GATATCATAC  TAAGTATTAA  CATCCTTATT  TAAGCTTT-A
Adia2 ----------  ----------  -------___  ----------  ----------
Set1  ----------  ----------  -------___  ----------  ----------
Set5  AAGGCTAAAG  TATATCATAC  TAAGTATAAA  CATCCTTATT  TAAGTTTTTA
Mim1  ----------  ----------  -------___  ----------  ----------

Het1  TACGTCGAAT  ATGTTTATTT  TCTAAGACTG  TCCCAAAAAA  GCTTTGACGT  1500
Het3  TACGCCAAAT  ATGTTACTTT  TTTAAGACTG  TTCGAATTAA  GCTTTGACAT
Het4  TACGTCAAAT  ATGTTAATTT  TCTAAGACTG  TCCCAAAAAA  GCTTTGACGT
Silv3 TACGGCAAAT  ATGTTA----  ----------  --TGAAATAA  GCTGTGACAT
Adia2 ----------  ----------  ----------  ----------  ----------
Set1  ----------  ----------  ----------  ----------  ----------
Set5  TAC-------  --GTTACTTT  TTTAAGACTG  CTTGAAATGA  GCTTTGACAT
Mim1  ----------  ----------  ----------  ----------  ----------
```

```
Het1  GTATTTTGGA  TATGTTT-CA  GTTTTTGACT  AATTTTAGTT  AAGTAATTAA  1550
Het3  GTATTTTGGA  TATGTTTTCA  GTTTTTGACT  AATTTTAATT  AATTAATTAA
Het4  GTATTTTGGA  TATGTTTTCA  GTTTTTGACT  AATTTTAGTT  AATTAATTAA
Silv3 ----------  ----------  ---TTTGA--  ----------  ----------
Adia2 ----------  ----------  ----------  ----------  ----------
Set1  ----------  ----------  ----------  ----------  ----------
Set5  GTATTTTGAA  TATGTTTTGC  -------ACT  A-TTTTAATT  AATTAATTAA
Mim1  ----------  ----------  ----------  ----------  ----------

Het1  TATTTTATTA  AAAACTAAAG  CTTTCTTTTC  AAACGTGATA  TAACATAAAA  1600
Het3  TATTTTAGTA  AAAACTAAAG  ATTATTTTTC  AAACATGATA  TAGCATGAAA
Het4  TATTTTATTA  AAAACTAAAG  TTTTCTTTTC  AAACGTGATA  TAACATAAAA
Silv3 ----------  ----------  ----------  ----------  -----TGAAA
Adia2 ----------  ----------  ----------  ----------  ----------
Set1  ----------  ----------  ----------  ----------  ----------
Set5  T-CTTTGATA  CAAAATAAAG  TTTTTTTCTT  AAACGTGACA  TAACATTAAA
Mim1  ----------  ----------  ----------  ----------  ----------

Het1  CATATTGGCA  TTTAAACATT  TTGAGTTTGT  TTCTTTGTTT  AAACCTTATA  1650
Het3  CAATTTGGCA  TTTAAACATT  TTGCATTTGT  TTCTTTGTTT  AAACTTTATA
Het4  CAAATTGGCA  TTTAAACATT  TTGCGTTTGT  TTCTTTGTTT  AAGCCTTATA
Silv3 CAAATTGGCA  CTTAAACATT  TTGCGTTTGT  TTCCTTGCTT  AGACTTTATA
Adia2 ----------  ----------  ----------  ----------  ----------
Set1  ----------  ----------  ----------  ----------  ----------
Set5  TAAATTGGCA  GTTAAACATC  TTGAGTTTGT  TTCCTTGTTG  AAACTTTACA
Mim1  ----------  ----------  ----------  ----------  ----------

Het1  GCACTTTAAA  TTTTTTGCTA  GAGACTGTTC  CAAATCAGCT  GGAAGACACT  1700
Het3  GCACTTTAAA  ATATTTGCTA  GGCGCTGTTC  CAGTTGAGCT  GTAAGACACT
Het4  GCACTTTAAA  TTTTTTGCTA  GACACTGTTC  CCAATACGCT  GCGAGACACT
Silv3 GCAATTTAAA  TTTTTTGCTA  GACACTGGTC  GAAATAAGGT  CTGAGACACT
Adia2 ----------  ----------  ----------  ----------  ----------
Set1  ----------  ----------  ----------  ----------  ----------
Set5  GCACTTTGAA  TTTTTTGCTT  GACACTGTTT  GAAATAAGCT  ------TATT
Mim1  ----------  ----------  ----------  ----------  ----------

Het1  GTATAA-CA-  AATTATTTAT  AT-   1723
Het3  GTATAT--AT  ATATATATAT  AT-
Het4  GTATAT-AAC  GAGTAGTTAC  A--
Silv3 GTATATGCAT  TTCTTTTTAT  GA-
Adia2 ----------  ----------  ---
Set1  ----------  ----------  ---
Set5  G-ATATGAGC  CATTGCTTGG  ATT
Mim1  ----------  ----------  ---
```

Figure 9: A 10 base pair imperfect duplication in isolates of *uhu* at base 1136 in shown comparison to the Tc1 element from *Caenorhabditis elegans*. The isolates of *uhu* shown are those in Figure 2 and three other isolates, one each from *D. silvestris* (Silv13) *D. planitibia* (Plan11) and *D. peniculipedis* (Pen4) where sequence information for this region is also available. The duplication results in an imperfect three amino acid duplication and a premature termination of the reading frame. The sequences of Tc1 from *C. elegans*, *Hb1* and *Bari1* from *D. melanogaster* and *Minos* from *D. hydei* are presented to show the extent of conservation of this region between the elements.

Nucleotide Sequence

Het 1	TAC CAG GAC AAC GAC CAG GAC AAC AAC CAA GCA TAA GTC
Het 3	TAC CAG GAC AAC GAC CAG GAC AAC AAC CAA GCA TAA GTC
Het 4	TAC CAG GAC AAC GAC CAG GAC AAC AAC CAA GCA TAA GTC
Silv 13	TAC CAG GAC AAC GAC CAG GAC AAC AAC CAA GCA TAA GTC
Plan 11	TAC CAG GAC AAC GAC TAG GAC AAC AAC CAA GCA TAA GTC
Silv 3	TAC CAG GAA AAC GAC C CC AAG CAT AAG TC
Set 5	TAC CAC GAC AAT GAC C CC AAG CAT AAG TA
Pen 4	TAC CAC GAC AAT GAC C CC CAG CAT AAG TC
Tc1	CAG CAG GAT AAC GAT C TC AAG CAT ACT TC
Hb1	CAA GAG GAT AAT GAT C AA AAA CGC AGA TG
Bari 1	CAG CAG GAC AAT GCT C CA TGC CAT AAG GG
Minos	CAG CAG GAC GGA GCA T CA TCG CAC ACA GC

Amino Acid Sequence

Het 1	Tyr-Gln-Asp-Asn-Asp-Gln-Asp-Asn-Asn-Gln-Ala-Stop
Het 3	Tyr-Gln-Asp-Asn-Asp-Gln-Asp-Asn-Asn-Gln-Ala-Stop
Het 4	Tyr-Gln-Asp-Asn-Asp-Gln-Asp-Asn-Asn-Gln-Ala-Stop
Silv 13	Tyr-Gln-Asp-Asn-Asp-Gln-Asp-Asn-Asn-Gln-Ala-Stop
Plan 11	Tyr-Gln-Asp-Asn-Asp-Stop
Silv 3	Tyr-Gln-Asp-Asn-Asp- Pro-Lys-His-Lys-Ser
Set 5	Tyr-Gln-Asp-Asn-Asp- Pro-Lys-His-Lys-Tyr
Pen 4	Tyr-Gln-Asp-Asn-Asp- Pro-Gln-His-Lys-Ser
Tc1	Gln-Gln-Asp-Asn-Asp- Leu-Lys-His-Thr-Ser
Hb1	Gln-Glu-Asp-Asn-Asp- Gln-Lys-Arg-Arg-Cys
Bari 1	Gln-Gln-Asp-Asn-Ala- Pro-Cys-His-Lys-Gly
Minos	Gln-Gln-Asp-Gly-Ala- Ser-Ser-His-Thr-Ala

Figure 10: Graphic representation of the different isolates of *uhu*. A consensus of all the *uhu* elements is presented at the top. The terminal repeats are indicated by T.R. The open reading frame is represented by the striped box labeled ORF. The diamond patterned box indicates the 400 bp region that was used for phylogenetic analysis of 24 isolates of *uhu*. *Het* is a consensus of the three *D. heteroneura* isolates. The new isolates are presented with deviations from the consensus indicated.

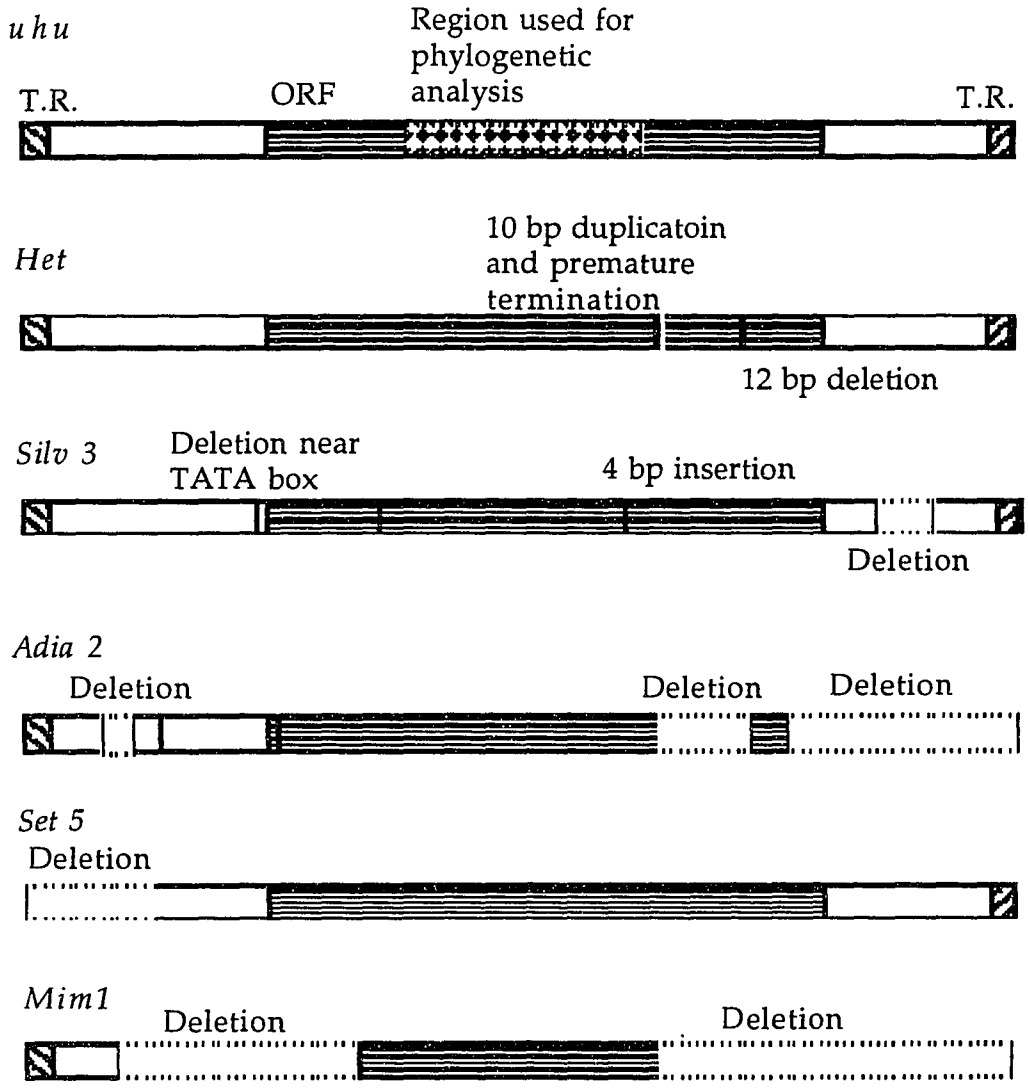

u h u
Region used for phylogenetic analysis
T.R.
ORF
T.R.
Het
10 bp duplicatoin and premature termination
12 bp deletion
Silv 3
Deletion near TATA box
4 bp insertion
Deletion
Adia 2
Deletion
Deletion
Deletion
Set 5
Deletion
Mim1
Deletion
Deletion

Figure 11: The synonymous/ nonsynonymous distance matrix for the seven isolates of *uhu*. The distances were obtained using Kimura's two parameter method which weights for transversions versus transitions. The synonymous rate of change is above the diagonal, the nonsynonymous is below the diagonal. Synonymous rates of change that are twice the nonsynonymous are underlined, those that are 2.5 times the nonsynonymous are shaded.

	HET1	HET3	HET4	SILV3	ADIA2	SET5	MIM1	
HET1	0	0.01463	0.05499	0.17683	0.16354	0.16677	0.44218	SYNONYMOUS
HET3	0.02301	0	0.04972	0.16855	0.17285	0.16981	0.4454	
HET4	0.02559	0.02304	0	0.19302	0.18088	0.1977	0.50582	
SILV3	0.07134	0.07204	0.06048	0	0.20966	0.21417	0.64233	
ADIA2	0.13243	0.14245	0.12792	0.14571	0	0.13497	0.51385	
SET5	0.11956	0.1189	0.1077	0.11283	0.13362	0	0.46857	
MIM1	0.18319	0.19863	0.18097	0.19142	0.24257	0.26536	0	
	NONSYNONYMOUS							

Figure 12: Comparison of the genetic distance of the 5' non-coding region of the uhu element with the synonymous and nonsynonymous distances of the open reading frame. Shading indicates a similar distance between the 5' region and the synonymous or nonsynonymous distance for the open reading frame.

	HET1	HET3	HET4	SILV3	ADIA2	SET5	MIM1	
HET1	0	0.01463	0.05499	0.17683	0.16354	0.16677	0.44218	Synonomous
HET3	0.062	0	0.04972	0.16855	0.17285	0.16981	0.4454	
HET4	0.0473	0.0923	0	0.19302	0.18088	0.1977	0.50582	
SILV3	0.1507	0.1957	0.1473	0	0.20966	0.21417	0.64233	
ADIA2	0.1369	0.1793	0.1497	0.2134	0	0.13497	0.51385	
SET5	0.1879	0.196	0.1888	0.2698	0.2395	0	0.46857	
MIM1	0.4724	0.5097	0.392	0.5264	0.4445	\\\\\\	0	

5' non-coding

	HET1	HET3	HET4	SILV3	ADIA2	SET5	MIM1	
HET1	0	0.02301	0.02559	0.07134	0.13243	0.11956	0.18319	Nonsynonomous
HET3	0.062	0	0.02304	0.07204	0.14245	0.1189	0.19863	
HET4	0.0473	0.0923	0	0.06048	0.12792	0.1077	0.18097	
SILV3	0.1507	0.1957	0.1473	0	0.14571	0.11283	0.19142	
ADIA2	0.1369	0.1793	0.1497	0.2134	0	0.13362	0.24257	
SET5	0.1879	0.196	0.1888	0.2698	0.2395	0	0.26536	
MIM1	0.4724	0.5097	0.392	0.5264	0.4445	\\\\\\	0	

5' non-coding

Figure 13: Terminal repeats of the isolates of *uhu* and other Tc1-like elements. Where available, both terminal repeats are shown for the *uhu* elements and Tc1. Only the 5' terminal repeat is shown for the other members of this family. The CAGTG or CAGTA motif is underlined, and the A-T rich region is stipple-underlined.

Het1 (6 Mismatches)

aataTATACAGTGTCTTACAGCTCAACTGGACCAGTGCCTAGCAAAAATTTTAA 5'

ttgtTATACAGTGTCTTCCAGCTCATTTGGACCAGTCTCTAGCAAAATTTTTAA 3'

Het3 (4 Mismatches)

tataTATACAGTGTCTCACAGCTCAACTGGACCAGTGCCTAGCAAAAAATTTAA 5'

tataTATACAGTGTCTTACAGCTCAACTGGACCAGCGCCTAGCAAATATTTTAA 3'

Het4 (4 Mismatches)

tataTATACAGTGTCTCGCAGCGTATTTGGACCAGTGCCTAGCAAAAATTTTAA 5'

gttaTATACAGTGTCTCGCAGCTCATTGGGAACAGTGCCTAGCAAAAATTTTAA 3'

Silv3 (1 Mismatch)

aataTGTACAGTGTCTCAGACCTTATTTCGACCAGTGTCTAGCAAAAAATTTAA 5'

atcaTATACAGTGTCTCAGACCTTATTTCGACCAGTGTCTAGCAAAAAATTTAA 3'

Adia2

tgtaTGCACAGTGACTCAGAGCTTATTTGGACCAGTGCCTAGCAAAATTTTTAA 5'

Set5

ctcaTAT-CAATA------AGCTTATTTCAAACAGTGTCAAGCAAAAAATTCAA 3'

Mim1

tgcaTATACAGTGACTCAGAACGTATCTGATGCTGTGTCAAGAAAAATTTTTTG 5'

Tc1 from *C. elegans* (Harris *et al.*, 1990)

CAGTGCTGGCCAAAAAGATATCCACTTTTGGTTTTTTGTGTGTAACTTTTTTCT 5'

||||||||||||||||||||||||||||||||||||||||||||||||||||||

CAGTGCTGGCCAAAAAGATATCCACTTTTGGTTTTTTGTGTGTAACTTTTTTCT 3'

Tcb1 from *C. briggsae* (Harris *et al.*, 1990)

CAGTACTGGCCATAAAGAATGCGACAACTTGTTTTTTGACGATAACTTTTTGAAAACTCAA

CTTTTCAACTCGAATTTTT

Bari1 from *D. melanogaster* (Ciazzi *et al.*, 1993)

CAGTCATGGTCAAAATTATTTTCACA

Minos from *D. hydei* (Franz and Savakis, 1991)

ACGAGCCCCAACCACTATTAATTCGAACAGCATGTTTTTTTTGCAGTGCGCAATGTTTAAC

ACACTATATTATCAATACTACTAAAGATAACACATACCAATGCATTTCGTCTCAAAGAGAA

TTTTATTCTCTTCACGACGAAAAAAAAAGTTTTGCTCTATTTCCAACAACAACAAAAATAT

GAGTAATTTATTCAAACGGTTTGCTTAAGAGATAAGAAAAAAGTGACCACTATTAATTCGA

ACGCGGCGTAA

Hb1 from *D. melanogaster* (Brierley and Potter, 1985)

AAATACAGCTGTGTTCAGAAAAATAGCAGTGC

Figure 14: The 5' end of the terminal repeats of other short inverted repeat transposable elements and the 5' splice sight of the small t-antigen intron from Simian Virus 40. The CAGTG or CAGTA motif is underlined.

Other Transposable elements with CAGTG or CAGTA motives in their terminal repeats.

Hobo from *D. melanogaster* (Strek *et al.*, 1986)

CAGAGAACTGCA

Pogo from *D. melanogaster* (Tudor *et al*, 1992)

CAGTATAATTCGCTTAGCTGCA

Tirant from *D. melanogaster* (Garrell and Modolell, 1990)

AGTTAAGTCTGTGATCGAGGGTGGAGCCTTTTGTGAAGATATGATGTGTAAATGTA

TGAAGGTTTTTAAGTTTCAGCAGATAGTTGTAGTGTGAAGATTTCCAATAAAGAAT

AC

Mariner from *D. melaongaster* (Medora *et al.*, 1991)

CCAGGTGTACAAGTAGGGAATGTCGGTT

Tam4 from *Antirrhinim majus* (Snapdragons) (Luo *et al.*, 1991)

CACTACAACAAAAA

Tgm1 from *Glysine max* (Soybean) (Rhodes and Vodkin, 1985)

CACTATTAGAAAATATGTTTTTTACATCGGTTATTTATG

Ac from *Zea mays* (Kunze and Starlinger, 1989)

CAGGGATGAAA

5' end of the small t-antigen intron of Simian Virus 40 (Lee and Barton, 1993)

TGTCTACAGTAAGTGAA

Bold face G is cleavage site of the intron.

Figure 15: Polymerase Chain Reaction amplification of *uhu* from plasmids using primers homologous to the terminal repeats. Isolates from *D. planitibia* (P1D) *D. heteroneura* (H3K) and *D. ornata* (O6E3) resulted in fragments near the expect 1.7 kb sized of *uhu*. An isolate from *D. setosimentum* (S5B) gave a 400 bp fragment. Sequence analysis varified that *uhu* is present in plasmids P1D, H3K and S5B.

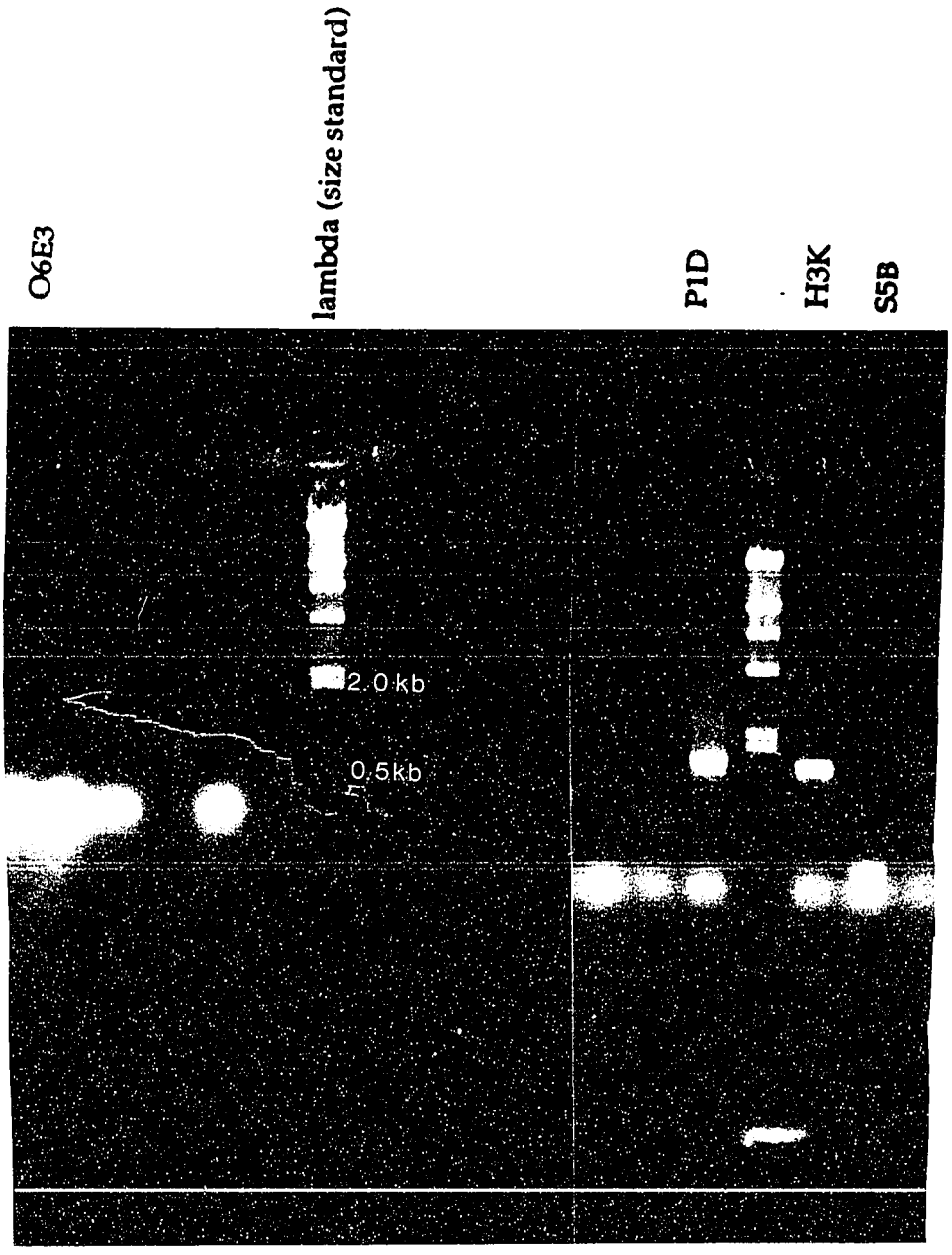
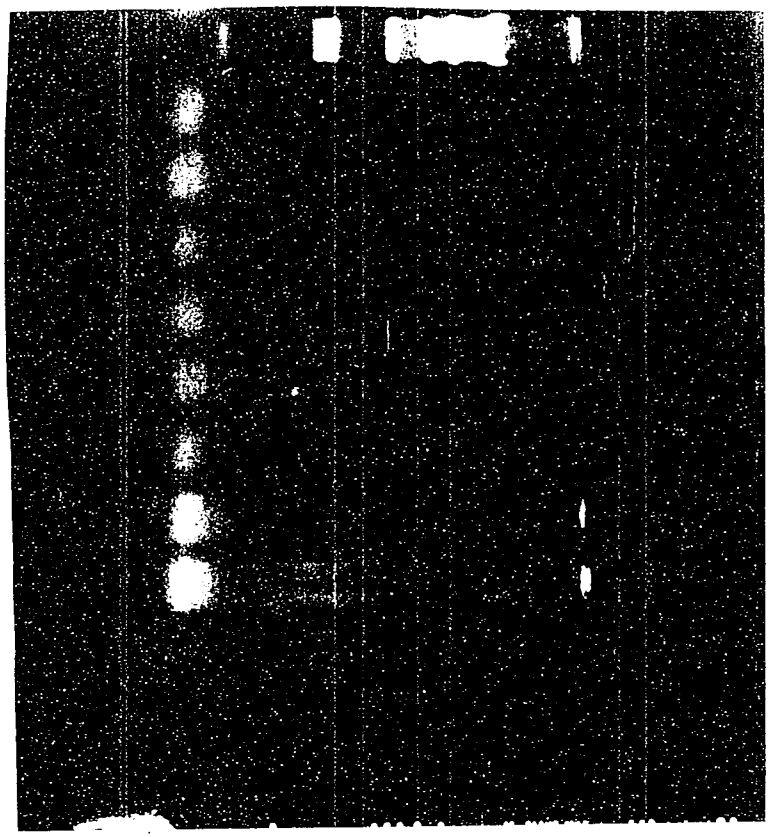

O6E3
lambda (size standard)
P1D
H3K
S5B
2.0kb
0.5kb

Figure 16: Polymerase Chain Reaction amplification of *uhu* from genomic DNA using primers homologous to the terminal repeats. Only DNA from *D. silvestris* gave the expected 1.7 kb fragment. Smaller bands are seen for both *D. silvestris* and *D. setosimentum.* Amount of DNA used as template in the reaction is in parenthesis.



D. setosimentum (0.01 µg)

D. setosimentum (0.1 µg)

D. silvestris (0.001 µg)

D. silvestris (0.01 µg)

D. silvestris (0.1 µg)

D. picticornis (0.02 µg)

D. picticornis (0.2 µg)

D. picticornis (2 µg)

Figure 17: The 402 bp region of the open reading frame used for phylogenetic analysis of *uhu*. This corresponds to the region between bases 676 and 1084 in Figure 2. Sequence Het1, Het3, Het4, Silv2, Silv3, Plan2, Plan3, Plan4, Diff1, Diff2, Pict1 and Pict4 are from Brezinsky *et al.*, (1993). ? indicates sequence that is was not possible to obtain, X indicates unresolved compressions and "-" indicate deletions. Deletions were inserted to maximize alignment.

```
HET1        AAAAAGTGCA  GTGTGCAAAC  TGTG-CGCCG  GGTTCTGCA-  CAACCATGAC
HET3        AAAAAGTGCA  GTGTGCAAAC  TGTG-CGCCG  GGTTCTGCA-  CAACCATGAC
HET4        AAAAAGTGCA  GTGTGCAAAC  TGTG-CGCCG  GGTTCTGCA-  CAACCATGAC
SILV2       ?????GTGCA  GTGTAGAAAC  TGTG-CGCCG  GGTTCTGCA-  CAACCATGGC
SILV3       AAAAAGTGCA  GTGTAGAAAC  TGTG-CGCCA  GGTTCTGCG-  CAACCATGAC
SILV13      ??????????  ??????????  ??????GCCG  GGTTCTGCA-  CAACCATGAC
PLAN2       AAAAAGTGGA  GTGTGGAAAC  TGTG-TGCCG  GGTTCTGCG-  CAACTATGAC
PLAN3       AAAAAGTGCA  GTGTGGAAAC  TGTG-TGCCG  GGTTCTGCG-  CAACTATGAC
PLAN4       AAAAAGTGCA  GTGTGGAAAC  TGTG-TGCCG  GGTTCTGCG-  CAACTATGAC
PLAN11      ??????????  ??????????  ?????CGCCG  GGTTCTGCC-  GAACCATTAC
DIFF1       AAAAAGTGCA  GTGTGGAAAC  TGTG-TGCCG  GGTTCTGCG-  CAACCATGAC
DIFF2       AAAAAGTGCA  GTGTGGAAAC  TGTG-TGCCG  GGTTCTGCG-  CAACTATGAC
PICT1       ??????????  ??GTGGAAAC  TGTG-CGCCG  GGTTCTGCG-  CAACCATGA-
PICT4       AAAAAGTGCA  GTGTGGAAAC  TGTG-CGTCG  GGTTCTGCG-  CGACCATGAC
ADIA1       ??????????  ??????????  ??TG-CACCT  GGTTCTGCA-  CAACCATGAC
ADIA2       AAAACGTGGA  GTGTAAAAAT  TGTG-CACCT  GGTTCTGCA-  CCACCATGAC
ADIA5       ??????????  ??????????  ??????????  ??????????  ??????????
PEN1        ??????????  ??????????  ??????????  ??????????  ??????????
PEN4        ??????????  ??????????  ?????????A  GGTTCTC---  ---CAATGAC
SET3        ??????????  ??????????  ??????????  ??????????  ??????GGAC
SET5        AAAAAGTGGA  GTGTATAAAT  TGCG-CGCCT  GGTTCTAAT-  CAACCATGAC
ORN1        ??????????  ??????????  ??????????  ??????????  ????????GC
MIM1        AACCGGTGCA  GGACG-AAAT  GGGGAAATCG  TGCTGTGCGG  TAACTAT-AC
MIM2        AAAAAGTGCA  GTGTGGAAAC  TGTGACGCCG  GGTTCTXCAG  CAACCATTAC

HET1        TTTAATGCCC  GAGTACCACG  GAAGAAGCCA  TTTATAA-GC  ACAAAAAATA
HET3        TTTAATGCCC  AAGTACCACG  GAAGAAGCCA  TTTATAA-GC  ACAAAAAATA
HET4        TTTAATGCCC  GAGTACCACG  GAAGAAGCCA  TTTATAA-GC  ACAAAAAATA
SILV2       TTTAATGCCC  GAGTACCACG  GAAGAAGCCA  TTTATAA-GC  GCAAAAAATA
SILV3       TTCAATGCCC  GAGTACCACG  GAAGAAGCCC  TTTATAA-GC  GCAAAAAATA
SILV13      TTTAATGCCC  GAGTACCACG  GAAGAAGCCA  TTTATAA-GC  ACAAAAAATA
PLAN2       TTTAATGCCC  GAGTACCACG  GAAGAAGCCC  TTTATAA-GC  GCAAGAAATG
PLAN3       TTTAATGCCC  GAGTACCACG  GAAGAAGCC-  ----------  --AAGAAATG
PLAN4       TTTAATGCCC  GAGTACCACG  GAAGAAGCCC  TTTATAA-GC  GCAAGAAATG
PLAN11      TTCAGTGCCC  GAGTACCACG  ----------  ---ATAA-GC  GCAAAAAATA
DIFF1       TTTAATGCCC  GAGTAGCACG  GAAGAAGCCC  TTTATAA-GC  GCAAGAAATG
DIFF2       TTTAATGCCC  GAGTACCACG  GAAGAAGCCC  TTTATAA-GC  GCAAGAAATG
PICT1       ----------  ----------  -AAGAAGCCC  TTTATTA-GC  GCAAAAAATA
PICT4       ----------  ----------  ----------  ----------  ----------
ADIA1       TTTAATACCC  GAGTACCACG  -AGTAAGCGT  TTTTTAA-GC  GCAAAAAATC
ADIA2       TTTAATGCCC  GAGAACCACG  GAAGAAGCGT  TTTATAA-GC  GCATAAAATA
ADIA5       TTAAATTCTC  GGGTACGGCG  GAAGAAGCCC  TTTTTAA-GT  GCAAAAAATA
PEN1        ???????CCC  GAGTACCACG  -AAGAAGCCC  TT-ATAA-GC  GCAAAAAACA
PEN4        TTTAATGCCC  GAGTACCACG  -AAGAAGCGT  TTTATAA-GC  GCTCAAAATA
SET3        ----------  ----------  ------GCGT  TTTATAA-GC  GCAAAAAATA
SET5        TTTAATGCCC  GAGTACCACG  AATGAAGCGT  TTTATAA-GC  GCAACAAATA
ORN1        TTTGACG---  ----------  ----------  ----------  ----------
MIM1        CTTAATTCCC  GAGTACTCCA  GAAGAAGCCC  TTCACAAAGA  ATAAAAAACT
MIM2        TTCAGTGCCC  GAGTACCACG  GA--------  ----TAAAGC  GCAAAAAATA
```

```
HET1        AAGGGA-CTA  GGATGACGTT  CGCCAA-AAC  CCACTTGGAC  AAGGATTTGG
HET3        AAGGGA-CTA  GGATGACGTT  CGCCAA-AAC  CCACTTGGAC  AAGGATTTGG
HET4        AAGGGA-CTA  GGATGACGTT  CGCCAA-AAC  CCACTTGGAC  AAGGATTTGG
SILV2       AAGGGA-CTA  GGATGAGGTT  CGCCAA-AAC  CCACTTGGAC  AAGGATTTGG
SILV3       AAGGGA-CTA  GGATGACGTT  CGCC------  -CACTTGGAC  AAGGATTTGG
SILV13      AAGGGA-CTA  GGATGAC?TT  CGCCAA-AAC  CCACTTGGAC  AAGGATTTGG
PLAN2       AAGGGA-TTA  GGATGACGTT  GGCCAA-AAC  CCACTTGGAC  AAGGATGTGG
PLAN3       AAGGGA-TTA  GGATGACGTT  GGCCAA-AAC  CCACTTGGAC  AAGGATGTGG
PLAN4       AAGGGA-TTA  GGATGACGTT  GGCCAA-AAC  CCACTTGGAC  AAGGATGTGG
PLAN11      AAGAGA-TTA  CGATGACGTT  CGCTAA-AAC  CCACTTGGAC  AAGGATTTCG
DIFF1       AAGGGA-TTA  GGATGACGTT  GGCCAA-AAC  CCACTTGGAC  AAGGATGTGG
DIFF2       AAGGGA-TTA  GGATGACGTT  GGCCAA-AAC  CCACTTGGAC  AAGGATGTGG
PICT1       AAGAAA-CTA  GGAT------  ----------  ------GGAC  AAGGATTTGG
PICT4       ----------  ----------  ----------  ----------  ----------
ADIA1       AAGGGG-CCA  GGATGACGTT  CGCCAA-AAT  CCTCTTGGGC  AAGGGTTTGG
ADIA2       AATGGA-CCA  GAATGACGTT  CTTCAA-ATC  CCACTTGGAC  AAGGATTTGA
ADIA5       AAGAAAACTA  GGATGACGTT  CGCCAA-AAT  CCACTTGAAC  AAGGATTTGG
PEN1        AAAGTA-CTA  TGATGACGTT  CGCCAA-AAC  CCACTTGGAC  AAGGATTTGG
PEN4        AAAGGA-CCA  GGATGACGTT  CTCCAA-ATC  CCGCTTGGAA  AAGAATTTGG
SET3        AAGGGG-CCA  GGACGACGTT  CTCCAA-ATC  TCACTTGGAC  AAGGATTTGA
SET5        AAGGGA-CCA  GGATGACGTT  CTCCAA-AAT  CCACTTGGAC  AAGAATTTAA
ORN1        ----------  ----------  ----------  ----------  ----------
MIM1        AACGGA-CTA  GGTTGACGTT  TGTCAACATC  CCACATGGAC  AAGGATTTGA
MIM2        CAGAGA-TTA  GGATGACXX-  CGCTAA-AAC  CCACTTGGAC  AAGGATTTG-

HET1        AGTTCTGG-A  ACACAATCAT  ATTTGAAGAT  GAGTCCAAAT  TCATAATTTT
HET3        AGTTCTGG-A  ACACAGTCAT  ATTTGAAGAT  GAGTCCAAAT  TCAACATTTT
HET4        AGTTCTGG-A  ACACAATCAT  ATTTGAAGAT  GAGTCCAAAT  TCATCATTTT
SILV2       AGTTCTGG-A  ACACAATCAT  ATTTGAAGAT  GAGTCCAAAT  TCATAATTTT
SILV3       AGTTCTGG-A  ACACAGTCAT  ATTTGAAGAT  GAGTCCAAAT  TTAACATTTT
SILV13      AGTTCTGG-A  ACACAATCAT  ATTTGAAGAT  GAGTCCAAAT  TCATAATTTT
PLAN2       AGTTCTGG-A  ACACAATCAT  ATTTGAAGAT  GAGTCCAAAT  TCATAATTTT
PLAN3       AGTTCTGG-A  ACACAATCAT  ATTTGAAGAT  GAGTCCAAAT  TCATAATTTT
PLAN4       AGTTCTGG-A  ACACAATCAT  ATTTGAAGAT  GAGTCCAAAT  TCATAATTTT
PLAN11      AGTTCTGG-A  ACACGGTCAT  ATTTGCAGAT  GAGTCCAAAT  TCAACATTTT
DIFF1       AGTTCTGG-A  ACACAATCAT  ATTTGAAGAT  GAGTCCAAAT  TCATAATTTT
DIFF2       AGTTCTGG-A  ACACAATCAT  ATTTGAAGAT  GAGTCCAAAT  TCATAATTTT
PICT1       AATTGTGG-A  ACACAGTCAT  ATTTGAAAAT  GAGTCAAAAT  TAAACATTTT
PICT4       ----------  ----------  ----------  ----------  ----------
ADIA1       AGTTCTGG-A  ACACAGTCAT  ATTCGAAGAT  GAGTCCAAAT  TCAAAGTTTT
ADIA2       AGTTCTGG-A  ACACACTCGT  ATTTGAAGAT  GAGTCCAAAT  TAAAAATTTT
ADIA5       AGTTCTGG-A  ATACAGTCAT  ATTTGCAAAT  GAGTCCAAAT  TCAACATTAT
PEN1        AGTTCTGG-A  ATACACTCAT  ATTTGCAGAT  CAGTCCAAAT  TCAAGATTTT
PEN4        AGTTGTGG-A  ACACAGTCAT  ATTTGCAAAT  GAATCCAA-T  TCAACA----
SET3        AGTTCTGG-A  ACACAGTCAT  ATTTGCAGAT  GAGTCCAAAT  TCAACATTTT
SET5        AGTTCTGG-A  ACACAGTCAT  AGTTGAAGAT  AAGTCCAAAT  TCAAAATTTT
ORN1        --TTCTGGCA  ACACAGTCAT  AGTTGCACAT  GAGTCCAAAG  GCAAAATTTT
MIM1        --TT------  ---TATCTAT  TTTTGAAGAC  TAGTCCAAAT  TCAAAATTTG
MIM2        AGTTCTGG--  ACATGGTCAT  ATTTAGCAGA  TGAXTCAA--  TCAACATTTT
```

```
HET1        TGGCTCGGAC  GGACGGAATT  ATGTGCGGCG  ACAGTCCAAT  ACTGAGCTGA
HET3        TGACTCGGAC  GGACGGAATT  ATGTGTGGCG  ACAGTCCAAT  ACTGAGCTGA
HET4        TGGCTCGGAC  GGACGGAATT  ATGTGTGGCG  ACAGTCCAAT  ACTGAGCTGA
SILV2       TGGCTCGGAC  GGACGGAATT  ATGTGCGGCG  ACAGTCCAAT  ACTGAGCTGA
SILV3       TGGATCGGAC  GGACTGAATT  ATGGGTGGCG  ACAATCCAAT  ACTGAGCCGA
SILV13      TGGCTCGGAC  GGACGGAATT  ATGTGCGGCG  ACAGTCCAAT  ACTGAGCTGA
PLAN2       TGGCTCGGAC  GGACGGAATT  ATGTGCGGCG  ACAGTCCAAT  ACTGAGCTGA
PLAN3       TGGCTCGGAC  GGACGGAATT  ATGTGCGGCG  ACAGTCCAAT  ACTGAGCTGA
PLAN4       TGGCTCGGAC  GGACGGAATT  ATGTGCGGCG  ACAGTCCAAT  ACTGGGCTGA
PLAN11      TGGCTCAGAC  GGACGGAATT  ATGTGTGGCG  ACAGTCCAAT  ACTGAGCTGA
DIFF1       TGGCTCGGAC  GGACGGAATT  ATGTGCGGCG  ACAGTCCAAT  ACTGAGCTGA
DIFF2       TGGCTCGGAC  GGACGGAATT  ATGTGCGGCG  ACAGTCCAAT  ACTGGGCTGA
PICT1       TGGCCCGGGC  GGACGAAATT  GTGTGTGGCG  ACAGTCCAAT  CATGAGCTGA
PICT4       ----------  ----------  ----------  --AGTCCAAT  GCTGAGCTCA
ADIA1       TGACTCGGTT  GGACGGAATT  ATAGGTGGCG  ACAATCCAAT  ACTGAXCTGA
ADIA2       TGGCTCGGAG  AGAGGGAATT  GTGTGTGGCG  ACAGTCCAAT  AC-GAGCTGG
ADIA5       TGGCTCGGAC  GGACGGAATT  ATGGGTGACG  ACAATTGAAT  GCTGAACTGA
PEN1        TGGCTCGGA-  -------ATT  ATGTGGGCCG  GCAGTCCAAT  ACTGAGCTAA
PEN4        -----CGGA-  GGA-GGAA--  --GTGTCATG  GTA-------  ----------
SET3        TGGCTCGAAC  AGCCGGCATT  ATGTGTGGCA  ACAGTCCAAA  GCTGAACTGA
SET5        TGGCTCGGAC  CGACGGCATT  ATGTGTGGCG  ACAGTCCAAT  ACTGAGCTGG
ORN1        TGGCTCAGAC  GGACGGCAT-  --GTGTGGCG  ACAGCTCAAT  ACTGAGCTGA
MIM1        TGGCTCGAAC  GGAAGGAAGT  TTGTGTGGCG  ACAGTC----  ----------
MIM2        --XCTCAGAC  G-ACGGAATT  ATGTGTGGCG  TCAGTCCAAT  --GAGCTTGA

HET1        ATCCCAAAAA  CCTAAAGGCA  ACAGTGAAGC  ACGGCGGAGG  AAGTGTCATG
HET3        ATCCGAAACA  CCTAAAGGCA  ACAGTGAAGC  ACGGCGGAGG  AAGTGTCATG
HET4        ATCCGAAACA  CCTAAAGGCA  ACAGTGAAGC  ACGGCGGAGG  AAGTGTCATG
SILV2       ATCCCAAAAA  CCTAAAGGCA  CCAGTGAAGC  ACGGGGAGGG  AAGTGTCATT
SILV3       ATCCCAAAGG  CTTAAAGCCA  CCGGTGAAGC  ACGGAGGAGG  AAGTGTCATG
SILV13      ATCCCAAAAA  CCTAAAGGCA  ACAGTGAAGC  ACGGCG-AGG  A-GTGTCATG
PLAN2       ATCCCAAAAA  CCTAAAGTCA  ACAGTGAAGC  ACGGCGGAGG  AAGTGTCATG
PLAN3       ATCCCAAAAA  CCTAAAGTCA  ACAGTGAAGC  ACGGCGGAGG  AAGTGTCATG
PLAN4       ATCCCAAAAA  CCTAAAGTCA  ACAGTGAAGC  ACGGCGGAGG  AAGTGTCATG
PLAN11      ATCCCAAAAA  CCTAAAGTCA  ACAGTCAAGC  ACGGCGGAGG  AAGTGTCATG
DIFF1       ATCCCAAAAA  CCTAAAGTCA  ACAGTGAAGC  ACGGCGGAGG  AAGTGTCATG
DIFF2       ATCCCAAAAA  CCTAAAGTCA  ACAGTGAAGC  ACGGCGGAGG  AAGTGTCATG
PICT1       ATCCCAAAAA  CCTAAAGGCA  ACAGTCAAGC  ACGGCGGAGG  AAGTTTCATG
PICT4       ATCCCAAAAA  CCTAAAGGTA  ACCGACAAGC  ACAGCAGCGG  AGGTGTCATG
ADIA1       ATCCCAAAAA  CGTAAAGGCA  ACAGTCAAGC  ACGGAGGAAG  AAGTGTCATG
ADIA2       ATCCCAAAAA  CGTAAAGGCA  ACAGTCAAGC  ACGGCGGAGG  AAGTGACATG
ADIA5       ATCCCAAAAA  CGTAAAGGCA  ACAGTCAACC  ACGGCGGAGG  AAATGACATG
PEN1        ATCC-AAAAA  CTTAAAGGCA  A---------  ----------  ----------
PEN4        ----------  ----------  ----------  ----------  ----------
SET3        ATCC-AAAAA  CGTAAAGGCA  ACAGTCAAAC  ACGGAGGAGG  AAGTGTCATG
SET5        ATCCCAAAAA  CGTGAAGGCA  ACAGTCAAGC  ACGGCGGAGG  AAGTGACATG
ORN1        ATCCC--AAC  CGTAAAGGCA  ACAGTCAAGC  ACGGCGGAG-  AAGTGTCATG
MIM1        ----------  --------CA  AC--------  -CGGT--AGG  AAATG-----
MIM2        ATCCXAAAAA  CCTAAAGTCA  ACAGTCAAG-  ACGGXGGAGG  AAGTGTCATG
```

```
HET1        GTATGGGCAT  GTATCTCCGC  AGCCAGCGTC  GGAAATTTGG  TGTGTATTGA
HET3        GTATGGGCAT  GTATCTCCGC  AGCCGGCGTC  GGAAATTTGG  TGTGTATTGA
HET4        GTATGGGCAT  GTATCTCCGC  AGCCGGCGTA  GGGAATTTGG  TGTTCATTGA
SILV2       GTATGGGCAT  GTATCTCGGC  AGCCAGCGTC  GGAAATTTGG  TGTGTATTGA
SILV3       GTATGGGCAT  GTATCTCGGC  AGCCGGCGTC  GGAAATTTGG  TGTTTATTGA
SILV13      GTATGGGCAT  GTATCTCGGC  AGCCGGCGTC  GGAAATTTGG  TGTGTATTGA
PLAN2       GTATGGGCAT  CTATCTCGGC  AGCCGGCGTC  GGAAATTTGG  TGTTCATTGA
PLAN3       GTATGGGCAT  CTATCTCGGC  AGCCGGCGTC  GGAAATTTGG  TGTTCATTGA
PLAN4       GTATGGGCAT  CTATCTCGGC  AGCCGGCGTC  GGAAATTTGG  TGTTCATTGA
PLAN11      GTATGGGCAT  CTATCTCGGC  AGCCGGCGTC  GGAAATTTGG  TGTTCACTGA
DIFF1       GTATGGGCAT  CTATCTCGGC  AGTCGGCGTC  GGAAATTTGG  TGTTCATTGA
DIFF2       GTATGGGCAT  CTATCTCGGC  AGCCGGCGTC  GGAAATTTGG  TGTTCATTGA
PICT1       GTATGGGCAT  CTATATC---  ----------  ----------  ----TATTAC
PICT4       GTAGGGGCAT  CTATCTCGCA  AGTCGTCGTC  GGAAATTTGG  TGTTCATCGA
ADIA1       GTGTGGGCTT  GTATCTCTAG  CAGCCGGC-X  XXGAATTTGG  TGTTTATTGG
ADIA2       GTATGGGCAT  CTATCTCGGC  AGCCAGCATC  GGGAATTTGG  TGTTCATTGA
ADIA5       GCATGGG---  ---------C  AGCCGG-GTC  GGGAATTTGG  T-TTCATTGA
PEN1        ----------  ----------  ----------  ----------  ----------
PEN4        ---TGGGCAT  GTATCTCGGC  AGCCGGAGTC  GGGAATTTGG  TGTTTATTGG
SET3        GTATGGGCAT  CTATCTCGGC  AGCCGACGTC  GGGAATTTGG  TGTTTAATTG
SET5        GTATGGGCAT  CTTTCTCGGC  AGCCAGAGTC  GGTAATTTGA  TGTTTATTGG
ORN1        G--------T  CTATCTCGGC  A??????????  ??????????  ??????????
MIM1        ----------  ------CGGC  AGCCAGTGTC  GGAAACTTGG  CGTTAATTGA
MIM2        GTATGGX-AT  CTATXTCGGC  TAGXCGXGTC  GGAAATTTGG  TGTTCACTGA

HET1        AACAACAACG  GACAGGAATG  TGGACCTCAG  TATATTAAAG  GAAAATTTAC
HET3        AACAACAACG  GACAGGAATG  TGGACCTCAG  AATATTAAAG  GAAAATTTAC
HET4        ACCAACAATG  GACAAGAATG  TATATCTCAA  TATACTAAAG  GAAAATTTAC
SILV2       AACAACAACG  GACAGGAATG  TGGACCT???  ??????????  ??????????
SILV3       AACAACAATC  GACAAGAATG  TGTACCTCAA  TATATTAAAG  GAAAATGTAC
SILV13      AACAACAACG  GACAGGAATG  TGGACCTCAG  TATATTAAAG  GAAAATTTAC
PLAN2       AACAACAATG  GACTTGAATG  TGTACCTCAA  TATATTAAAG  GAAAA?????
PLAN3       AACAACAATG  GACATGAATG  ??????????  ??????????  ??????????
PLAN4       AACAACAATG  GACATGAATG  TGTACCTCAA  TATATTAAAG  GAAA??????
PLAN11      AACAACACTG  GACAAGAATG  TGTACCTCAA  TATATTAAAG  GAAAATGTAC
DIFF1       AACAACAATG  GACATGAATG  TGTACCTCAA  TATATTAAAG  GAAAATTTA?
DIFF2       AACAACAATG  GACATGAATG  TGTACCTCAA  TATATTAAAG  GAAAATTT??
PICT1       AACAACAATG  GACCAGAATA  TGGATCTCAA  TATATTAAAG  GAAAATTT??
PICT4       AACAACAATG  GACAAGAATG  TGTACCTCAA  TATATTAAAG  GAAAATGTAC
ADIA1       AACAACAATG  GACAAGAATG  TGTACTTTAA  TATTAACATG  GAAAAT-TAC
ADIA2       AACAACAATG  GACAAGAATG  TGTTCTTCAA  TATATTAAAG  GGAAATTTAC
ADIA5       AACAATAA??  ??????????  ??????????  ??????????  ??????????
PEN1        ----------  ----------  ----------  ----------  ----------
PEN4        AACAAGAATG  GACAAGAATG  TGTACTTCAA  TATATTAAAG  GAAAATTTTC
SET3        GACAGCAATG  GACAAGAATG  TGGACTTCAT  AATATTAAAG  GAAATTTGAC
SET5        AACAAAAATG  AACAAGAATG  TGTACTTCAA  TATATAAAAG  AAAAATTTAC
ORN1        ??????????  ??????????  ??????????  ??????????  ??????????
MIM1        GACAACAATG  GATAGGAATA  TGTATCTTAA  TATACT---T  TAAAAT-TAA
MIM2        AACAACACTG  GACAAGAATG  TGTACCTCAA  TATATTAAAG  GAAAATGTAC
```

```
HET1        T-CCAAAGT   409
HET3        T-CCAAAGT
HET4        T-CCAAAGT
SILV2       ?????????
SILV3       T-CCAAAGT
SILV13      T-CCAAAGT
PLAN2       ?????????
PLAN3       ?????????
PLAN4       ?????????
PLAN11      T-CCAAAGT
DIFF1       ?????????
DIFF2       ?????????
PICT1       ?????????
PICT4       T-CCAAAGT
ADIA1       T-CCAAAGA
ADIA2       T-CCAAAGT
ADIA5       ?????????
PEN1        ---------
PEN4        T-TCAAAGT
SET3        T-CCAAAGT
SET5        T-CCAAAGA
ORN1        ?????????
MIM1        TACAAAAGC
MIM2        T-CCAAAGT
```

Figure 18: The synonymous/ nonsynonymous distance matrix for the 24 isolates of *uhu*. The distances were obtained using Kimura's two parameter method which weights for transversions versus transitions. The synonymous rate of change is above the diagonal, the nonsynonymous is below the diagonal. Synonymous rates of change that are twice the nonsynonymous are underlined, those that are 2.5 times the nonsynonymous are shaded.

	HET1	HET3	HET4	SILV2	SILV3	SILV13	PLAN2
HET1	0	0.02424	0.10038	0.05331	0.15113	0.01412	0.03748
HET3	0.02562	0	0.07528	0.08172	0.14646	0.0434	0.06423
HET4	0.02893	0.03546	0	0.15623	0.2348	0.13541	0.11905
SILV2	0.02505	0.05096	0.05465	0	0.15008	0.01529	0.0666
SILV3	0.07254	0.07032	0.05297	0.07745	0	0.14691	0.13727
SILV13	0.00368	0.02604	0.02988	0.01649	0.06843	0	0.02934
PLAN2	0.06882	0.09053	0.06138	0.08175	0.08077	0.05588	0
PLAN3	0.05719	0.07707	0.05681	0.07482	0.08262	0.04503	0.00738
PLAN4	0.06554	0.08713	0.05796	0.07817	0.07733	0.05602	0.00993
PLAN11	0.08744	0.08717	0.06765	0.10617	0.07296	0.08116	0.07511
DIFF1	0.06429	0.08545	0.0569	0.07799	0.07942	0.05893	0.01659
DIFF2	0.06495	0.08633	0.05744	0.07817	0.07661	0.05543	0.00993
PICT1	0.10677	0.10639	0.10578	0.11553	0.11285	0.11123	0.11739
PICT4	0.14826	0.1547	0.11916	0.20579	0.13653	0.13873	0.12869
ADIA1	0.16071	0.16002	0.14079	0.16434	0.1565	0.16841	0.18537
ADIA2	0.13067	0.14546	0.11912	0.14678	0.14381	0.12792	0.1504
ADIA5	0.12184	0.12421	0.12426	0.13747	0.12207	0.12032	0.14357
PEN1	0.06657	0.08457	0.07743	0.06659	0.09575	0.06752	0.10432
PEN4	0.152	0.14022	0.11728	0.16432	0.10622	0.14635	0.15649
SET3	0.13949	0.13018	0.12946	0.1581	0.14618	0.14039	0.15151
SET5	0.14211	0.15104	0.13021	0.15598	0.14535	0.13861	0.15006
ORN1	0.13539	0.13403	0.13524	0.16532	0.17504	0.14138	0.13577
MIM1	0.23572	0.25379	0.22632	0.23937	0.24378	0.18959	0.25392
MIM2	0.13931	0.13653	0.11782	0.1589	0.13045	0.15418	0.13436

	PLAN3	PLAN4	PLAN11	DIFF1	DIFF2	PICT1	PICT4
HET1	0.02735	0.03743	0.11846	0.03712	0.03727	0.19206	0.15975
HET3	0.0564	0.06407	0.12037	0.06352	0.06379	0.19381	0.18499
HET4	0.07147	0.11847	0.18484	0.11748	0.11796	0.19476	0.29772
SILV2	0.05617	0.06641	0.14561	0.06671	0.06641	0.27326	0.1872
SILV3	0.15236	0.13618	0.18532	0.13493	0.13546	0.27121	0.19884
SILV13	0.01665	0.02935	0.10905	0.02905	0.0292	0.23733	0.13431
PLAN2	-0.00005	-0.00004	0.10778	-0.00004	-0.00004	0.17573	0.11849
PLAN3	0	0	0.1184	0	0	0.14365	0.13201
PLAN4	0.00368	0	0.10767	0	0	0.16549	0.11751
PLAN11	0.07923	0.07521	0	0.1065	0.10717	0.20229	0.13457
DIFF1	0.01109	0.0133	0.07762	0	0	0.16519	0.11554
DIFF2	0.00368	0	0.07444	0.01318	0	0.16435	0.1166
PICT1	0.12839	0.12493	0.10753	0.11305	0.12341	0	0.20261
PICT4	0.13189	0.12234	0.12567	0.10329	0.12018	0.13261	0
ADIA1	0.17165	0.18059	0.17387	0.17842	0.17865	0.19425	0.2306
ADIA2	0.15151	0.15465	0.14848	0.15156	0.15316	0.15435	0.18169
ADIA5	0.14685	0.1495	0.1263	0.15467	0.1495	0.14763	0.22234
PEN1	0.11183	0.11236	0.08564	0.11202	0.11236	0.14557	0.08613
PEN4	0.1675	0.15005	0.1454	0.16401	0.14785	0.15198	0.13663
SET3	0.1479	0.15188	0.1321	0.15772	0.15495	0.15986	0.19311
SET5	0.15133	0.1543	0.14622	0.15127	0.15282	0.19037	0.22722
ORN1	0.13577	0.14628	0.11553	0.13577	0.14628	0.1747	0.27612
MIM1	0.23692	0.24132	0.25458	0.2571	0.23795	0.25863	0.36806
MIM2	0.1398	0.11778	0.0701	0.12656	0.11643	0.16083	0.12832

	ADIA1	ADIA2	ADIA5	PEN1	PEN4	SET3	SET5
HET1	0.20724	0.10701	0.28447	0.23435	0.25719	0.20994	0.13632
HET3	0.229	0.12281	0.27625	0.22365	0.23552	0.17451	0.16107
HET4	0.31005	0.16608	0.25146	0.22005	0.3368	0.27148	0.24373
SILV2	0.20452	0.11672	0.33657	0.23578	0.22232	0.21173	0.14644
SILV3	0.24196	0.15498	0.30759	0.26426	0.27697	0.21714	0.20484
SILV13	0.23585	0.09313	0.29865	0.24218	0.23445	0.19811	0.11685
PLAN2	0.23086	0.0887	0.24073	0.20011	0.25896	0.22012	0.14826
PLAN3	0.19901	0.06208	0.23287	0.21804	0.24605	0.20656	0.14573
PLAN4	0.23188	0.08847	0.24042	0.19815	0.26057	0.21989	0.14778
PLAN11	0.27764	0.11117	0.28111	0.31884	0.30354	0.21768	0.14693
DIFF1	0.23063	0.08774	0.24129	0.20011	0.25634	0.21795	0.14716
DIFF2	0.23063	0.08809	0.24042	0.19815	0.25858	0.21848	0.26013
PICT1	0.38554	0.21625	0.23647	0.37884	0.5076	0.35685	0.25742
PICT4	0.39542	0.15841	0.33156	0.33174	0.33347	0.33442	0.22654
ADIA1	0	0.19231	0.30645	0.33851	0.29454	0.25123	0.22654
ADIA2	0.18397	0	0.23565	0.28048	0.16377	0.13708	0.10941
ADIA5	0.17698	0.1825	0	0.30605	0.43504	0.24601	0.33552
PEN1	0.16846	0.14149	0.13933	0	0.33452	0.35747	0.33793
PEN4	0.20658	0.14833	0.1667	0.18173	0	0.20058	0.14835
SET3	0.19756	0.14064	0.17157	0.13436	0.16352	0	0.16722
SET5	0.16031	0.12595	0.1511	0.12455	0.15733	0.14458	0
ORN1	0.17461	0.16196	0.16793	0.17381	0.23829	0.12813	0.11611
MIM1	0.34289	0.27831	0.23235	0.2038	0.27726	0.27914	0.32356
MIM2	0.21661	0.20695	0.19869	0.22562	0.21557	0.2025	0.20295

	ORN1	MIM1	MIM2
HET1	0.12704	0.45931	0.18931
HET3	0.16729	0.46528	0.17345
HET4	0.20071	0.52506	0.25258
SILV2	0.16308	0.46315	0.23418
SILV3	0.32237	0.59338	0.21825
SILV13	0.09731	0.44304	0.20769
PLAN2	0.09234	0.40595	0.16227
PLAN3	0.09234	0.36568	0.17996
PLAN4	0.09142	0.40134	0.1705
PLAN11	0.09627	0.5361	0.08453
DIFF1	0.09234	0.39719	0.15934
DIFF2	0.09142	0.39719	0.16952
PICT1	0.15898	0.66343	0.17935
PICT4	0.17631	0.67351	0.13142
ADIA1	0.26267	0.64461	0.36168
ADIA2	0.08049	0.59543	0.20248
ADIA5	0.2546	0.56439	0.40557
PEN1	0.33347	0.38182	0.46638
PEN4	0.29709	0.77796	0.34084
SET3	0.15916	0.65653	0.30784
SET5	0.09469	0.59095	0.26141
ORN1	0	0.60076	0.42613
MIM1	0.28638	0	0.61043
MIM2	0.27481	0.34883	0

Figure 19: The consensus tree from a bootstrap analysis using Kimura's two parameter distance (scale at the bottom). Branch lengths indicate relative distances. Numbers at the nodes indicate the number of times out of a hundred trees all the species to the right of the node occurred together (how many times that node occurred). An * after a species indicates isolates that are known not to have the 10 base pair duplication (Figure 3), † indicates isolates that are known to have the duplication.

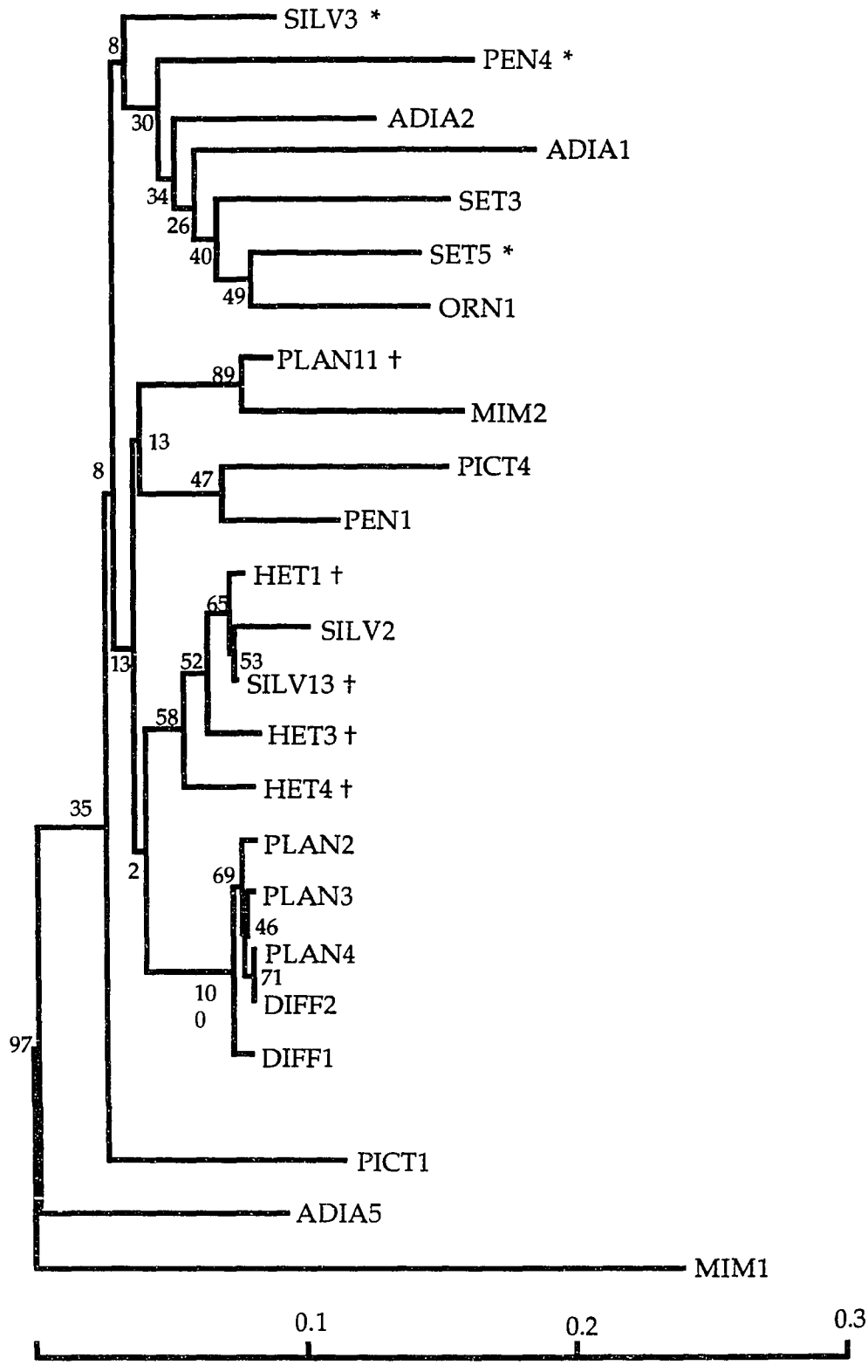

SILV3 *
PEN4 *
ADIA2
ADIA1
SET3
SET5 *
ORN1
PLAN11 †
MIM2
PICT4
PEN1
HET1 †
SILV2
SILV13 †
HET3 †
HET4 †
PLAN2
PLAN3
PLAN4
DIFF2
DIFF1
PICT1
ADIA5
MIM1
8
30
34
26
40
49
89
13
8
47
65
13
52
53
58
35
2
69
46
71
10
0
97
0.1
0.2
0.3

Figure 20: A Maximum Likelihood phylogeny of the 24 isolates of *uhu*. Distances were calculated using Kimura's two-parameter method. Open boxes indicate the 95% confidence limits of the nodes. An * after a species indicates isolates that are known not to have the 10 base pair duplication (Figure 3), † indicates isolates that are known to have the duplication.

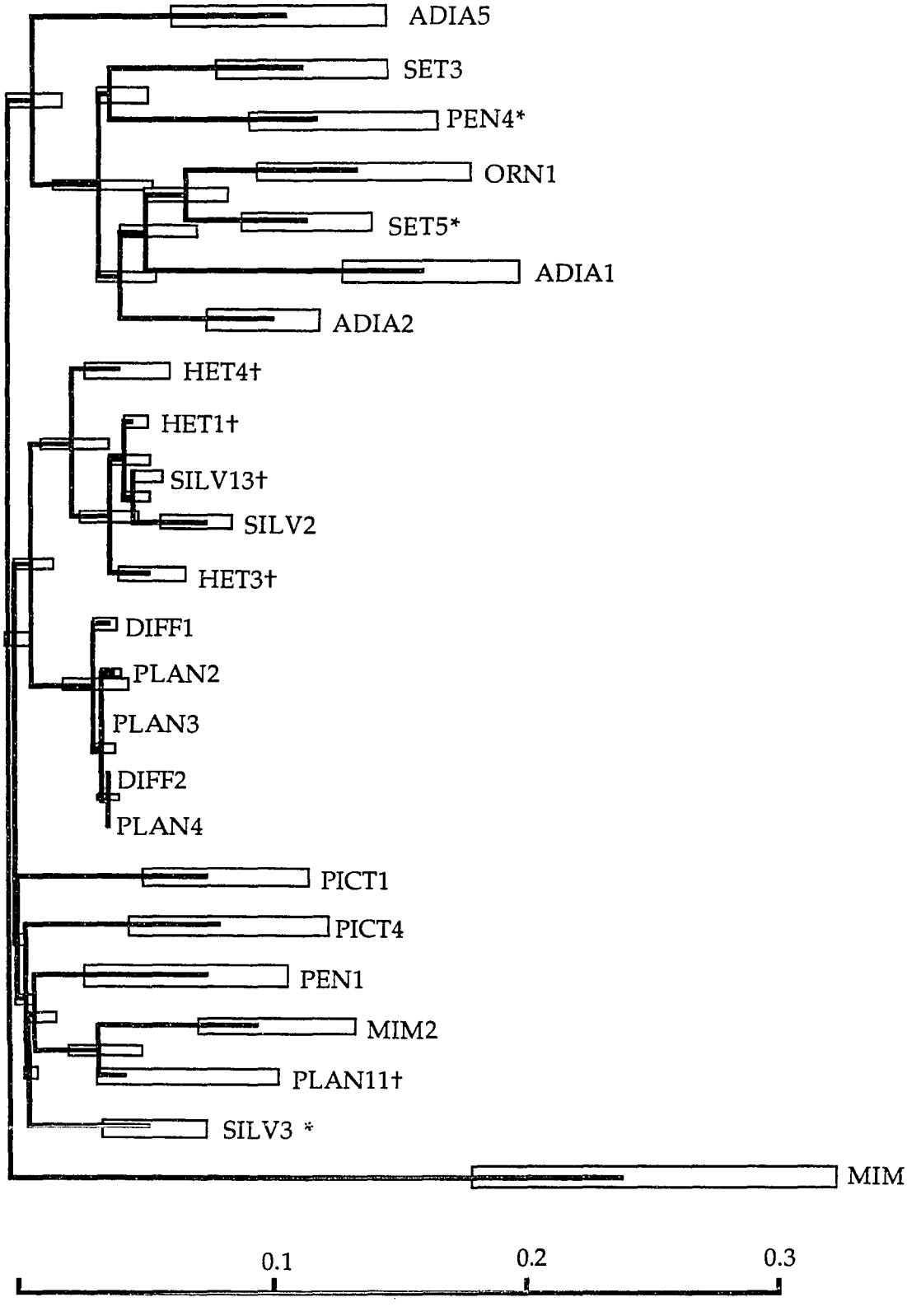

ADIA5
SET3
PEN4*
ORN1
SET5*
ADIA1
ADIA2
HET4†
HET1†
SILV13†
SILV2
HET3†
DIFF1
PLAN2
PLAN3
DIFF2
PLAN4
PICT1
PICT4
PEN1
MIM2
PLAN11†
SILV3 *
MIM
0.1
0.2
0.3

Figure 21: One of the best trees obtained by parsimony analysis for the 24 isolates of *uhu*. An * after a species indicates isolates that are known not to have the 10 base pair duplication (Figure 3), † indicates isolates that are known to have the duplication.

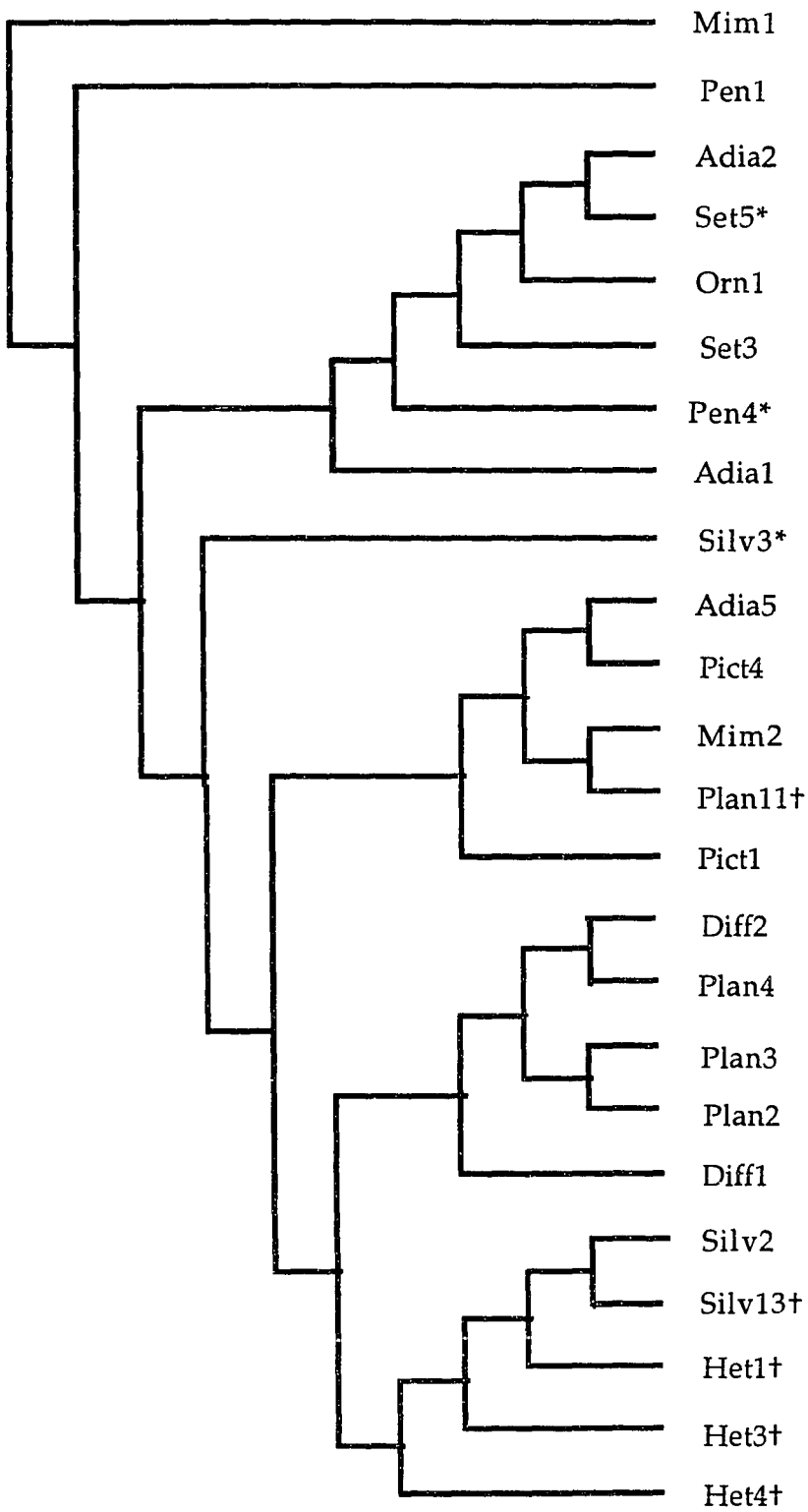
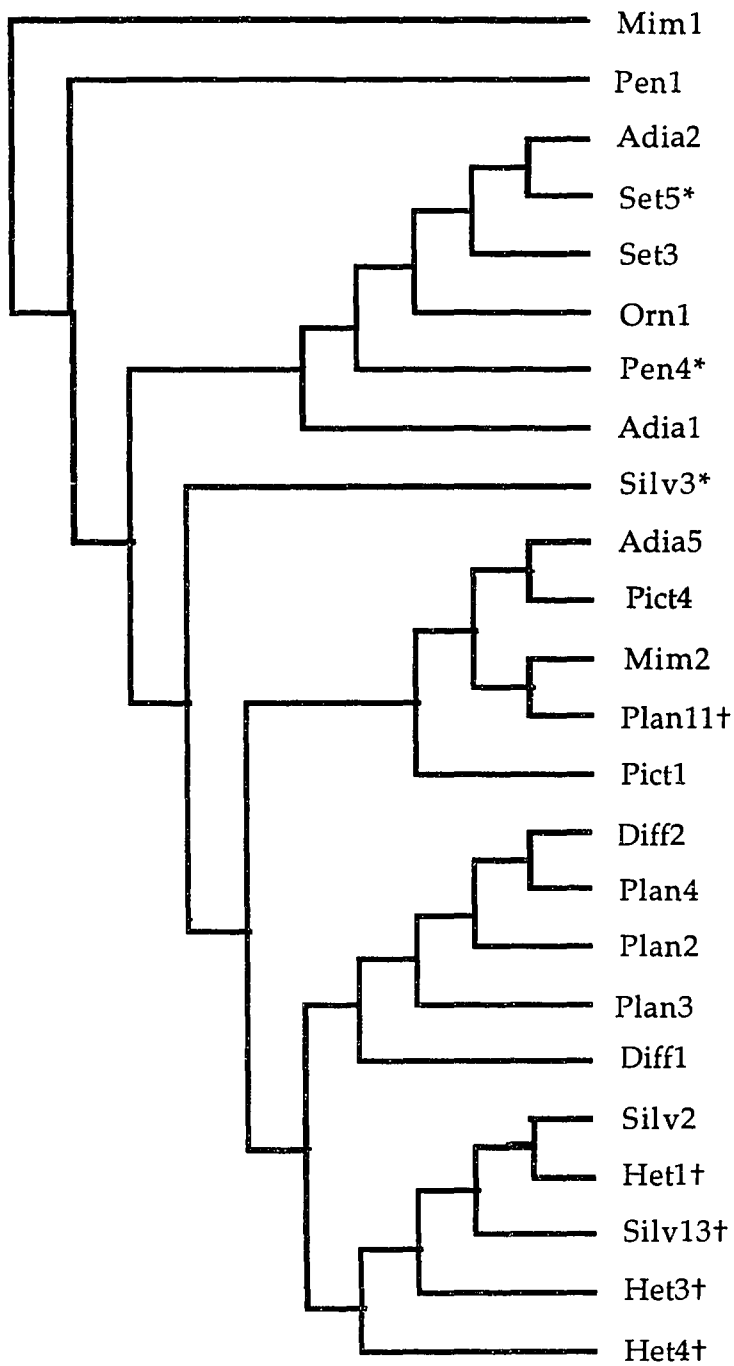
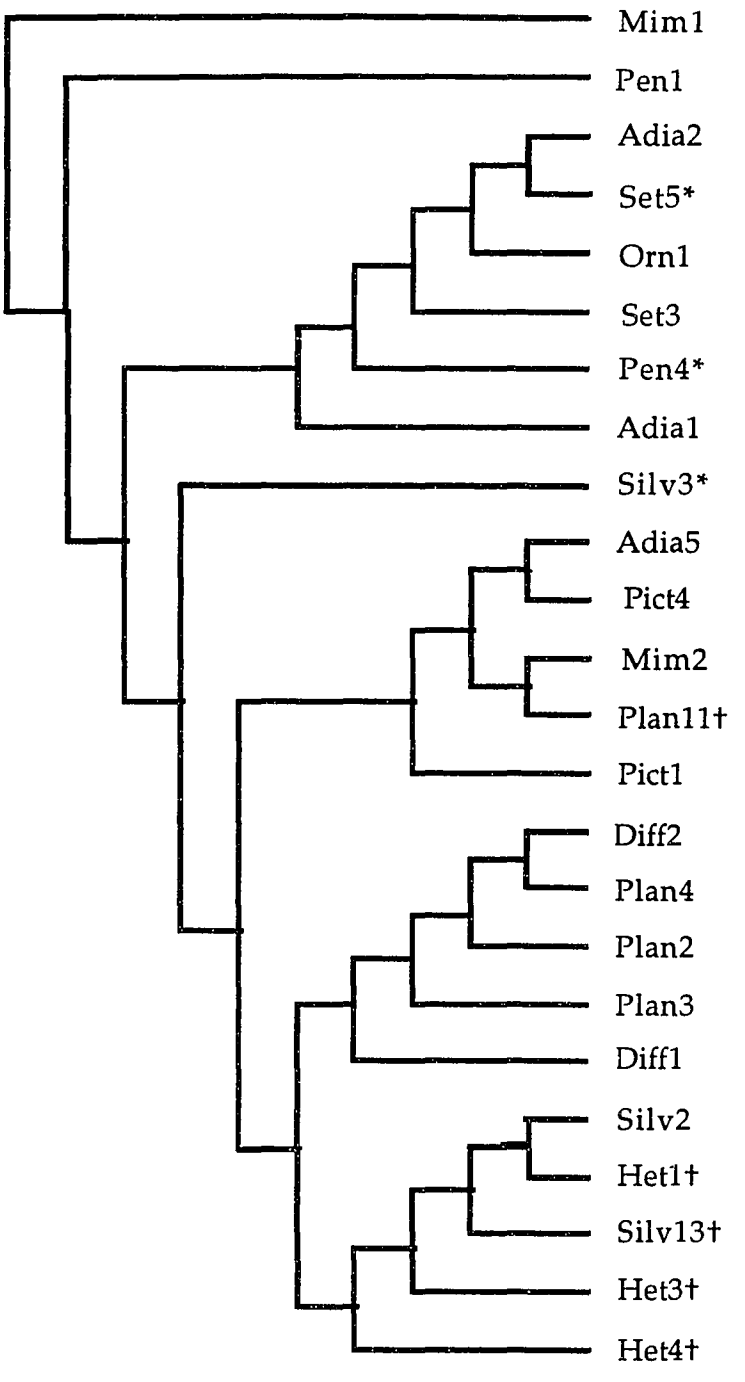
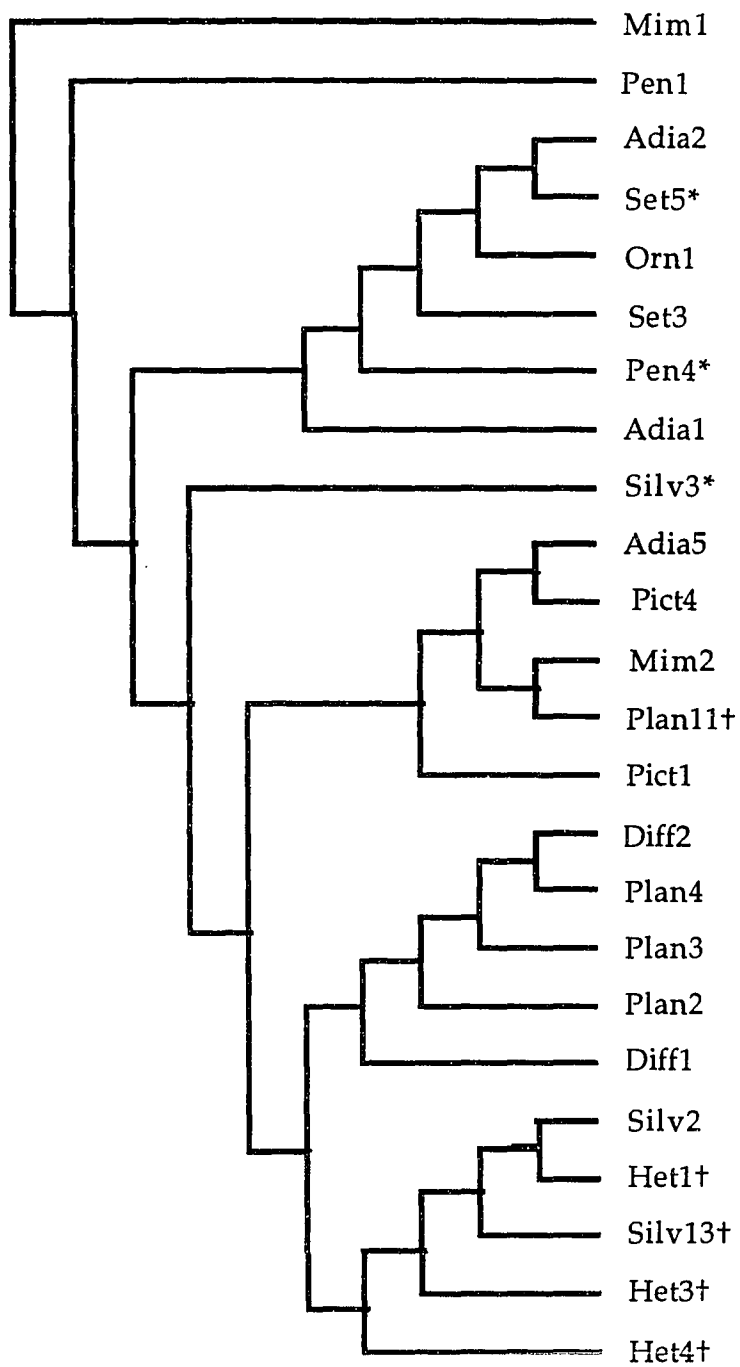
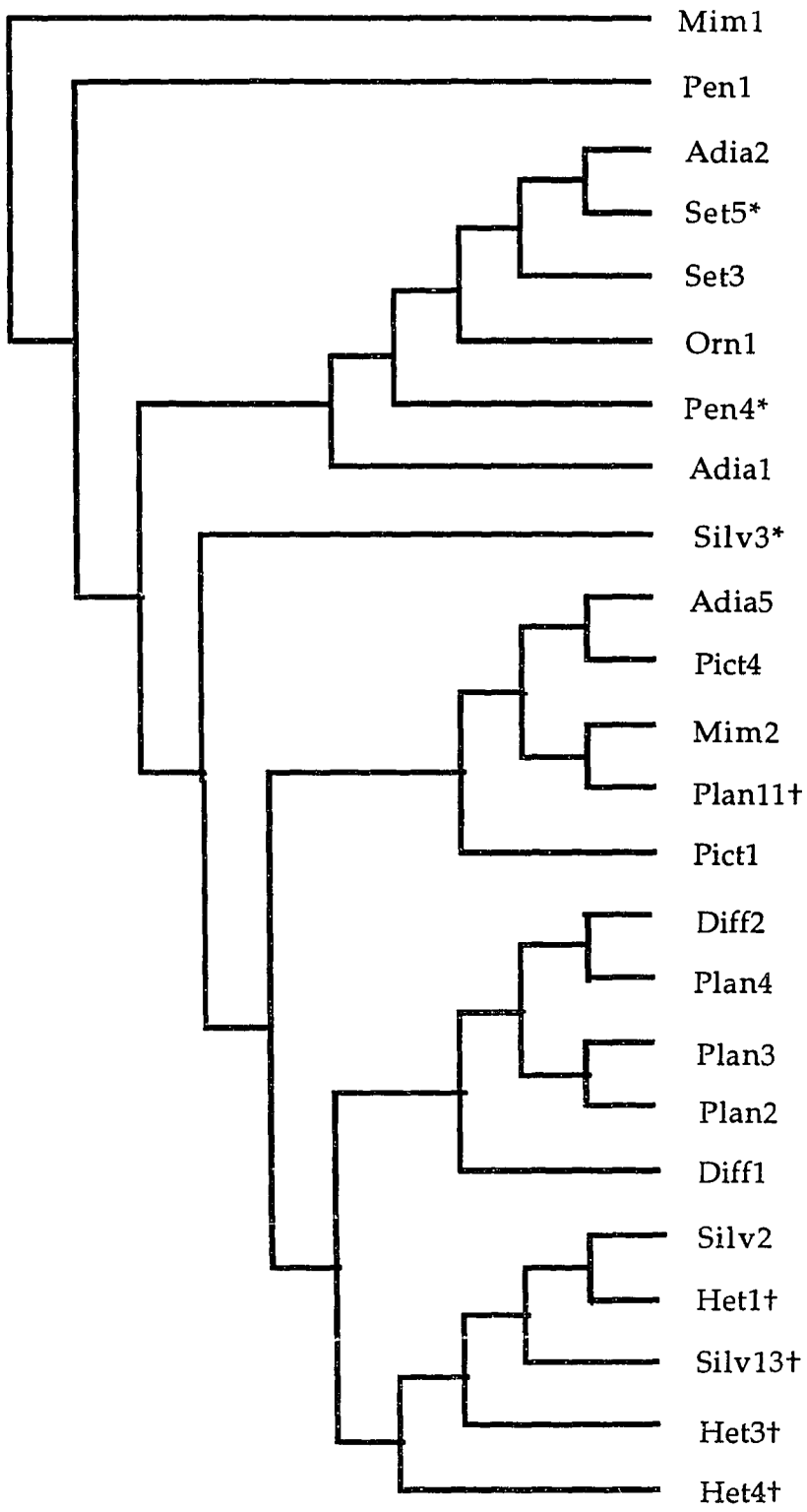

Mim1
Pen1
Adia2
Set5*
Orn1
Set3
Pen4*
Adia1
Silv3*
Adia5
Pict4
Mim2
Plan11†
Pict1
Diff2
Plan4
Plan3
Plan2
Diff1
Silv2
Silv13†
Het1†
Het3†
Het4†

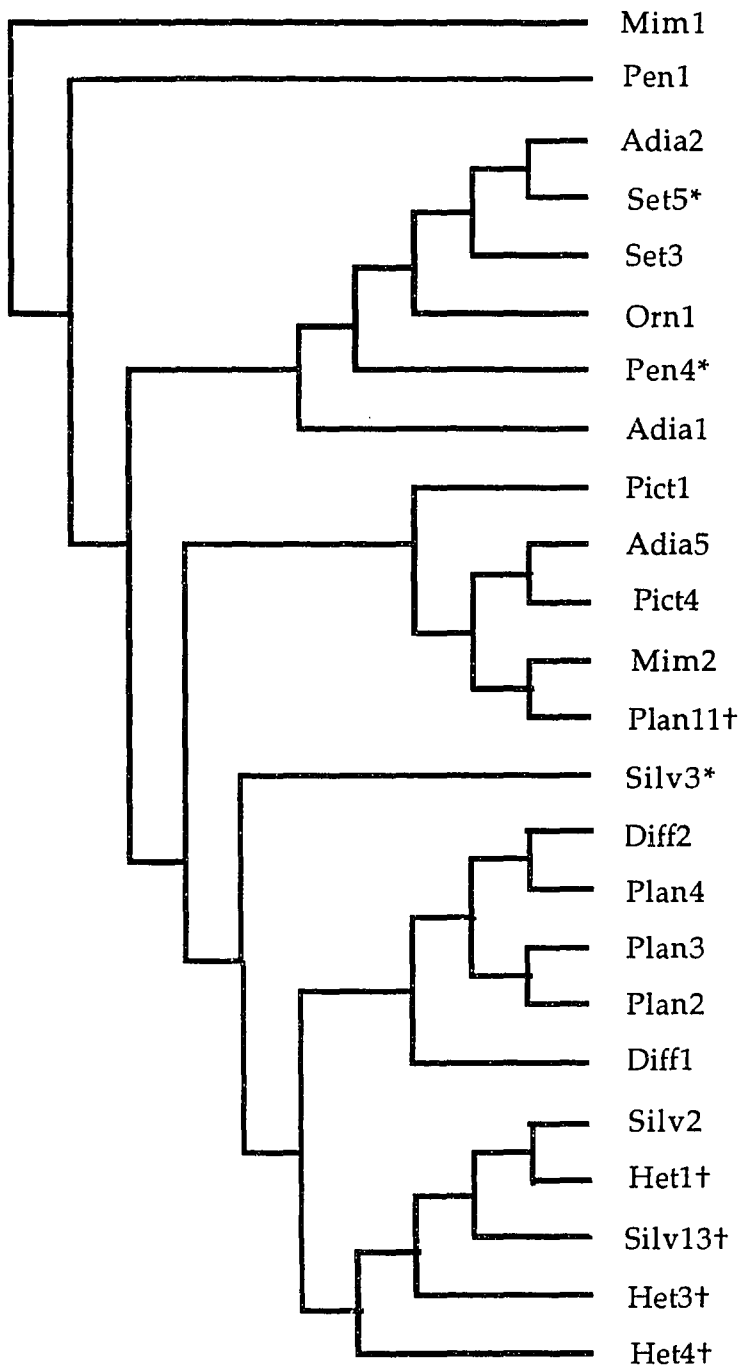
Figure 22: Eight additional trees that were obtained by parsimony analysis for the 24 isolates of *uhu*. These trees are of equal distance to the one in Figure 23. They differ primarily in the branching order within the lower two clusters, the branching order of Set3 and Orn1 or the placement of Silv3.

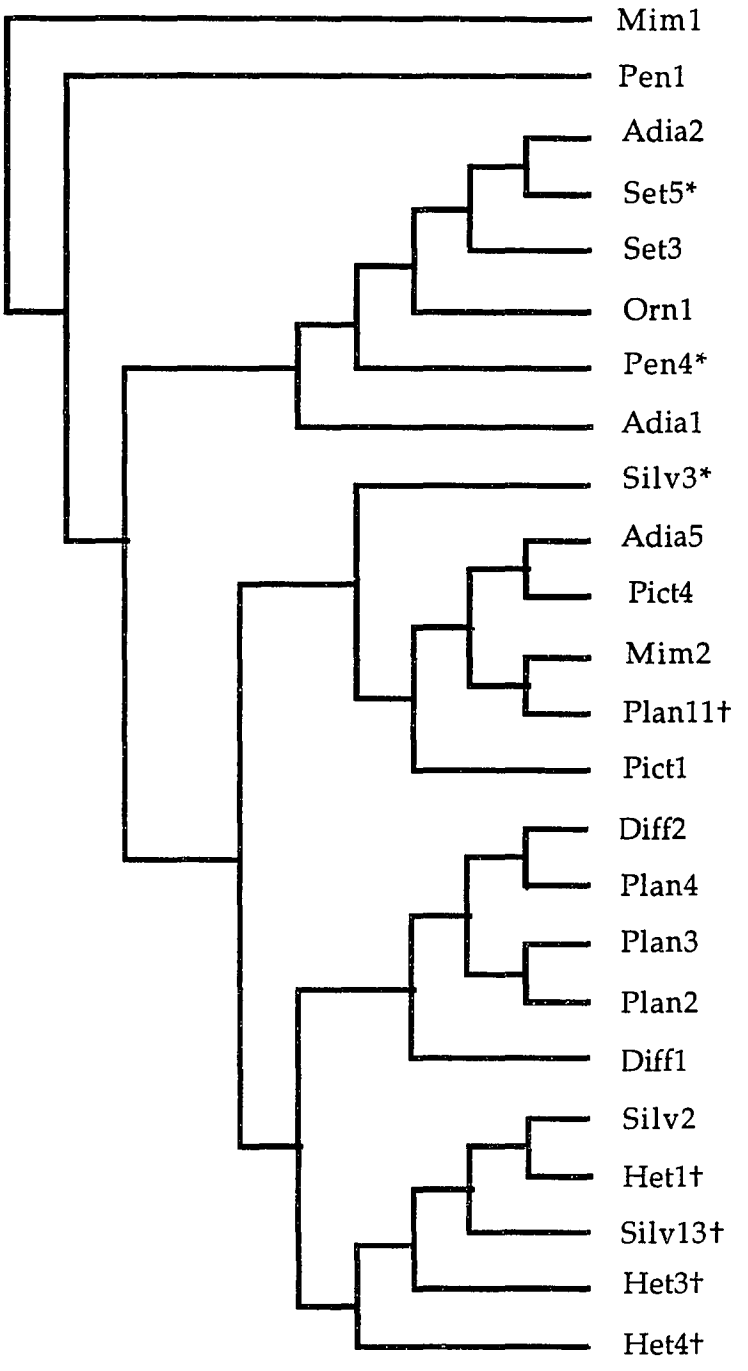

Mim1
Pen1
Adia2
Set5*
Set3
Orn1
Pen4*
Adia1
Silv3*
Adia5
Pict4
Mim2
Plan11†
Pict1
Diff2
Plan4
Plan2
Plan3
Diff1
Silv2
Het1†
Silv13†
Het3†
Het4†

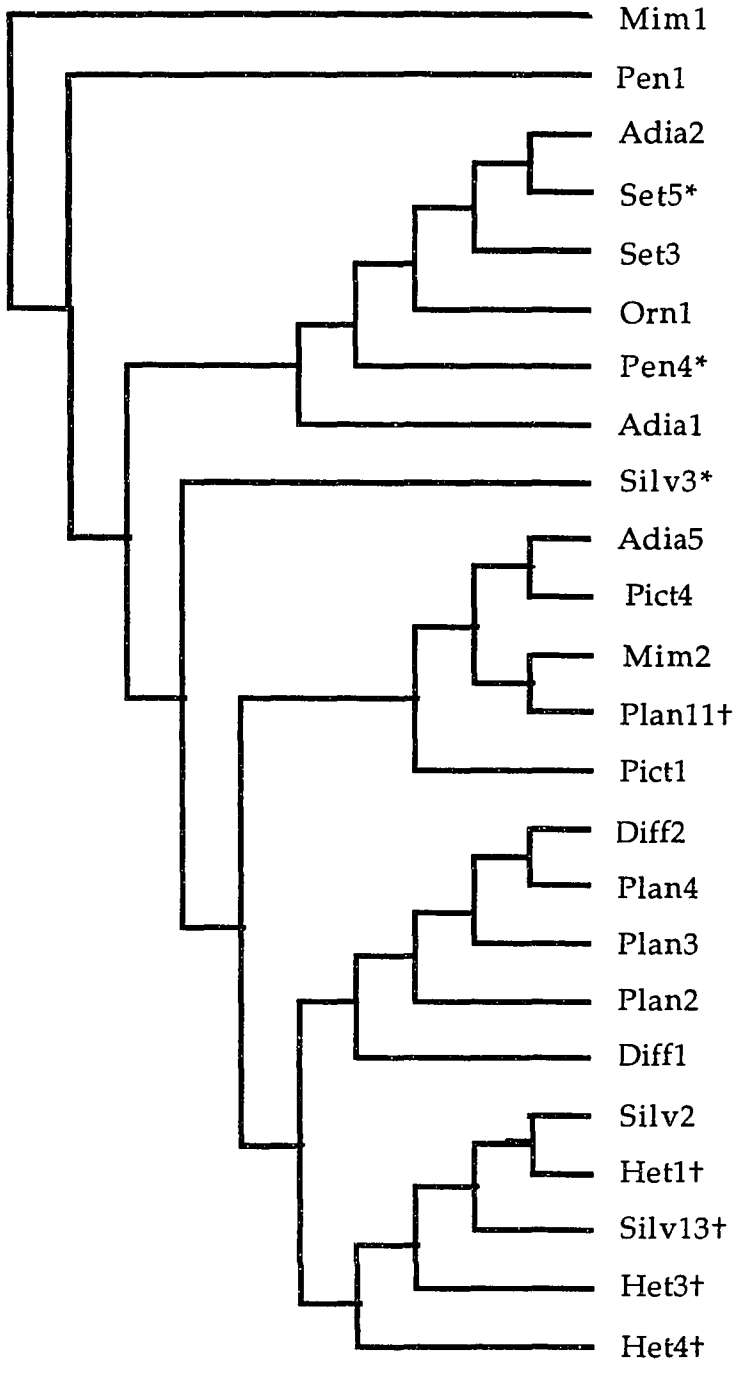

Mim1
Pen1
Adia2
Set5*
Orn1
Set3
Pen4*
Adia1
Silv3*
Adia5
Pict4
Mim2
Plan11†
Pict1
Diff2
Plan4
Plan2
Plan3
Diff1
Silv2
Het1†
Silv13†
Het3†
Het4†

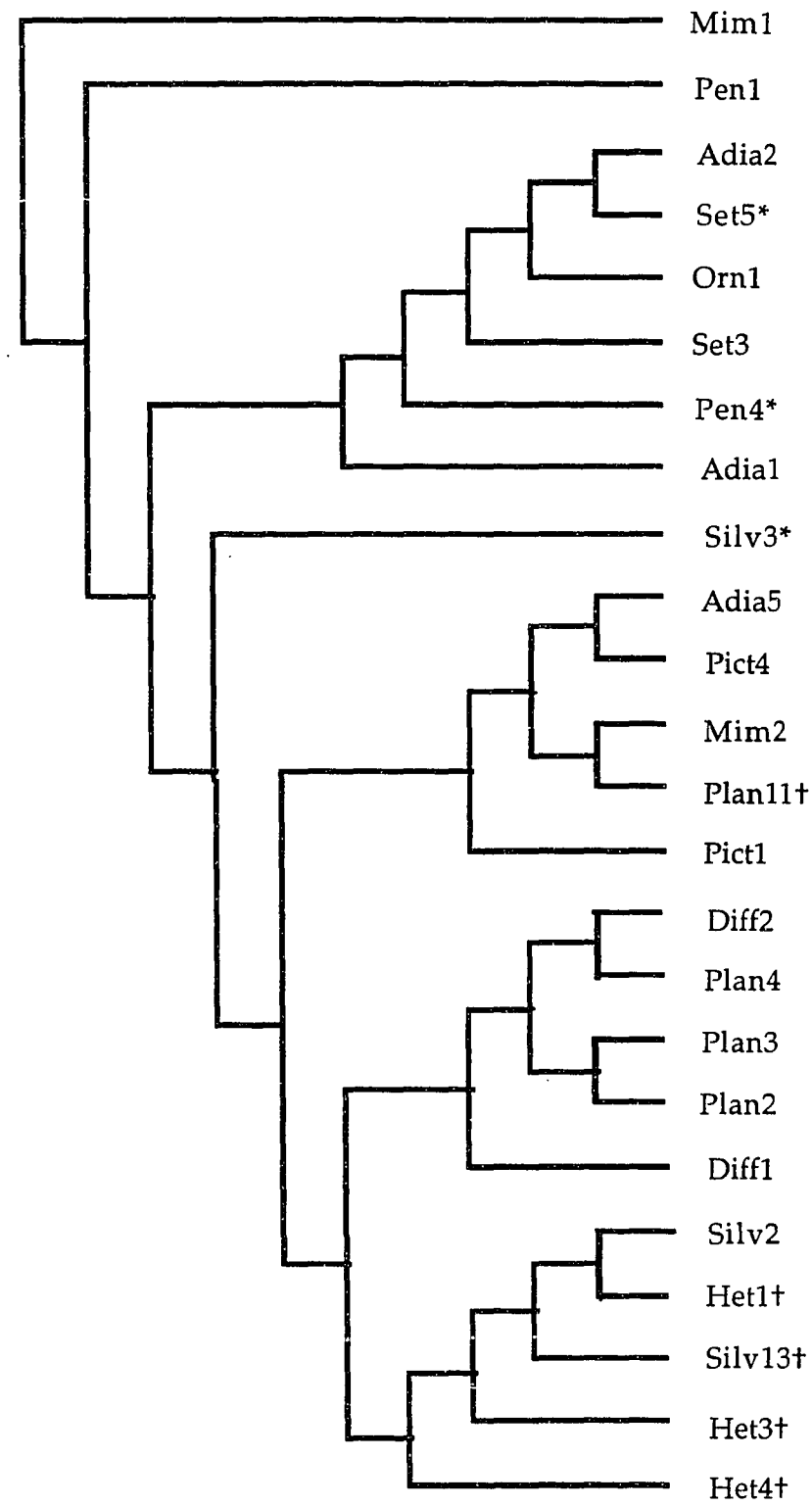

Mim1
Pen1
Adia2
Set5*
Orn1
Set3
Pen4*
Adia1
Silv3*
Adia5
Pict4
Mim2
Plan11†
Pict1
Diff2
Plan4
Plan3
Plan2
Diff1
Silv2
Het1†
Silv13†
Het3†
Het4†


Mim1
Pen1
Adia2
Set5*
Set3
Orn1
Pen4*
Adia1
Silv3*
Adia5
Pict4
Mim2
Plan11†
Pict1
Diff2
Plan4
Plan3
Plan2
Diff1
Silv2
Het1†
Silv13†
Het3†
Het4†


Mim1
Pen1
Adia2
Set5*
Set3
Orn1
Pen4*
Adia1
Pict1
Adia5
Pict4
Mim2
Plan11†
Silv3*
Diff2
Plan4
Plan3
Plan2
Diff1
Silv2
Het1†
Silv13†
Het3†
Het4†


Mim1
Pen1
Adia2
Set5*
Set3
Orn1
Pen4*
Adia1
Silv3*
Adia5
Pict4
Mim2
Plan11†
Pict1
Diff2
Plan4
Plan3
Plan2
Diff1
Silv2
Het1†
Silv13†
Het3†
Het4†


Mim1
Pen1
Adia2
Set5*
Set3
Orn1
Pen4*
Adia1
Silv3*
Adia5
Pict4
Mim2
Plan11†
Pict1
Diff2
Plan4
Plan3
Plan2
Diff1
Silv2
Het1†
Silv13†
Het3†
Het4†


Mim1
Pen1
Adia2
Set5*
Orn1
Set3
Pen4*
Adia1
Silv3*
Adia5
Pict4
Mim2
Plan11†
Pict1
Diff2
Plan4
Plan3
Plan2
Diff1
Silv2
Het1†
Silv13†
Het3†
Het4†

APPENDIX C

Isolation of DNA from Drosophila

Modified from Rivin *et al.,* (1982)

1) Using sterile shilled apparatus and homogenization solution. In 5 to 10 ml grinding buffer from mototized mortear and pestil, use 10 -15 strokes over ice. Strain through sterile gauze. Wash apparature in 5-10 ml buffer.

2) Spin homogenate 10' @ 10,000 rpm.

3) Retain pellet, lyse pellet with lysis buffer @ RT.

4) Allow debris to float to top. Draw off liquid from below. Add 0.1 volume 10% sacosyl. Add 1 g/ml CsCl. Mix gently by inversion.

5) Transfer to ultracentriguge tube.

top with 500 µl 10 mg/ml EtBr.

final top with mineral oil, cap

spin @ 50,000 for 48-60 hours. @ 18-24 °C

6) Using needle, pierce tube below DNA band (illuminated with UV light). Collect DNA band. Wash with equal volume IAA until pink color is gone.

7) Transfer to dialysis tubing. Dialize in a large volume ice cold 10 -4 EDTA. Changing buffer every twelve hours.

8) EtOH ppt.

Grinding Buffer

0.3 M sucrose

50 mM Tris- HCl pH 8.0

5 mM $MgCl_2$

Lysis Buffer

20 mM EDTA

50 mM Tris pH 8.0

1 % sarcosyl

APPENDIX D

Lifton Method for rapid Drosophila DNA isolation (variation)

1) Grind in 5 mls buffer for ~300 flies (~300 mg) (~100 Hawaiian Drosophila)
   2.5 mls buffer for 80-100 flies (~20 Hawaiian Drosophila).
   add DEPC (0.5% v/v) prior to grinding unless DNA is to be cloned
   Homogenize by hand a minimal amount.

2) Gently strain through sterile polyfil (pillow stuffing) in syrings.

3) Add ≥ 200 µg/ml Porteinase K to sample and Heat at 65°C, 1 hr.

4) Add 750 µl 8M KAc to 300 flies/ 375 µl KAc to 80-100 flies
   Incubate on ice at least 1 hour (Okay @ -20°C o/n)

5) Spin 15': 10,000 rpm.

6) Decant supernatent, add 2X volume 95% EtOH at r.t.. Mix and spin immediately.

7) Discard supernatent, resuspend pellet in 500µl TE for 300 flies/ 250µl for 80-100 flies
   (Note, adjuct volume of TE to size of pellet.)

8) Add ~50 µg/ml RNase. Incubate at r.t. 15-30'.

9) Phenol extract (1X vol. buffered phenol, 2 times);

back extract (1X vol TE to first phenol wash, 1 time);

combine aqueous layers from phenol and back extractions;

chloroform extract (1X vol. 24:1 chloroform:IAA, 2 times)

10) Add 1/10 vol 3M NaAc pH 6.0, mix;

2.5 vol ice cold 95% EtOH.

1 hr to o/n at -20°C

11) microfuge 10', resuspend pellet in TE or sddH_2O.

Lifton Grind Buffer:	for 100 mls
0.2 M sucrose	6.8g Sucrose
50 mM EDTA	10 mls 0.5 M EDTA
100 mM Tris pH 9	10 mls 1.0 M Tris pH 9
0.5% SDS	2.5 mls 20% SDS
	sddH_2O to volume

APPENDIX E

DNA isolation from one to ten flies

Modification (combination) of Junakovic/Angelucci and Lifton Methods

1) Homogonize fly (flies) in 100 µl Lifton Grind Buffer in microfuge tube.

2) Spin down debry. Take supernatent add ≥200 µg/ml (i.e. 20 µg) Proteinase K,
   Heat @ 65°C, 1 hr.

3) Add 30 µl 8 M KAc. Incubate on ice 30' to an hour.

4) Centrifuge 10'.

5) Add 2X vol 95% EtOH at r.t. Mix and spin immediately.

6) Resuspend pellet in 50 µl TE.
   Add ~50 µg/ml RNase (1 µl), incubate r.t. 30'to an hour).

7) Phenol extract (1X vol. buffered phenol, 2 times);
   back extract (1X vol TE to first phenol wash, 1 time);
   combine aqueous layers from phenol and back extractions;
   chloroform extract (1X vol. 24:1 chloroform:IAA, 2 times)

8) Add 1/10 vol 3M NaAc pH 6.0, mix;

2.5 vol ice cold 95% EtOH.

1 hr to o/n at -20°C

9) microfuge 10', resuspend pellet in TE or sddH$_2$O.

DNA should be suitable for cloning.

Lifton Grind Buffer:	for 100 mls
0.2 M sucrose	6.8g Sucrose
50 mM EDTA	10 mls 0.5 M EDTA
100 mM Tris pH 9	10 mls 1.0 M Tris pH 9
0.5% SDS	2.5 mls 20% SDS
	sddH$_2$O to volume

APPENDIX F

DIGOXIGENIN- LABELING REACTION

Denature 1 mg linear DNA (10 ng to 3 mg) (10' boiling H_20, 5' on ice).

LS-Buffer for Digoxi.:	100 µl	
	40 µl	1M HEPES pH 6.6
	40 µl	TM salt buffer
	10 µl	primer (Hexanucleotide)
		(Promega calf-thymus primer 7.5 mg/ml)
	10 µl	20X dNTP's

20X dNTP's:		10 µl
	2mM dATP	1 µl 20 mM
	2mM dCTP	1 µl 20 mM
	2mM dGTP	1 µl 20 mM
	1.3mM dTTP	0.65 µl 20 mM
	0.7mM Dig-dUTP	7 µl 1mM/ml

20 µl reaction mix:

10 µl LS

2 µl BSA 1mg/ml nuclease free

7 µl denatured DNA & ddH_20

0.5 µl Klenow (~3 units)

>60 mins. 37° C

precipitate:

to reaction mix add:

5 µl tRNA 25mg/ml

3 µl 3M NaAc

70 µl 95% ETOH

20 mins -70° C

centrifuge 10 mins.

wash pellet w/ 70% ETOH

air dry

resuspend for probing filter (100 µl TE) or in situ to polytene chromosomes (61 µl H_2O).

Filter Digoxigenin Hybrization and Detection

Prehybridization

>1 hr, 60° C in an open container

Solution=	20 ml	50 ml
	1 ml 100X Denhardt's	5 ml 50X
	5 ml 20X SSC	12.5 ml
	11.2 ml ddH2O	29.5 ml
	2.4 ml 0.4 M or 2.0 ml 0.5 M Phosphate Buffer	2.5 ml
1M	200 µl 10% SDS	250 µl 20%
	200 µl ssDNA denatured	500 µl

Hybridization

Digoxi-probe: 10' boil, 5' ice

mix with 5-10 ml prehybe-sol. seal in bag.

min: 20µl/cm^3 filter

max: 1mg DNA/µl, normal 10-50 ng/µl

overnight 60°C H_2O-bath

wash 15' RT 2X SSC 0.1% SDS

2X 30' 60°C 0.2X SSC 0.1% SDS (for genomic blots & cross hybridization: use 0.5X SSC).

2X 5' RT 2X SSC

Blocking Buffer:	500	ml
100 mM Tris-HCl pH 7.5	50	ml 1M
100 mM NaCl	10	ml 5M
3 mM $MgCl_2$	1.5	ml 1M
0.5% Tween 20 (v/v) Sigma	2.5	ml

30 mins RT 25 ml in open dish

Detection (30-1 hr)

30 min RT 5 ml in sealed bag

Vortex Anti-digox.-AP-Complex

dilute 1:5000 = 1 µl complex: 5 ml dil. buffer (buffer 1)

dilution buffer (buffer 1): 100 mM Tris-HCl pH 7.5

150 mM NaCl

(if necessary, add 1% blocking reagent)

wash: 3x 10' RT 200 ml dilution buffer

equilibrate 2X 2' in 20 ml buffer 3: 1L

100mM Tris-HCl pH 9.5	100ml 1M
100mM NaCl	20ml 5M
50mM $MgCl_2$	50ml 1M

color reaction: hours - 1 day at 37°c (IN THE DARK!!!!!)

10 ml fresh prepared:

45 µl NBT (Stock solution)

35 µl X-phosphate (stock solution)

10 ml buffer 3

stop reaction wash 5' in TE

Stock solutions

NBT 75mg/ml in 70% Dimethylformamid (700 µl dimethylformamid + 300µl H_2O, dissolves only with water added)

BCIP 50mg/ml in 100 % dimethylformamid

in situ hybridization to polytene chromosomes with Digoxigenin-labeled probe

Heat pretreatment of slides

30' 2X SSC 65-70°C

10' 70% ETOH 65°C --alcohol containers removed to RT

10' 70% ETOH

2X 10' 95% ETOH

slides are selected at this stage and can be kept

hybridization mix:	4X SSC	
	50% Formamide	
	0.3 mg/µl salmon sperm DNA	
per slide (15µl)	7.5 µl Formamide	(15 slides) 112.5 µl
	3 µl 20X SSC	45 µl
	0.5 µl ssDNA	6.75 µl
	4 µl labeled DNA in ddH2O (~ 75ng DNA)	61 µl

Denature chromosomes with either NaOH or heat, not both.

1) NaOH denaturization

2.5-6 mins. (for silvestris) 0.07 N NaOH RT (0.28 g/100ml ddH_2O)

2X 5' 70% ETOH

2X 5' 95% ETOH

air dry

Hybridization

probe 15µl/slide (boil 10'/ ice 5')

cover with coverslip, seal with rubber cement. o.n. 37°C in moist chamber (on soaked paper towels)

2) Heat denaturization

place probe (15µl) on slide,

cover with coverslip, seal with rubber cement. No need to denature probe.

Let rubber cement dry completely.

Place slide in preheated moist chamber (on soaked paper towels) at 80°C (in a waterbath) for 10 mins. - 1 hr.

remove from 80°C and place at 37°C overnight

washing

2X 10' 2X SSC RT

2X 10' 2X SSC 37°C

5 min 2X SSC RT

3 min 1X PBS + 0.1% triton RT

3X 5 ' 1X PBS RT

(note, use 4XSSC for cross hybridization)

blocking 1X PBS + 0.5% TWEEN 20, 10-30 mins.

detection

add 50µl/slide of diluted complex (1µl complex: 250µl blocking buffer)

cover with coverslip, incubate 30' at 37°C in dark, moist chamber.

wash away coverslips in 2X SSC

3X 5' 2X SSC RT

1X 5' 1X PBS RT

1X 2' in buffer 3

color reaction

add 0.5-1ml color reaction buffer /slide

color reaction buffer = 1 ml buffer 3 pH 9.5

4.5 µl NBT (stock soln.)

3.5 µl BCIP (= x-phosphate)

wash intensively with dH_2O

cover with cover glass and seal with fingernail polish.

APPENDIX G

Hybridization —SDS method

No prehybridization of filter needed!!!!!!

Hybridization solution 0.5M Na Phosphate buffer pH 7.5, 7%SDS

denature probe, boil for 10' quick cool on ice.

$1X10^5$ to $1X10^6$ cpm/ ml

hybridize for 12 hours at 65° C in H_2O bath.

Wash filter

2 times, 20' 2XSSC, 2%SDS 65°C

or

2 times20' 0.5XSSC, 0.5%SDS 65°C higher stringency.

dry and expose

Na- Phosphate buffer pH 6.8

1) basic 0.5M Na_2HPO_4 ~pH 9

2) acidic 0.5M $Na_2H_2PO_4$ ~pH 4.5

add 2 into 1 until pH 7.5

APPENDIX H

Phage miniprep

The Aloha Method

1) Grow phage on small petri dishes to confluence.

2) Add 3-5 ml SM buffer. Leave on for at least 1 hour or o/n (yield is higher if SM is on o/n) Suck off SM with pipet.

3) Transfer 1 ml supernatent (SN) to microfuge tube.
   spin down debry 5'.

4) Take 800 µl SN, and 600 µl DEAE 52 cellulose. invert tube 30-50 times.
   spin 5'.

5) Take 800 µl SN, add 150 µl phage lysis buffer.
   incubate 15' at 67°C.

6) add 200 µl 8M ammonium acetate (or KAC or NaAC)
   incubate 15' on ice.

7) spin 5'
   take 900µl SN, add 600µl isopropanol, 15' r.t.

8) spin 10'. discard SN (pellet is mostly invisiable on very loose).

9) wash pellet 2 times in 500 µl 70% ice cold EtOH. (spin ~ 2' after each wash)

10) Dry pellet. Dissolve pellet in TE of sddH_2O. usually 30-60 µl depending on pellet size. Usually enough DNA for 2-6 digests

Treatment of DEAE 52 cellulose

put DEAE 52 cellulose in 0.05 N HCl (2.5 ml conc. HCl to 500 ml).
With gently stirring add NaOH till pH approaches 7. Let resin settle and decant. Add LB, let resin settle and decant. Change LB several times until pH is ~7. resuspend finally to ~75% resin/25% LB. Keep frozen.

Phage Lysis Buffer
0.25 M EDTA
0.5 M Tris pH 8.5
2.5% SDS

APPENDIX I

Phage Miniprep

1) Phage suspension: 1 plaque in 500 µl phage buffer.

2) Bacteria culture in NCZYM, overnight.

3) 25-60 µl phage suspension (depending on titre), 25-30 µl bacteria culture.
20' 37°C

4) Add to 12 ml NCZYM. 4hr to overnight 37°C, 200 rpm

5) When lysis occurs add 200 µl chloroform and shake for another 10'.

6) Centrifuge 5', 10,000 rpm.

7) 10 ml Supernatent (SN), 50 µl DNase, 50 µl Rnase (both 10 mg.ml) Incubate 45' in 37°C, between 4°C o/n.

8) Add 2 ml phage lysis buffer, prewarmed to 70°C. Incubate 30', 70°C.

9) add 2.5 ml 8 M KAc, 15-30' on ice. Centrifuge 20', 13,000 rpm, 4°C. (pellet is SDS and KAc)

10) Take SN, add 8 ml Isopropanol., 10' r.t., centrifuge 10', 10,000 rpm. air dry. pellet should be invisiable.

11) Add 0.3 M NaAc to pellet. Phenol extract (1 vol buffered phenol 2 times, 1 volume 24:1 chloromoform:IAA, 2 times).

12) Take aqueous layer and add 300 µl isopropanol. 10; r.t. 5' centrifuge.

13) Wash 2X in 70% EtOH, dry.

14) Resuspend pellet in TE of sddH$_2$O.

Phage Buffer

10 mM $MgCl_2$.

20 mM NaCl

10 mM Tris pH 7.5

Phage Lysis Buffer

0.25 M EDTA

0.5 M Tris pH 8.5

2.5% SDS

APPENDIX J

Transformation

Making competitent cells.

1) Select a single colony of DH5a, grow in sterile 2XYT overnight (o/n) 37°C.

2) Inocculate 500 ml 2XYT with 500 µl overnight culture. Shake at 37°C at 2,000 rpm. about 5-6 hours (until $OD_{600} = 0.5 \approx 5 \times 10^8$ cells/ml).

3) Spin down cells in sterile tubes.

4) Resuspend in icecold, sterile 50 mM $CaCl_2$, 1/2 original volume. There should be no clumping of cells.

5) Leave at 4°C for ~12 hours (Don't bring cultures above 4°C and don't exceed time too much).

6) Pellet cells next morning. Resuspend in 25 ml (1/20 volume of original culture) 50 mM $CaCl_2$. Leave overnight at 4°C.

7) Add 2.5 ml sterile glycerol (1/200 volume of original culture). Aliquot and freeze at -70°C.

Transforming

8) Thaw cells at RT, then leave 5′ on ice.

9) Add 5-20 ng DNA (not more than 20 µl) to 50 µl cell suspension.

10) Leave on ice for at least 30′, not exceeding 1 hour.

11) heatshock for 2′-5′ at 42°C.

12) Leave on ice for 5′.

13) Add 200 µl 2XYT.

14) incubate 1-1.5 hour at 37°C.

15) Plate between 50 µl and whole culture on selective plates (usually Amp 50 µg/ml).

If doing subcloning, spread X-gal (40 µl [20mg/ml]) and
IPTG (4 µl[200mg/ml]) on plates at least 20′ before plating cells.

16) Incubate o/n at 37°C, up-side down.

APPENDIX K

RAPID BOILING PLASMID PREP

1) 2 ml 2xYT (or other media) + Amp (50 µg/ml).

2) pick colony w/ toothpick, throw toothpick into media. Shake overnight.
   37° C.

3) Fill eppendorf tube (~1.8 ml) Microfuge 15-30 seconds.

4) Remove Supernatant. Use kimwipes to remove droplets.

(It would be a very good idea to startboiling the water about now!!!!)

5) Add 300 µl STET. suspend pellet using pipet-men.

6) Add 15 µl lysozyme (freshly made 20mg/ml 0.01M Tris pH 8.0). Vortex
   gently
   On ice 5' to 10'
   40 seconds. in boiling water.
   microfuge 15 minutes+. (RT)

7) Pull out snot with toothpick, add 300µl isopropanol, let sit 10 minutes. RT.
   centrifuge 10' cold, discard supernatant

8) Wash pellet in 70 % ETOH. Centrifuge 2'.

9) Remove liquid by placing in vacuum 8-10 minutes.

10) Suspend pellet in 50 µl H_2O or TE. 1.5-2.0 µl/ digest.

STET

8% sucrose

50 mM Tris pH 8.0

50 mM EDTA

5% Triton X100

Sterile Filter Sterilize

LITERATURE CITED

Bennett, M.D. (1982) The neucleotypic basis of the spatial ordering of chromosomes in eukaryotes and the implications of the order for genome evolution and phenotypic variation. In G.A. Dover and R.B. Flavell (Eds), *Genome Evolution*, Academic Press, London, pp. 239-260.

Bishop, J.G. III, and J.A. Hunt (1988) DNA divergence in and around the alcohol dehydrogenase locus in five closely related species of Hawaiian *Drosophila*. *Molecular Biology and Evolution* 5:415-431.

Brezinsky, L., G.V.L. Wang, T. Humphreys and J. Hunt (1990) The transposable element *Uhu* from Hawaiian *Drosophila* -- member of the widely dispersed class of Tc1-like transposons. *Nucleic Acid Research* 18:2053-2059.

Brezinsky, L., T.D. Humphreys and J.A. Hunt (1993) Evolution of the transposable element *Uhu* in five species of Hawaiian *Drosophila*. *Genetica* 86:21-36.

Brierley, H.L. and S.S. Potter (1985) Distinct characteristics of loop sequences of two *Drosophila* foldback transposable elements. *Nucleic Acids Research* 13:485-500.

Britten, R.J. and E.H. Davidson (1969) Gene regulation for higher cells: A theory. *Science* 165:349-357.

Britten, R.J. and E.H. Davidson (1971) Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *Quarterly Review of Biology* 46:111-138.

Brookfield, J.F.Y. (1986) A model for DNA sequence evolution within transposable element families. *Genetics* 112:191-407.

Capy, P., J.R. David and D.L. Hartl (1993) Evolution of the transposable element *mariner* in the *Drosophila melanogaster* species group. *Genetica* 86:37-46.

Carson, H.L (1976) Inference of the time of origin of some *Drosophila* species. *Nature* 259:395-396.

Carson, H.L. (1983) Chromosomal sequences and interisland colonizations in Hawaiian *Drosophila*. *Genetics* 103:465-482.

Carson, H.L. (1987) Tracing ancestry with chromosomal sequences. *Trends in Ecology and Evolution* 2:203-206.

Carson, H.L., D.E. Hardy, H.T. Spieth, and W.S. Stone (1970) The evolutionary biology of the Hawaiian Drosophilidae. In Hecht, M.K. and W.C. Steere (eds.) *Essays in Evolution and Genetics in Honor of Th. Dobzhansky*. pp. 437-543. Appleton-Century-Crofts, New York

Carson, H.L., J.P. Lockwood and E.M. Craddock (1990) Extinction and recolonization of local populations on a growing shield volcano. *Proc. Natl. Acad. Sci USA* 87:7055-7057.

Chao, L., C. Vargas, B.B. Spear and E.C. Cox (1983) Transposable elements as mutator genes in evolution. *Nature* 303:633-635.

Charlesworth, B. (1986) Genetic divergence between transposable elements. *Genet. Res. Camb.* 48:111-118.

Charlesworth, B. (1988) The maintenance of transposable elements in natural populations. Biol. Sci. 189-212.

Charlesworth, B. and C.H. Langley (1989) The population genetics of *Drosophila* transposable elements. *Annual Review of Genetics* 23:251-287.

Charlesworth, B., A. Lapid and Canada D. (1992) The distribution of transposable elements with and between chromosomes in a population of *Drosophila melanogaster*. I. Element frequencies and distribution. *Genetical Research* 60:103-114.

Chow, C.S. and J.K. Barton (1992) Transition metal complexes as probes of nucleic acids. *Methods in Enzymology*. 212:219-242.

Church, G. M. and W. Gilbert (1984) Genomic Sequencing. *Proc. Natl. Acad. Sci USA* 81:1991-1995

Ciazzi, R., C. Caggese and S. Pimpinelli (1993) *Bari-1*, a new transposon -like family in *Drosophila melanogaster* with a unique heterochromatic organization. Genetics 133:315-334.

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, 726-727.

Craddock, E.M. and H.L. Carson (1989) Chromosomal inversion patterning and population differentiation in a young insular species, *Drosophila silvestris*. *Proc. Natl. Acad. Sci. USA* 88:4798-4802.

Cullis, C.A. (1988) Control of variation in higher plants. in *Evolutionary process and metaphors*. (eds) Ho, M.-W. and S.W. Fox pp. 49-61. John Wiley and Sons, Ltd.

DeSalle, R., & Giddings, L. V. (1986). Discordance of nuclear and mitochondria DNA phylogenies in Hawaiian Drosophila. *Proc. Natl. Acad. Sci. USA*, 83, 6902-6906.

Di Franco, C., D. Galuppi and N. Junakovic (1993) Genomic distribution of transposable element among individual of an inbred *Drosophila* line. *Genetica* 86:1-12.

Doolittle, W.F. (1985) The evolutionary significance of middle-repetitive DNAs. in *Evolution of Genome Size*. (ed) Cavalier-Smith, T. pp. 443-487. John Wiley and Sons, Ltd

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

Eanes, W.F., C. Wesley, J. Hey, D. Houle and J.W. Ajioka (1988) The fitness consequences of *P* element insertion in *Drosophila melanogaster*. *Genet. Res. Camb.* 52:17-26.

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

Engels, W.R., D.M. Johnson-Schlitz, W.B. Eggleston and J. Sved (1990) High-frequency P element loss in *Drosophila* is homolog dependent. *Cell* 62:515-525.

Felger, L.I. (1988) Organisation und chromosomale verteilung von mittel-repetitiver DNA bei zwei *Drosophila-arten: D. subobscura* und *D. silvestris*. Dissertation zur Erlangung des Grades eines Dokors der Naturwissenschaften, der Fakultat fur Biologie der Eberhard-Karls-Uneversitat Tubingen.

Felger, I. and J.A. Hunt (1992) A non-LTR retrotransposon from the Hawaiian *Drosophila:* The LOA element. *Genetica* 85:119-130.

Felsenstein, J. (1993) PHYLIP (Phylogeny inference Package) Version 3.5c University of Washington, Seattle.

Finnegan, D.J. (1985) Transposable elements in eukaryotes. *Int. Rev. Cytol.* 93:281-326.

Finnegan, D.J. and D. H. Fawcett (1986) Transposable elements in *Drosophila melanogaster. Oxford Survey of Eukaryotic Genes* 3:1-62 and appendix.

Franz, G. and C. Savakis (1991) Minos, a new transposon-like element from *Drosophila hydei,* is a member of the Tc1-like family of transposons. *Nucleic Acids Research* 19:6646.

Garrell, J. and J. Modolell (1990) The Drosophila extramacrochaetae locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop-helix protein. *Cell* 61:39-48.

Hardy, D. E. and K. Y. Kaneshiro (1981) Drosophilidae of Pacific Oceania. in *The Genetics and biology of* Drosophila. M Ashburner, H.L. Carson and J.N. Thompson, Jr. (eds) Academic Press, London pp 309-348.

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. Nucleic Acids Research* 16:5991-5998.

Harris. L.J., S. Prasad and A.M. Rose (1990) Isolation and sequence analysis of *Caenorhabditis briggsae* repetitive elements related to the *Caenorhabditis elegans* transposon Tc1. *Journal of Molecular Evolution* 30:359-369.

Higgins, D.G. and P.M. Sharp (1988) Clustal: a package fro performing multiple sequence alignment on a microcomputer. Gene 73:351-359.

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. PNAS* 81:7146-7150.

Hunt, J.A., K.A. Houtchens, L. Brezinsky, F. Shadravan, and J.G. Bishop III. (1989) Genomic DNA variation within and between closely related species of Hawaiian *Drosophila*. In Giddings, L.V., K.Y. Kaneshiro, and W.W. Anderson (eds). *Genetics, Speciation and the Founder Principle.* pp.167-180. Oxford University Press. New York, NY

Hunt, J.A., T.J. Hall and R.J. Britten (1981) Evolutionary distances in Hawaiian *Drosophila* measured by DNA reassociation. *J. Mol. Evol.* 17:361-367.

Junakovic, N. and V. Angelucci (1986) Polymorphisms in the genomic distribution of *copia-like* elements in related laboratory stocks of *Drosophila melanogaster. J. Mol. Evol.* 24:83-88.

Junakovic, N., C.D. Franco, P. Barsanti, and G. Palumbo (1986) Transposition of *copia-like* nomadic elements can be induced by heat shock. *J. Mol. Evol.* 24:89-93.

Kam, M. (1978) The biosystematics of the *mimica* subgroup of the "modified mouthparts" species group of Hawaiian *Drosophila.* M.S. Thesis, Univ. Hawaii, Honolulu, 62pp.

Kaneshiro, K.Y. (1976) Ethological isolation and phylogeny in the plantibia subgroup of Hawaiian *Drosophila.* Evolution 30:740-745.

Kidwell, M.G. (1993) Horizontal transfer of *P* elements and other short inverted repeat transposons. *Genetica* 86:275-286

Kidwell, M. G. and J.B. Novy (1979) Hybrid dysgenesis in *Drosophila melanogaster*: Sterility resulting from gonadal dysgenesis in the P-M system. *Genetics* 92:1127-1140.

Kidwell, M.G., J.F. Kidwell and J.A. Sved (1977) Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. *Genetics* 86:813-833.

Kimura, M. (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* 16:111-120.

Kirshenbaum, M.R., R. Tribolet and J.K. Barton (1988) $Rh(DIP)_3^{3+}$: A shape-selective metal complex which targets cruciforms. *Nucleic Acids Research* 16:7943-7960.

Krause, M., J. Harwood, J. Fierer and D. Guiney (1991) Genetic analysis of the homology between the virulence plasmids of *Salmonella dublin* and *Yersinia pseudotuberculosis*. *Infect. Immun.* 59:1860-1863.

Kunze, R. and P. Starlinger (1989) The putative transposase transposable element *Ac* from *Zea mays* L. interacts with subterminal sequences of *Ac*. *EMBO Journal* 8:3177-3185.

Lansman, R.A., R.O. Shade, T. A. Grigliatti and H.W. Brock (1987) Evolution of P transposable elements: Sequences of *Drosophila nebulosa* P elements. *PNAS USA* 84:6491-6495.

Lee, I. and J.K. Barton (1993) A distinct intron-DNA structure in Simian Virus 40 T-antigen and Adenovirus 2 E1A genes. *Biochemistry* 32:6121-6127.

Lewin, B. (1990) *Genes IV*. Oxford University Press, Oxford, U.K.

Li, W. -H., C. -I. Wu and C. -C. Luo (1985) A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Molecular Biology and Evolution* 2:150-174.

Lim, J.K. (1988) Interchromosomal rearrangements mediated by *hobo* transoposons in *Drosophila melanogaster*. PNAS 85:9153-9157.

Luo, D., E.S. Coen, S. Doyle and R. Carpenter (1991) Pigmentation mutants produced by transposon-mutagenesis in *Antirrhinum majus*. *Plant Journal* 1:59-69

Lyttle, T.W. and D.S. Haymer (1993) The role of the transposable element *hobo* in the origin of endemic inversions in wild population of *Drosophila melanogaster*. *Genetica* 86:113-126.

MacKay, T. F. C. (1985) Transposable element-induced response to artificial selection in *Drosophila melanogaster*. *Genetics* 111:351-374.

Manning, J.E., C.W. Schmid, and N. Davidson (1975) Interspersion of repetitive and nonrepetitive DNA sequences in the *Drosophila melanogaster* genome. *Cell* 4:141-155.

Manuelidis, L. (1982) Repeated DNA sequences and nuclear structure. In G.A. Dover and R.B. Flavell (Eds), *Genome Evolution*, Academic Press, London, pp. 263-285.

Marck, C. (1986) Fast analysis of DNA and protein sequence on Apple IIe: restriction sites search, alignments of short sequences and dot matrix analysis. *Nucleic Acids Research* 14:583-590.

Martin, G., D. Wiernasz, P. Schedl (1983) Evolution of *Drosophila* repetitive-dispersed DNA. *J. Mol Evol* 19:203-213.

McClintock, B. (1950) Mutable loci in maize. *Carnegie Inst. of Wash. Year Book* 49:157-167.

McDonald J.F. (1989) The potential evolutionary significance of retroviral-like transposable elements in peripheral populations. In Fontdevila, A. (ed.) *Genetics of Transient Unstable Populations.* Springer-Verlag Berlin pp. 190-205.

McDougall, I. (1979) Age of the shield-building volcanism of Kauai and linear migration of volcanism on the Hawaiian Island chain. *Earth Planet. Sci.* Lett. 46:31-42.

Medhora, M., K. Maruyama and D.L. Hartl (1991) Molecular and functional analysis of the *mariner* mutator element *Mos1* in Drosophila. *Genetics* 128:311-318.

Misra, S. and D.C. Rio (1990) Cytotype control of Drosophila P element transposition: The 66 kd protein is a repressor of transposase activity. *Cell* 62:269-284.

Montchamp-Moreau, C., S. Ronsseray, M. Jacques, M. Lehmann and D. Anxolabéhère (1993) Distribution and conservaton of sequences homologous to the *1731* retrotransposon in *Drosophila. Molecular Biology and Evolution.* 10:791-803.

Ohta, T. (1985) A model of duplicative transposition and gene conversion for repetitive DNA families. *Genetics* 110:513-524.

Orgel, L.E. and F.H.C. Crick (1980) Selfish DNA: the ultimate parasite. *Nature* 284:604-607.

Paquin, C.E, and V.M. Williamson (1984) Temperature effects on the rate of Ty transposition. *Science* 236:53-55.

Perlman, J. (1983) One side of a deletion breakpoint from the *Drosophila melanogaster* genome contains a transposable element. *Gene* 21:87-94.

Perna, N.T., M.A. Batzer and P.L. Deininger (1992) Alu insertion polymorphism: A new type of marker for human population studies. *Human Biology* 64:641-648.

Plasterk, R.H.A. and J.R.M. Groenen (1992) Targeted alterations of the *Caenorhabditis elegans* genome by transgene instructed DNA double strand break repair following Tc1 excision. *EMBO Journal* 11:287-290.

Pyle, A.M. and J.K. Barton (1990) Probing nucleic acid with transition metal complexes. *Progress in Inorganic Chemistry* 38:413-475.

Rhodes, P.R. and L.O. Vodkin (1985) Highly structured sequence homology between an insertion element and the gene in which it resides. *Proc. Natl. Acad. Sci. USA* 82:493-497.

Rivin, C.J., E.A. Zimmer and B. Walbot (1982) Isolation of DNA and DNA recombinates from Maize. in W.F. Sheridan (ed) *Maize for Biological Research.* Plant Molecular Biology Association. Charlottesville, VA.

Robertson, H.M. (1993) The *mariner* transposable element is widespread in insects. *Nature* 362:241-245

Rogers, J.H. (1985) The origin and evolution of retroposons. *Int. Rev. Cytol.* 93:188-280.

Rowan, R.G. and J.A. Hunt (1991) Rates of DNA change and phylogeny from the DNA sequences of the Alcohol Dehydrogenase gene for five closely related species of Hawaiian *Drosophila. Molecular Biology and Evolution* 8:49-70.

Rubin, G.R. (1983) Dispersed Repetitive DNAs in *Drosophila.* In Shapiro, J.A. (ed) *Mobile Genetic Elements.* Academic Press, New York, NY

Sambrook, J., E.F. Fritsch, T. Maniatis (1989) *Molecular Cloning, A Laboratory Manual, 2nd. ed.* Cold Spring Harbor Laboratory Press.

Sanger, F., S. Nicklen and A.R. Coulson (1977) DNA sequencing with chain-terminating inhibitors. *PNAS USA* 74:5463-5467.

Sankaranarayanan, K. (1988) Mobile genetic elements, spontaneous mutations and the assessment of genetic radiation hazards in man. In Lambert, M.E., J.F. McDonald and I.B. Weinstein (eds.) *Eukaryotic Transposable Elements as Mutagenic Agents.* Cold Spring Harbor Press, Cold Spring Harbor, NY.

Schachat, F.H. and D.S. Hogness (1974) Repetitive sequences in isolated Thomas circles from *Drosophila melanogaster. Cold Spring Harbor Symp. Quar. Biol.* 38:371-381.

Schukkink, R., & R.H.A., Plasterk (1990) TcA, the putative transposase of the C. elegans Tc1 transposon, has an N-terminal DNA binding domain. *Nucleic-Acids-Res,* 18, 895-900.

Schulze, D.H. and C.S. Lee (1986) DNA sequence comparison among closely related *Drosophila* species in the Mulleri complex. *Genetics* 113:287-303.

Schwarz-Sommer, Z. and H. Saedler (1987) Can plant transposable elements generate novel regulatory systems? Mol. Gen. Genet. 209:207-209.

Silber, J., C. Bazin, F. Lemeunier, S. Aulard, and M. Volovitch (1989) Distribution and conservaton of the foldback transposable element in *Drosophila. Journal of Molecular Evolution* 28:220-224.

Slatkin, M. (1985) Genetic differentiation of transposable elements under mutation and unbiased gene conversion. *Genetics* 110:154-158.

Sprandling, A.C. and G.R. Rubin (1981) *Drosophila* genome organization: Conserved and dynamic aspects. *Ann. Rev. Genet.* 15:219-264.

Stamatis, N., M. Monastirioti, G. Yannopoulos and C. Louis (1989) The P-M and the 23.5 *MRF (hobo)* systems of hybrid dysgenesis in *Drosophila melanogaster* are independent of each other. *Genetics* 1233:379-387.

Strand, D.J. and J.F McDonald (1989) Insertion of a *copia* element 5' to the *D. melanogaster* alcohol dehydrogenase gene (*Adh*) is associated with altered developmental and tissue-specific patterns of expression. *Genetics* 121:787-794.

Strand, D.J. and J.F. McDonald (1985) *Copia* is transcriptionally responsive to environmental stress. *Nucleic Acids Research* 13:4401-4410.

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

Syvanen, M. (1984) The evolutionary implications of mobile genetic elements. *Ann. Rev. Genet.* 18:271-293.

Templeton, A.R. (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of humans and apes. *Evolution* 37:221-244.

Thomas, R. and J.A. Hunt (1991) Molecular evolution of alcohol dehydrogenase locus and the phylogeny of Hawaiian *Drosophila*. *Molecular Biology and Evolution* 8:687-702.

Throckmorton, L.H. (1975) The phylogeny, ecology, and geography of *Drosophila*. in *Handbook of Genetics, volume 3*. R.C. King (ed) Plenum Press, New York and London. pp. 421-470.

Titus, D.E. (1991) *Promega Protocols and Application Guide. 2nd Edition.* Promega Corporation. Madison WI

Triantaphyllidis, C.D. and R.H. Richardson (1980) The DNA sequences of six endemic Hawaiian *Drosophila* species. *Sci. Annals, Fac. Phys. and Methem. Univ. Thessaloniki* 20a:93-113.

Tudor, M., M. Lobocka, M. Goodell, J. Pettitt and K. O'Hare (1992) The pogo transposable element family of *Drosophila melanogaster*. *Molecular and General Genetics* 232:126-134.

Vanlerberghe, F., F. Bonhomme, C.A. Hutchison III and M.H. Edgell (1993) A major difference between the divergence patterns within the Lines-1 families in mice and voles. *Molecular Biology and Evolution* 10:719-731.

von Sternberg., R.M., G.E. Novic, G. -P. Gao and R.J. Herrera (1993) Genome canalization: the coevolution of transposable and interspersed repetitive elements with single copy DNA. *Genetica* 86:215-246.

Wensink, P.C., S. Tabata, C. Pachl (1979) The clustered and scrambled arrangement of moderately repetitive elements in *Drosophila* DNA. *Cell* 18:1231-1246.

Wilson, J.T. (1963) A possible origin for the Hawaiian Islands. *Canadian Journal of Physics* 41:135-138.

Yoon, J.S., D. Resch, and M.R. Wheeler (1972) Cytogenetic relationships in Hawaiian species of *Drosophila.* II. the *Drosophila mimica* subgroup of the "modified mouthparts" species group. Studies in Genetics VII Univ. of Texas Publ. 7213:201-212.

Young, M.W. and H.E. Schwartz (1980) Nomadic gene families in *Drosophila. Cold Spring Harbor Symp Quart Biol.* 45:629-640.