THE DISTRIBUTION AND PHYLOGENY OF *HET-A* TRANSPOSABLE ELEMENTS WITHIN SELECTED SPECIES OF *DROSOPHILA*

Melissa J. Bentley
WSU Zoology Thesis
Spring Semester 2002

THE DISTRIBUTION AND PHYLOGENY OF *HET-A* TRANSPOSABLE ELEMENTS WITHIN SELECTED SPECIES OF *DROSOPHILA*

Melissa J. Bentley
WSU Zoology Thesis
Spring Semester 2002

WEBER STATE UNIVERSITY

We hereby recommend that the thesis prepared under our supervision by Melissa J. Bentley entitled "The Distribution and Phylogeny of *Het-A* Transposable Elements Within Selected Species of *Drosophila*" be accepted as fulfilling partial requirements for the thesis in Zoology.

Thesis Committee

Advisor

Department Chair

Abstract

The ends of eukaryotic chromosomes are capped by telomeres. These structures serve two major functions: they prevent the progressive shortening of chromosomes due to incomplete DNA replication and they enable the cell to distinguish a legitimate chromosome end from DNA lesion. Investigations of the structure of telomeres is one of the more fascinating areas of current genetic research because of the association of telomeres with aging and cancer. Most eukaryotic telomeres are characterized by tandemly repeated, short DNA sequences, which are added to chromosome ends by the enzyme telomerase. In contrast, neither the simple tandem repeats nor the enzyme telomerase have been identified in *Drosophila melanogaster*, a model genetic organism. Instead, *Drosophila* relies on a system of telomere-specific transposable elements (TEs) for telomere maintenance. The best characterized telomere TE is *HeT-A*, which comprises 6000 bp and includes both coding and non-coding regions.

The purpose of this study was to identify *HeT-A* gene sequences in several species of *Drosophila*. Using PCR, *HeT-A* was identified in four species of *Drosophila*, three are from the *melanogaster* species subgroup and one from the *suzukii* subgroup. This represents the first time that *HeT-A* was described in three of these species. *HeT-A* elements from each species were sequenced and the phylogeny examined using parsimony analysis. Although much is known about the molecular structure of *HeT-A* in *D. melanogaster*, this is the first phylogenetic analysis of *HeT-A* in *Drosophila*. Evidence for different subfamilies of *HeT-A* elements within the same genome was revealed. These results suggest that, as is the case for most TEs, the evolutionary history of *HeT-A* may be complex.

Introduction

Eukaryotic DNA is organized into linear chromosomes, the ends of which are capped by telomeres. The existence of telomeres was initially proposed to explain the absence of terminal deletions among induced rearrangement in the offspring of irradiated *Drosophila* (Mason, 1995). It is now known that telomeres perform at least two vital functions. First, they prevent end-to-end chromosome fusions and attack of chromosome ends by degradative enzymes. The prevention of these processes is crucial because lesions in DNA are subject to repair to avoid cell cycle arrest. Telomere-capped chromosomes are not subject to repair, nor do they interrupt the normal cell cycle. Second, telomeres prevent the loss of DNA which would be expected to occur over time due to the inability of DNA polymerase to completely replicate linear chromosome ends (Levis, 1993).

The uniqueness of *Drosophila* telomeres in itself makes them a tantalizing subject of study. However, interest in telomeres continues to grow as new studies implicate them in cellular processes, particularly senescence and cancer. Recent work by Karlseder, Smorgorzewska, and de Lange (2002) involved isolation of a protein, TRF2, which binds to telomeric DNA. Telomeres progressively shorten with replication. It appears that the TRF2 protein normally provides a protective function at the telomere, but is no longer able to do so when telomeres

reach a critically short length.

Beginning in the 1990s, a link was made between the enzyme telomerase and cancer. This increased interest in telomeres greatly, and research involving links between cancer and telomerase has continued. Telomerase activity has been detected in up to 90% of human cancers. It has also been demonstrated that tumor cell growth can be inhibited by blocking telomerase activity (Reported by Marx, 2002).

In most eukaryotic organisms, telomeric DNA is characterized by short, tandemly repeated sequences, which are evolutionarily conserved (Mason, 1995). These sequences are typically G-C rich and between two and eight nucleotides in length (Beissmann, 1992). For example, human telomeres consist of the sequence $(TTAGGG)_n$ (Henning, 1998). The block of tandem terminal repeats at the telomere may vary in length from hundreds to thousands of nucleotides. Telomere length is thought to be maintained in part by the reverse transcriptase telomerase. This enzyme has two key components: 1) an RNA template and 2) a reverse transcriptase that reads the template RNA and adds complementary DNA to the chromosome ends. The enzyme telomerase has been isolated in cell cultures of mice, humans, and protozoa. (Levis, 1993) and its activity has been correlated with cell proliferation, immortality, and division (Greider, 1998). It is thought that telomerase is responsible for lengthening chromosomes, or maintaining chromosome length in replicating cells, through the addition of tandemly repeated telomere sequences.

The telomeres of *Drosophila* differ from those of most eukaryotes because they lack the tandemly repeated DNA. In addition, telemorase has not been detected in Drosophila cells (Labrodor and Corces, 1997). Instead of repeated telomeric sequences, *Drosophila*

chromosomes appear to be maintained by at least two different groups of transposable elements, termed *HeT-A* and *TART*. Approximately 10% of the genome of *Drosophila melanogaster* is composed of transposable elements, and 30-50 different families of elements have been identified within this species. Because of their mobility, most TEs are scattered throughout the genome. However, both *HeT-A* and *TART* are clustered at the heterochromatic telomeric regions and apparently are confined to this location (Pimpinelli, 1994).

Transposable elements (TEs) are found throughout eukaryotic genomes and are of two main types. Class II elements move via a transposase that is responsible for cleaving the TE out of the chromosome and reinserting it at another location within the genome. Most class II elements are flanked by inverted repeat sequences, that are necessary for excision and insertion. (Nakatsu 1991). Class I elements, called retrotransposable elements or retrotransposons, require an RNA intermediate for movement. Class I elements contain sequences with significant homology to genes encoding reverse transcriptase. It is known that infection and transformation of cells by RNA viruses requires reverse transcriptase activity, and it is assumed that a similar enzyme is needed for the mobile RNA transcript of a class I transposable element to reinsert into the genome as DNA.

Transposable elements often act as mutagens and typically have a negative effect on the host organism. An active TE may insert within the coding region of a gene, resulting in the production of a mutant polypeptide. In addition, a TE may insert in the regulatory region of a gene, thereby altering gene expression. However, in the case of *HeT-A*, its confinement to heterochromatic regions of the chromosome precludes the possibility for significant gene interruption. It has instead been proposed that the *HeT-A* element is beneficial and, in fact,

essential to its host due to its role in telomere maintenance.

The *HeT-A* element in *Drosophila* comprises approximately 6000 kb (Danilevskaya et al., 1994) (Fig.1). The elements characterized thus far are found in head-to-tail arrangement in telomeric heterochromatin (Fig. 2). They contain repeated patterns of A-rich sequences and nearly half of the element consists of non-coding sequences, primarily in the 3' region. A shorter untranslated region is found at the 5' end of the gene. These 3' non-coding sequences have also been found unassociated with *HeT-A* at telomeric regions of *Drosophila* chromosomes. An unusual feature of the *HeT-A* genes is that the promoter is found in the 3' region of the gene. This promoter drives the transcription of the adjacent *HeT-A* gene, rather than its own transcription. Transcription of a *HeT-A* element in *D. melanogaster* begins at a promoter sequences located at either -62 or -31 bp within the 3' region of the element immediately upstream of the element being transcribed (Danilevskaya, 1998).

The coding region of the *HeT-A* element consists of two open reading frames (ORFs). ORF1 encodes a protein which resembles the Gag protein of retro viruses. This protein is known to perform several functions in viruses. It assists in assembly of virus particles, helps incorporate viral genomic RNA into virus particles, and is involved in uncoating of the viral core to facilitate reverse transcription (Srivastava, 1999). Although ORF 1 is expressed within *Drosophila* cells, it is not yet known what role this protein plays in *HeT-A* mobility. Recent work by Rashkova, Karam, and Pardue (2002) showed that *HeT-A* Gag proteins are transported efficiently to the nucleus. Their intranuclear location and number was consistent with that expected for proteins progressing to final localization at the telomeres. It was also noted that all *HeT-A* elements that had transposed successfully had maintained the ability to produce an intact Gag protein, which

suggests this protein is in some way necessary for the conversion of *HeT-A* RNA into the chromosomal DNA.

The sequence of ORF2 resembles that of genes encoding reverse transcriptase, but a true reverse transcriptase domain has not been detected in any *HeT-A* element (Sheen, 1994).

Evidence indicates that *HeT-A* elements transpose within the genome through a polyadenylated RNA intermediate. Danilevskaya et al.. (1994) isolated such a poly-A RNA intermediate from *Drosophila melanogaster*. As previously noted, *HeT-A* elements do not encode reverse transcriptase. Because of this, it has been proposed that these elements evolved from telomerase and that their mechanism of transposition relies upon the cellular gene for the catalytic subunit of telomerase, which is related to reverse transcriptase of retrotransposons. The 3' and 5' noncoding regions of *HeT-A* appear to be involved in binding and localization at the telomere (Danilevskaya et al., 1998).

This mechanism of telomere-maintenance through *HeT-A* transposition in *Drosophila* is unusual, raising several questions: Why is *Drosophila* unique among organisms examined in utilization of this TE system? Where did these transposable elements come from? What is the mechanism by which *HeT-A* is confined to the ends of chromosomes? How did these TEs come to supplant the tandem repeat system that characterizes telomeres in other eukaryotes?

In order to answer some of these questions, it is necessary to characterize these transposable elements in other species. Thus far, *HeT-A* has been characterized in detail only in *D. melanogaster*. Additional *HeT-A* sequences are available from only three other species, *D. yakuba* (Danilevskaya 1998), *D. simulans* (Pardue et al., 1998), and *D. teissieri* (Ostler, 2001), and little is known about their genomic distribution and expression within these three species.

This study concerns sequence characterization of the *HeT-A* element in those species most closely related to *D. melanogaster*. *D. melanogaster* is one of about 150 species that belong to the *melanogaster* species group. Within this group are eight subgroups, each containing several closely-related species. This study focused on the *melanogaster* subgroup, which includes eight species: *D. erecta, D. melanogaster, D. mauritiana, D. orena, D. sechellia, D. simulans, D. teissieri,* and *D. yakuba* (Bock and Wheeler, 1972). Also included in the study was one less-closely related species, *D. lucipennis*, which belongs to the *suzukii* subgroup (Lemeunier et al.. 1986).

The objectives of this study were to examine the distribution of the *HeT-A* element among selected flies of the *melanogaster* subgroup, to obtain nucleotide sequences of *HeT-A* from these species, and to determine the phylogeny of the *HeT-A* elements. Because of their mobility within the genome and their presence in multiple copies, the phylogeny of TEs does not always trace the phylogeny of their host species. This is in contrast to non-mobile genes, which are often used in evolutionary genetics to determine phylogenetic relationships. The well-studied phylogeny of *Drosophila* makes comparison of the *HeT-A* phylogeny to that of its host possible. A TE phylogeny paralleling that of its host species would indicate that there is strong selection maintaining the function of *HeT-A* within the genome, which would explain the observed evolutionary conservation.

Materials and Methods

DNA extraction.

DNA for each species was extracted from 25-50 adult flies using a standard *Drosophila* DNA extraction protocol. Recipes for all reagents used are given in Appendix A. Flies were homogenized in a micro centrifuge tube with 800 µl of homogenization buffer. A lysis buffer (200 µl) and 7 µl of proteinase K (10mg/ ml) were added and the tube was incubated at 50° C for 15 minutes. Following lysis and protein degradation, 225 µl of 5M potassium acetate was added and the tube was incubated on ice for 15 minutes. Debris and detergent were then precipitated by centrifuging at 17000 g for 20 minutes.

The supernatant was removed following centrifugation, was added to an equal volume of phenol:chloroform (1:1), and the sample was again centrifuged (12000 g; 5 minutes). The upper phase was removed and combined with an equal volume of chloroform:isoamyl alcohol (24:1). The tube was spun again at 12000 g for 5 minutes, and once again the upper phase was removed to a clean tube. Two volumes of absolute ethanol were added and the sample was stored on ice for at least 30 minutes. Nucleic acids were pelleted by centrifugation at 17000 g for 20 minutes, and the pellet was washed with cold 70% ethanol and air-dried. The nucleic acids were resuspended in 125 µl TE [8.0] buffer and stored overnight at 4°C.

RNase A (2 µl of a 10mg/ml solution) was added and the sample was incubated for 30 minutes at 37°C. After RNA degradation, the sample was extracted with phenol and chloroform and the DNA was precipitated as described above. Finally, DNA was pelleted at 10000 rpm and washed with 125 µl 70% ethanol. It was then air-dried and resuspended in 80 µl TE [8.0].

A 2 µl aliquot of the sample was examined on a 0.8% agarose gel to verify DNA integrity and purity. Concentration of DNA was determined using spectrophotometric analysis and the DNA was stored at -20°C until used.

Polymerase Chain Reaction.

The presence of *HeT-A* elements was detected using the polymerase chain reaction (PCR). The primers used were complementary to a region at the 5' end of the *HeT-A* gene. They are called 5'UTR-1F and ORF-3R, named for the specific regions of the gene to which they bind, with sequences AAAGCCACCCTATTTATACC and GNGTCATRTCYTTNGCYTG (N= A, C, G, or T; Y= C or T; R= A or G), respectively. A standard 25 µl reaction was performed in separate reactions with DNA from each species. Reaction volumes were as follows:

15.625 µl water
2.0 µl dNTPs (10mM ea.)
1.25 µl 5'UTR 1F (5 µM)
1.25 µl ORF-3R (5 µM)
2.5 µl 10x Buffer
1.25 µl $MgCl_2$ [50 mM]
0.125 µl Taq polymerase[5U/µl]
1 µl DNA template
25 µl total reaction volume

Template DNA concentrations ranged from 10 ng/µl to 1000 ng/µl. In some cases twice the volume of $MgCl_2$ was used to obtain optimum PCR amplification.

One drop of mineral oil was added to each tube and the reaction was placed into a pre-heated (94°C) thermocycler block. A step-cycle file with 30 cycles was used. Each cycle consisted of: one minute at 94°; one minute at 50°; one minute at 72°. Products were analyzed by electrophoresis through a 1% agarose mini-gel.

Cloning and Transformation.

PCR products were cloned using the Invitrogen TOPO TA Cloning® kit following the accompanying instruction manual. The volume of PCR product used for ligation varied based on estimated concentrations of DNA from analysis of the mini-gel containing the PCR product. For the 1700 bp *HeT-A* fragment, 20-50 ng/µl was optimal. Following transformation of *E. coli*, 50 and 200 µl aliquots were plated to ensure colonies with proper spacing. The plates contained Xgal (20mg/ml in DMSO) and ampicillin (50µg/ml) to select for those plasmids carrying the TOPO plasmid with the *HeT-A* insert.

Plasmid Purification and Sequencing.

Ten to 15 white colonies and several blue colonies (control) were selected and grown in 2 ml of LB-Amp solution shaken at 37° overnight. The samples were then purified using the QIAprep® Miniprep Kit following the accompanying protocol. The purified plasmid was resuspended in 20-30 µl of elution buffer, and the concentration was estimated by agarose gel electrophoresis. The cloned PCR products were sequenced using an ABI-Prism automated DNA sequencer.

Phylogenetic Analysis.

Electronic sequence files were analyzed using alignment data generated by the Gene Stream program from the Institute of Human Genetics, Montpellier, France (http://www2.igh.cnrs.fr/bin/align-guess.cgi). Sequence from 12 species were aligned and the master sequence file stored in the NEXUS format. Phylogenetic analysis was performed on *HeT-A* sequences using the program PAUP 4.0 (Swofford 1991). Both heuristic and branch-and-bound searches were performed. For the latter, 10 random sequence additions were used to reduce the influence of taxa order on the analysis. To evaluate the statistical significance of the phylogeny, bootstrapping was performed. One thousand bootstrap replicates were examined using the heuristic option, with ten random sequence additions. For all searches, the sequences for *D. lucipennis* were used as outgroups.

Results

Following isolation, the genomic DNA from each species was quantified and qualified using spectrophotometry. Figure 3 gives the absorption curve obtained for a *D. yakuba* sample; similar spectra were obtained for other species (See Table 1for species list and Table 2 for DNA concentrations).

Species	DNA Concentration
Drosophila mauritiana	3000 ng/µl
Drosophila lucipennis	562.5 ng/µl
Drosophila yakuba	1125 ng/µl

Table 2. DNA yield from species used in this study.

The 5'UTR-1F and ORF-3R primer pair was used in detection of *HeT-A* elements using PCR. Figure 3 shows the location of these primers relative to the *HeT-A* element. 5'UTR-1F is complementary to a conserved sequence at the end of the 5'UTR of the element. ORF-3R is complementary to the middle of the coding region of the *gag* ORF. In *D. melanogaster*, these primers amplify a 1730 bp fragment. For each of the three species studied, PCR yielded a fragment of approximately 1700 bp size, which is similar to size expected for *D. melanogaster*. The fragments were visualized through 1% agarose gel electrophoresis as shown in Figure 3.

PCR products were cloned into the TOPO cloning vector (Fig. 5) and transformed into *E. coli* cells. This synthetic vector contains a gene for ampicillin resistance, which allows for selection of transformed cells. The site at which PCR product is inserted into the vector is flanked by *Eco* RI restriction enzyme sites, facilitating the removal of the fragment following transformation and plasmid isolation. Following inoculation and growth of cultures, plasmid DNA was isolated from the *E. coli* cells and the plasmids were digested using *Eco* RI to excise the *HeT-A* fragment. As seen in Figure 7, a fragment of approximately 1700 bp is released from the plasmid following digestion with *Eco* RI; this corresponds to the *HeT-A* gene.

Plasmids were purified and quantified for sequencing. Two sequences were obtained from *D. lucipennis* and one each from *D. yakuba* and *D. mauritiana*. PCR products were sequenced from the 3' end using a primer which anneals to the Sp6 promoter of the TOPO® vector; they were also sequenced in the reverse direction using a primer to the T7 promoter (See Figure 6). An example of the chromatogram from a sequencing reaction is shown in Figure 7. It can be seen that there is a limit of 600-700 bp of unambiguous sequence information that is generated from each sequencing reaction. Since the *HeT-A* fragment of interest is approximately 1700 bp, reliable sequence information is not possible for the central portion of the *HeT-A* fragment. Thus, sequence comparisons were limited to about 1400 bp, representing 700 bp from each end of the PCR product.

Fragment sequences were aligned using the GeneStream program, and the alignment refined by eye. The complete aligned data set is shown in Appendix B. Also included in the alignment is a single published sequence from *D. yakuba* (Danilevskaya et al., 1998), the two partial sequences from *D. teissieri* (Ostler, 2001), three complete sequences from *D.*

melanogaster (Biessmann et al., 1994; Danilevskaya et al., 1994), and two partial sequences from *D. simulans* (Clark et al., unpublished). Due to length variation among the various sequences, gaps are necessary to achieve proper alignment of all sequences.

Table 3 gives a distance matrix for the aligned sequences. Values range from 0.5 divergence between the two *HeT-A* sequences from *D. teissieri* and 1.6% divergence between the two sequences from *D. lucipennis*, to 38% divergence between the *HeT-A* sequences for *D. lucipennis* and *D. teissieri*. The small degree of intraspecies divergence is expected for *HeT-A* sequences from the same genome. However, the extensive divergence of some *HeT-A* sequences between species is characteristic of multiple subfamilies of *HeT-A* within the genome of some species.

	1	2	3	4	5	6	7	8	9	10	11	12
1 *D. lucipennis* 3	-	0.016	0.243	0.366	0.370	0.238	0.238	0.242	0.108	0.107	0.381	0.379
2 *D. lucipennis* 6	21	-	0.240	0.364	0.369	0.243	0.242	0.249	0.118	0.117	0.380	0.378
3 *D. mauritiana* 6	298	295	-	0.314	0.314	0.060	0.242	0.191	0.234	0.236	0.344	0.342
4 *D. yakuba* 4	442	442	385	-	0.014	0.302	0.305	0.297	0.345	0.350	0.137	0.135
5 *D. yajuba* AF	447	448	386	18	-	0.305	0.307	0.298	0.347	0.351	0.140	0.137
6 *D. melanogaster* 9D4	286	294	74	367	370	-	0.047	0.166	0.214	0.219	0.338	0.336
7 *D. melanogaster* 23Zn-1	288	295	90	374	376	59	-	0.173	0.220	0.223	0.339	0.338
8 *D. melanogaster* 17B3	297	307	238	369	370	207	218	-	0.223	0.229	0.327	0.327
9 *D. simulans 2*	137	151	278	410	412	256	265	273	-	0.071	0.354	0.353
10 *D. simulans* 5	133	147	274	407	408	257	263	274	89	-	0.362	0.361
11 *D. teissieri* 1	458	459	418	170	173	411	416	403	422	422	-	0.005
12 *D. teissieri* 2	456	457	416	167	170	409	414	402	421	421	7	-

Table 3: Pairwise distances between taxa. Data below diagonal represent the number of differences out of 1430 total positions; data above the diagonal are mean distances, excluding gaps. (Based on 1430 character sequences used for phylogenetic tree determination- see Figures 8 and 9).

The phylogeny of the *HeT-A* sequences was examined using parsimony analysis. Figure 9 is a phylogeny that includes the published sequences for *D. yakuba* and *D. melanogaster* cited above, and those obtained for the three species of this study. The sequences from *D. lucipennis* were used as an outgroup in all analyses. *D. lucipennis* belongs to the *suzukii* subgroup and as such is clearly distinct from the other species, all of which belong to the *melanogaster* subgroup. Although the *melanogaster* species group arose in Africa, extant species are found distributed throughout the Old World. Whereas the flies of the *melanogaster* subgroup are found primarily within Africa, *D. lucipennis* and other members of the *suzukii* subgroup are found in central and eastern Asia, reflecting the diversification of these flies during evolution (Bock, 1980). The parsimony analysis produces a single tree of 1153 steps. The outgroup branch roots two major clades. One is formed by two sequences from *D. melanogaster*, 9D4 and 23Zn-1, and a *D. mauritiana* sequence. The second clade is formed by the two *D. yakuba* sequences and a third *D. melanogaster* sequence, 17B3. The bootstrap values show that support for these 2 clades is high, indicating they represent real biological phenomena.

Figure 10 shows the phylogenetic tree obtained for analysis of all known *HeT-A* gene sequences, including those published, three from this study, and additional sequences obtained from the DNA laboratory. These results are consistent with the tree using fewer species but a bit more resolution is provided for certain relationships. From the root, the two major clades are one that includes 2 sequences from *D. simulans*, and one that includes the rest of the sequences from the *melanogaster* subgroup. Within this well-supported clade, the major division is a clade with sequences for *D. yakuba* and *D. teissieri*, and a clade consisting of sequences from *D. mauritiana*,

and *D. melanogaster*. The relatively low bootstrap value for the latter (62%) reflects the inclusion of the more divergent *D. melanogaster* 17B3 sequence. When compared to Fig. 9, the affiliation of the *D. yakuba HeT-A* sequences also switches, from the *D. melanogaster* 17B3 sequence to the two sequences from *D. teissiere*. This is consistent with the close evolutionary relationship between *D. teissieri* and *D. yakuba* (Tsakas and Tsacos, 1984; Jeffs et al., 1994).

A comparison of the phylogeny generated using *HeT-A* sequence information and the known species phylogeny for the *melanogaster* subgroup and *D. lucipennis* is shown in Figure 11. Different copies of *HeT-A* within the same genome are generally quite similar. For example, the two *HeT-A* sequences for *D. teissieri* differ by only 0.5% over the 1400 bp region. Similar values are seen for comparisons within the genomes of *D. lucipennis*, *D. simulans*, and *D. yakuba*. However, there appear to be two distinct subfamilies of *HeT-A* within the genome of *D. melanogaster*. Elements 9D4 and 23Zn-1 are 95% identical, whereas element 17B3 from the same species is only 83% identical to the other two elements. In contrast to intraspecies comparisons, interspecies comparisons show considerable divergence of *HeT-A* elements as shown in the distance matrix generated from analysis of all species (Table 3).

Discussion

The transposable elements *HeT-A* and *TART* are unique among eukaryotic transposons because they provide the only unambiguous example of a beneficial role for a transposable element. Until very recently, the mechanism of chromosome end protection through transposition of DNA to the telomeric regions had only been observed in *Drosophila*. However, Arkhipova and Morrison (2001) reported the existence of two non-LTR retrotransposon families in *Giardia*

lamblia that are located at the ends of the majority of chromosomes, evidence that they may be beneficial to the host in protection from terminal chromosome degradation. A distinction between these TEs and *HeT-A* and *TART* is that the former have not entirely replaced telomeric repeats in *Giardia*. In *Drosophila*, *HeT-A* and *TART* elements form the telomeric DNA and no sequences similar to the repetitive end DNA of other eukaryotes have been found in the *Drosophila* telomeres.

The evolutionary origin of this unprecedented telomere-maintenance mechanism in *Drosophila* is largely unknown and raises several questions. Is this the ancestral mechanism for telomere maintenance or is it a derived system that replaced the telomerase system of most other eukaryotes? If derived, how recently did it arise in *Drosophila.* An additional important question, which is the purpose of this research effort, is whether the *HeT-A* system is common to all species of the genus *Drosophila.*

The determination of *HeT-A* element distribution and evolutionary history is essential in answering such questions. Danilevskaya, et al. (1998) performed Southern blot analyses using a *HeT-A* probe from *D. melanogaster*. They found that cross hybridization drops drastically with species distance A *HeT-A* probe for *D. melanogaster* is only able to detect *HeT-A* in *D. mauritiana,* and *D. simulans*, two species that are closely related to *D. melanogaster*. This method was able to detect only faint signals in *D. yakuba* and *D. teissieri*, and was unable to detect *HeT-A* at all in other species. In contrast, using PCR amplification, *HeT-A* genes were found in all *Drosophila* species examined in this study; this is a significant finding. In addition, this report has identified the first *HeT-A* sequence outside of the *melanogaster* subgroup, namely, the sequences from *D. lucipennis*. The primers used for PCR contained a certain degree of

degeneracy to facilitate identification of divergent *HeT-A* genes. This is the advantage of using PCR, the reason for choosing this experimental protocol. (See Materials and Methods for primer sequences).

Within a species, sequences of amplified fragments showed considerable conservation (See Table 3 and Appendix B). Prior to this study, intraspecific data on *HeT-A* was only available for *D. melanogaster* (Danilevskaya, et al., 1994) and *D. yakuba* (Danilevskaya 1996). *D. melanogaster* elements 9D4 and 23Zn-1 were found to be 95% identical with respect to the coding region, and each was only about 83% identical to 17B3 from the same species. Thus, there are clearly different subfamilies of *HeT-A* elements within *D. melanogaster*, a finding first reported by Biessmann et al. (1994). This is consistent with the conclusions of phylogenetic analysis of the *P* transposable element within *Drosophila* (Clark and Kidwell, 1997) and the widespread TE *mariner* (Roberstson and MacLeod, 1993). In contrast to the *P* element, where many subfamilies are now incomplete or inactive, 9D4, 17B3, and 23Zn-1 have each been shown to be complete, active *HeT-A* elements (Pardue et al., 1996). Such intraspecific variation was not detected in the species examined in this study, but that is probably a reflection of the limited sampling done thus far. In the only other PCR-based survey reported, four sequences for *D. simulans* were found to be 98% identical, values similar to those reported here for intraspecific variation.

This study is the first phylogenetic analysis of *HeT-A* sequences reported. Although it is preliminary because it is based on a relatively small number of sequences, the findings merit attention. One interesting result is the placement of the *HeT-A* sequences from *D. mauritiana* relative to those from *D. melanogaster*. The *D. mauritiana* sequence is clearly more closely

related to *D. melanogaster* 9D4 and 23Zn1 than either of these are to the third sequence from *D. melanogaster,* 17B3. This reflects the close relationship between these two species and the conservation of *HeT-A*. Additional PCR sampling may reveal *HeT-A* sequences from *D. mauritiana* that are also similar to 173B from *D. melanogaster*. This sequence conservation between species is seen in comparisons of *HeT-A* elements between *D. teissieri*, and *D. yakuba*. Within the *melanogaster* subgroup these two species are closely related and distinct from the *melanogaster-simulans-mauritiana* trio (Tsakas and Tsacos, 1984; Jeffs et al., 1994). Thus, this clade of *HeT-A* sequences appears to trace the evolutionary history of the host species perfectly.

The conservation of *HeT-A* element sequences within a species is itself of interest. The *HeT-A* sequence obtained from *D. yakuba* in this study (*D. yakuba* 4) was remarkably similar to the published GenBank sequence (*D. yakuba* AF) (Danilevskaya, 1998). However, the two sequences clearly do not represent the same *HeT-A* gene. They differ at 22 sites, of which 19 represent transversions and 3 transitions. The differences between these sequences are similar to those seen when comparing very similar sequences from the genomes of *D. teissieri*, *D. simulans*, and *D. lucipennis*. Thus the PCR screen is clearly amplifying different *HeT-A* genes that are present in the same genome.

The phylogenetic placement of *D. simulans* elements was not expected, given the close relationship among *D. melanogaster*, *D. simulans*, and *D. mauritiana*. Based on molecular, morphological, and biogeographical data, these three sibling species have diverged from one another only in the past 2 million years or so (Bock and Wheeler, 1972). Sequence comparisons of the alcohol dehydrogenase gene from these three species provide support for this close relationship. The *Adh* genes from these three species differ by only about 1% at the nucleotide

level (Jeffs et al., 1994). Somewhere in the *D. simulans* genome there are likely *HeT-A* sequences which are closely related to those of *D. melanogaster* and *D. mauritiana*, but which were simply not sampled in this study. Similarly, there may be *HeT-A* genes within the genomes of *D. melanogaster* and *D. mauritiana* that are closely related to those of *D. simulans*. The divergent *D. simulans* sequences provide support for the idea of divergent subfamilies of *HeT-A* within a genome. Danilevskaya et al. (1998) report on additional *HeT-A* sequences from *D. melanogaster* (sequences unpublished) that differ from 9D4, 17B3, and 23Zn-1 by greater than 20%. This value is similar to the divergence of the *D. simulans* elements from those of *D. melanogaster* and *D. mauritiana* (See Table 2). Different subfamilies of TES in the same genome, each with a distinct evolutionary history, have been reported for other TES, including the *P* element from *Drosophila* (Clark and Kidwell, 1997) and the *mariner* element (Robertson and MacLeod, 1993). In order to determine the extent of *HeT-A* diversity in a genome, additional sequences will need to be isolated from each species, thus increasing sample size.

The different subfamilies of TES within a genome often represent inactive elements that are analogous to molecular fossils. *HeT-A* is interesting for comparison because it is a TE that is apparently performing an essential function. Thus, it is expected that there would be some selection against mutation of the *HeT-A* gene. However, if a particular *HeT-A* gene were to suffer a mutation, there are many other functional copies of the gene available in the genome and the host would not suffer deleterious consequences of the mutation. These mutant elements would then be free to accumulate additional mutations and the sequences would change dramatically over time. This is thought to be the source of the different subfamilies that are characteristic of TEs within the genome.

Transposable elements often show considerable sequence variation due to high mutation rate and mobile nature within the genome (Kidwell, 1997). The finding of this study that there is conservation within a genome hints at selection operating on these genes, but it should be stressed that the current analysis is limited by sample size. In spite of this, when combined with the observations that telomerase is not produced by *Drosophila* species (Labrodor and Corces, 1997) and that *HeT-A* elements are restricted to telomeric regions (Danilevskaya et al., 1994), it is increasingly likely that these transposable elements do indeed play a vital role in telomere maintenance.

Considerable work regarding the *HeT-A* element in *Drosophila* remains. More sequences should be obtained for each species, and sequence data should be obtained from species outside of the *melanogaster* subgroup, and then from other species groups of *Drosophila*. Intracellular flourescent tagging of Gag proteins associated with *HeT-A* showed that the protein is processed in *D. melanogaster* for telomeric destinations (Rashkova, Karam, and Pardue, 2002). Chromosomal location of the *HeT-A* genes within other species should be determined with *in situ* hybridization studies. This would provide information about whether *HeT-A* is confined to the chromosome ends within other species of *Drosophila*, or if these TEs are also found in other locations in the genome. It may be that a population of *HeT-A* is found at telomeres and that these sequences are conserved, while other populations of *HeT-A* are found elsewhere along the chromosomes and represent more divergent subfamilies.

Literature Cited

Arkhipova, I. R., and Morrason, H.G. (2001). Three retrotransposon families in the genome of *Giardi lamblia*: Two telomeric, one dead. Proc. Natl. Acad. Sci. USA *98* (25), 14497-14502.

Beissmann, H., Valgeirsdottir, K., Lofsky, A., Chin, C., Ginther, B., Levis, R.W., and Pardue, M. (1992). HeT-A, a Transposable Element Specifically Involved in "Healing" Broken Chromosome Ends in *Drosophila melanogaster*. Mol. Cell. Biol. *12*, 3910-3918.

Bock, I.R. 1972. Current status of the *Drosophila melanogaster* species group. Syst. Entomol. *5*, 341-356.

Bock, I. R., and Wheeler, M. R. (1972). The *Drosophila melanogaster* species group. Univ. Tex. Publ., *7213*, 1-102

Clark, J.B., Kim, P.C., and Kidwell, M.G. (1998). Molecular Evolution of *P* Transposable Elements in the Genus *Drosophila.* III. The *melanogaster* Species Group. Mol. Biol. Evol. *15* (6) 746-755.

Clark, J.B., and Kidwell, M.G. (1997). A phylogenetic perspective on *P* transposable element evolution in *Drosophila*. Proc. Natl. Acad. Sci. USA *94* (21) 11428-11433.

Danilevskaya, O.N., Slot, F., Traverse, K.L., Hogan, N.C., and Pardue, M.L. (1994). Drosophila Telomere Transposon HeT-A Produces a Transcript with Tightly Bound Protein. Proc. Natl. Acad. Sci. USA *91* (14), 6679-6682.

Danilevskaya, O.N., Tan, C., Wong, J., Alibhai, M., and Pardue M. (1998). Unusual Features of the *Drosophila melanogaster* Telomere Transposable Element Het-A are Conserved in *Drosophila yakuba* Telomere Elements. Proc. Natl. Acad. Sci. USA *95* (7), 3770-3775.

Greider, C.W. (1998). Telomerase activity, cell proliferation, and cancer. Proc. Natl. Acad. Sci. USA *95*, 90-92.

Henning, K., Moskowitz,N., Ashlock, M., Liu, P. (1998). Humanizing the Yeast Telomerase Template. Proc. Natl. Acad. Sci. USA. *95* (10), 5667-5671.

Jeffs, P.S., Holmes, E.c., and Aschburner, M. 1994. The molecular evolution of the alcohol dehydrogenase-related genes within the *Drosophila melanogaster* species subgroup. Molec. Biol. Evol. *11*, 287-304.

Karlseder, J., Smogorzewska, A., de Lange, T. (2002). Senescence Induced by Altered Telomere State, Not Telomere Loss. Science *295*, 2446-2449.

Kidwell, M.G. (1992). Horizontal transfer. Curr. Opinion Genetics Devel. 2, 868-873

Kidwell, M.G. and Lisch, D. (1997). Transposable elements as Sources of Variation in Animals and Plants. Proc. Natl. Acad. Sci. USA 94 (15), 7704-7711.

Lemeunier, F., David, J. R., Tsacas, L., and Ashburner, M. (1986). The *melanogaster* species group. In M. Ashburner, H.L. Carson, and J. N. Thompson, eds. The genetics and biology of Drosophila. Academic Press, London. *3e.*, 147-256.

Levis, R.W., Ganesan, R., Houtchens, K., Tolar, L.A., and Sheen, F. (1993).

Transposons in Place of Telomeric Repeats at a Drosophila Telomere. Cell *75*, 1083-1093.

Marx, Jean (2002) Chromosome End Game Draws a Crowd. Science *295*, 2348-2351).

Mason, J.M., and Biessman, H. (1995) The unusual telomeres of *Drosophila*. Trends in Genetics *11* (2), 58-62.

Nakatsu, C., Ng, J., Singh, R., Straus, N., and Wyndham, C. (1991). Chlorobenzoate Catabolic Transposon Tn5271 is a Composite Class I Element with Flanking Class II Insertion Sequences. Proc. Natl. Acad. Sci. USA *88* (19), 8312-8316.

Ostler, Chris (2001). Identification of a *HeT-A* Transposable Element in *Drosophila teissieri*. Senior Thesis. Weber State University, Ogden, UT.

Pardue, M., DeBaryshe, P.G., and Lowenhaupt, K. (2001). Another protozoan contributes to understanding telomeres and transposable elements. Proc. Natl. Acad. Sci USA *95* (25), 14195-14197.

Pimpinelli, S., Berloco, M., Fanti, L., Dimitri, P., Bonaccorsi, S., Marchetti, E., Caizzi, R., Caggese, C., and Gatti, M. (1995). Transposable elements are stable structural components of *Drosophila melanogaster* heterochromatin. Proc. Natl. Acad. Sci. USA *92* (9), 3804-3808.

Robertson, H.M., MacLeod, E.G. 1993. Five major subfamilies of *mariner* transposable elements in insects, including the Mediterranean fruit fly, and related arthropods. Insect Molec. Biol. *2*, 125-139.

Sheen, F. and Levis, R.W. (1994). Transposition of the LINE-Like Retrotransposon TART to Drosophila Chromosome Termini. Proc. Natl. Acad. Sci. USA *91* (26), 12510-12514.

Srivastava, M., Cartas, M., Rizvi, R.A., Singh, S.P., Serio, D., Kalyanaraman, V.S., Pollard, H.B., and Srinivasan, A. (1999). HIV-1 Gag Shares a Signature Motif with Annexin (Anx7), Which Is Required for Virus Replication. Proc. Natl. Acad. Sci. USA *96* (6), 2704-2709.

Tsakas, S.C. and Tsacas, L. 1984. A phenetic tree of eighteen species of the *melanogaster* group of *Drosophila* using allozyme data as compared with classification based on other criteria. Genetics *64*, 139-144.

Appendix A :Recipes for Buffers and Reagents

Homogenization Buffer

0.1 M NaCl
0.2 M sucrose
0.01 M EDTA
0.03 M Tris/ pH 8.0

Lysis Buffer

20 % SDS
0.25 M EDTA
0.1 M Tris/ pH 8.0

TE Buffer

0.01 M Tris
0.001 M EDTA/ pH 8.0

Appendix B

```
#NEXUS
BEGIN DATA;
DIMENSIONS NTAX=12 NCHAR=1430;
FORMAT DATATYPE=DNA GAP=- MISSING=? INTERLEAVE;
MATRIX

D._lucipennis_3        ATGTCCA---------GTTCTGACCACCTTTTCTCTGAAGACGAGGTACTTTCAATTTCCTCTAGCG-------AGCGATCTTCCCCAATTCATGTTAATATTTCACCTATGTCTCACGGGTCTGGCAAT
D._lucipennis_6        ATGTCCA---------GTTCTGACCACCTTTTCTCTGAAGACGAGGTACTTTCAATTTCCTCTAGCG-------AGCGATCTTCCCCAATTCATGTTAATATTTCACCTATGTCTCACGGGTCTGGCAAT
D._mauritiana_6        ATGTCCA---------TGTCCGACAACCTATTTTCTGACGATGAGGTACTTTCAATTTCCTCAAGCCCAGAA-CAGCNATCTTCCCCATTCTACTTAAATATATCGCCCATGTCCCAACGGATCAGATAAT
D._yakuba_4            ATGGCCCAAGTAATTCTCTCTGACGAC---TCTTCAAATGATGAGGTGCTCTCTCTTTTCTCGAGCCCAGAA-AGTCAAAATACCCCTTTCTACCTGGAAATCTCGCCCATGTCCCAAAATTCTAACAAC
D._yakuba_AF           ATGGCCCAAGTAATTCTCTCTGACGAC---TCTTCAAATGATGAGGTGCTCTCTCTTTTCTCGAGCCCAGAA-AGTCAAAATACCCCTTTCTACCTGGAAATCTCGCCCATGTCCCAAAATTCTAACAAC
D._melanogaster_9D4    ATGTCCA---------TGTCCGACAACCTATTTTCTGACGATGAGGTACTTTCAATTTCCTCAAGCCCAGAA-CAGCGATCTTCCCCTTTCTACTTAAATATATCGCCCATGTCCAACGGATCAGATAAT
D._melanogaster_23Zn1  ATGTCCA---------TGTCCGACAACCTTTTTTCTGACGATGAGGTACTTTCAATTTCCTCAAGCCCAGAA-CAGCGATCTTCTCCGTTCTACCTCAATATATCGCCCATGTCCCACGGATCAGACAAT
D._melanogaster_17B3   ATGTCCA---------CGTCCGACAACCTTTTTTCTGACGATGAGGTTCTTGCAATTTCATCAAGCCCAGAG-CAGAGATCATCCCCCTTCTATCTTAATATATCGCCCATGTCCCACGGATCTGACAAT
D._simulans_2          ATGTCCA---------GTTCTGACCACCTTTTCTCTGAAGACGAGGTACTTTCAATTACCTCTAGTCCAGAA-GAGCGATCTTCCCCTTTTCATTTAAATAT----------TCAATTGG----------
D._simulans_5          ATGTCCA---------GCTCTGACCACCTTTTCTCTGAAGACGAGGTGCTTTCAATTACCTCTAGC-------GAGCGATCTTCCCCTATTCACGTAAATATTTCACCTATGTCTCACGGGTCTGGCAAT
D._teissieri_1         ATGGCCAATATAATACTCTCTGACG--CTTCT-TCAAACGACGAAGCGCTCACCCTATTCTCGAGCCCAGAAAGACAAAATACCCCTTTCTACCTGGAAATATCGCCCATGTCCCATGATTCTAACAACT
D._teissieri_2         ATGGCCAATATAATACTCTCTGACG--CTTCT-TCAAACGACGAAGCGCTCACCCTATTCTCGAGCCCAGAAAGACAAAATACCCCTTTCTACCTGGAAATATCGCCCATGTCCCATGATTCTAACAACT

Dlu3   TCCCTGGTTAATACCGTCAATATTAACCAGAAGAAATTGCCTGCAAATCGA---AATAAAAGCCCAATTAACTCTTCTTGG------TAGCCTTTAATTTCTTTAATTCCCTTTCTCTAAATATG
Dlu6   TCCCTGGTTAATACCGTCAATATTAACCAGAAGAAATTGCCTGCAAATCGA---AATAAAAGCCTAATTAACTCTTCTTGG-----GTAGCCTTTAATTTCTTTAATTCCCTTTCTCTAAATATG
Dma6   TCTCAGATTAATACTATAATTTTTAATCAGAAGAAATTGCCCTCAAATCAAGCAAACATAAGTCTAAAAAACTCTTCTGGG-----GCTGCTATAAAAAATTGTTAGTTCCCTTTCACATAAGAAG
Dya4   TCTCAGTTTAATATCAGTATAATAAACATGAAAAAATCGTCTTCCAACTCTGCAATTAACAGTCTAAAAAACCCTTCCGAG-----GCTGCTATAAAAATTATAAATTCACTTACATATAAGGGG
DyaAF  TCTCAGTTTAATATCAGCATAATAAACATGAAAAAATCGTCTTCCAACTCTGCAATTAACAGTCTAAAAAACCCTTCCGAG-----GCTGCTATAAAAATTATAAATTCACTTACATATAAGGGG
Dme9D  TCTCAGATTAATACTATAATTTTTAATCAGAAGAAATTGCCCTCAAATCAAGCAAACATAAGTCTAAAAAACTCTTCTGGG-----GCTGCTATAAAAATTGTTAGTTCCCTTTCACATAAGAAG
DmeZn  TCTCAGATTAATACAGTCATCATTAATTCGAAGAAATTGCCCTCAAATCAAGCAGACATAAGTTTAAAAAACTCTTCTGGG-----GCTGCTATAAAAATTGTTAATTCCCTTTCACACAAGAAG
Dme17  TCTCAGATTAATACAATCATTATTAATCAGAAGCAATTGTCCTCTAGCCAAGCAAATAAAAGTCCAAAAAACTCTTCAGAG-----GCTGCCATAAAAATTGTTAATTCCCTTTCACATAAGGGG
Dsim2  -CTGGCAATAATACCGTCAATATTAACCAGAAGAAATTGCCTTCAAATCAA---AAGAAAAGCCCAAATAACTCTTCTTTAATTCTTCTGCCTTTAATTTCTTTAATTCCCTTTCTCTAAATATG
Dsim5  TCCCTGGTTAAAACGACCTATATTAACCAGAAGAAATTGCCTCCAAATCAA---AAGAAAAGCCCAATTATCTCTTCTTGG-----GT---------AGCCTTTAATTTCCTTTCTCTCAATAAG
Dtei1  -CCCAGTTCAATATCAACATAATTAACATGAAAAAACTGTCTTCAAACCCTGAAATTAACAGTCTAAAAAACCCTTCCGAG-----GCTGCTATAAAATTCTTAAATCCACTTACATATAAGGGG
Dtei2  -CCCAGTTCAATATCAACATAATTAACATGAAAAAACTGTCTTCAAACTCTGAAATTAACAGTCTAAAAAACCCTTCCGAG-----GCTGCTATAAAATTCTTAAATTCACTTACATATAAGGGG

Dlu3   AAAGAGAT-ATAAATGCCAAAAATGCCCAAAGAGACCCCCTCTCGATCACCAATACAGCTGCAGCCAATGTTGGCGCCAAAAGCAGCATCTCGAAGGGGAAA
Dlu6   AAAGAGAATATAAATGCCAAAAATGCCCAAAGAGACCCCCTCTCGATCACCAATACAGCTGCAGCCAATGTTGGCGCCAAAAGCAGCATCTCGAAGGGGAAA
Dma6   AAAGAGAACACAAACACTAATAATGCCCAAAAAGACCCCCTCTCACTCACCAATACTACTGCAAGCACTTCTGGCGCCAAAAGAAGCATCTCGGAGGGGAAA
Dya4   AAAGAGAATAGAGAAAACAAAAATGCCCAAAAAGACCCGCTCTTCCTTACTAATACTAATAAAGAAACAGCTGGCGCCAAAAGTAGCGTCTCAAATGGGCCA
DyaAF  AAAGAGAATAGAGAAAACAAAAATGCCCAAAAAGACCCGCTCTTCCTTACTAATACCAATAAAGAAACAGCTGGCGCCAAAAGTAGCGTCTCAAATGGGCCA
Dme19  AAAGAGAACTCAAACACTAATAATGCCCAAAAAGACCCCCTCTCACTCACCAATACTACTGCAAGCACTTGTGGCGCCAAAAGCAGCATCACAGAGGGGAAA
DmelZ  AAAGAGAACACAAACGTTAATAATGCCCAAAAAGACCCCCTCTCACTCACCAATACTACTGCAAGCACTTGTGGCGCCAAAAGCAGCATCTCAGAGGGGAAA
DmelB  AAAGAGAAGACAAACACCAATAATGCCCAAAAAGACCCCCTCTCGCTCACTAATACTGTTGCAGCCACTGTTGGCGCCAAAAGCAGCATCACGAAGGGGAAA
Dsim2  AAAGAGAATATAAACGCCAAAAATGCCCAAGA-GACCCCCTCTGGATCAGT--------------CAATGTTGGCGCCAAAAGCAGCATCTCGAAGGGGAAA
Dsim5  AAAGAGAATATAATCGCC-------------------------------AGT--.----------CAATGTTGGCGCCAAAAGCAGCATCTCGAAGGGGAAA
Dtei1  AAAGAGAATAGGAAAAACGAAAATGCCCAAAAAAACCCGCTCTCCCTTACTAATACTAATAAAGAAACAGCTGGCGCCAAAAGTAGCGTCTCAAATGGGTCA
Dtei2  AAAGAGAATAGGAAAAACGAAAATGCCCAAAAAAACCCGCTCTCCCTTACTAATACTAATAAAGAAACAGCTGGCGCCAAAAGTAGCGTCTCAAATGGGTCA

Dlu3   ACAGAATTAAGAGGCGCCAAACCGAGCGACGCTATGGGAAATTTCCCCTCTCTCCCACACACCGACA---GCGCAGAGAAAAATCTGAGCTC
Dlu6   ACAGAATTAAGANGCGCCAAANCGAGCGACGCTATGGGAAATTTTCCCTCTCTCCCACACACCGACA---GCGCAGANAAAAATCTGAGCTC
Dma6   ACAGACTTGAGAGGCGCCAAAACGAGCGATGCAATGGGAAGTTTCCCCTCTCTCTCGCACAGCGACAATAGCATAGAGAAAAACCTCAGCTC
Dya4   ATGGAT---------GACAAAAAGAGCGGCGCTCTTAGAAATTTCCCTTTCCCCACACATGAAGACAACAGCATGGAGAGAAACCTCAGCTC
DyaAF  ATGGAT---------GACAAAAAGAGCGGCGCTCTTAGAAATTTCCCTTTCCCCACACATGAAGACAACAGCATGGAGAGAAACCTCAGCTC
Dme19  ACAGATTTGAGAGGCGCCAAAACGAGCGATGCAATGGGAAGTTTCCCCTCTCTCTCACACAGCGACAGTAGCATAGAGAAAAATCTCAGCTC
DmelZ  ACAGACTTTAGAGGCGCCAAAACGAGCGATGCAATGGGAAGTTTCCCCTCTCTCTCGCACAGCGACAATAGCATAGAGAAAAATCTGAGTTC
DmelB  GCAGAATTAAAAAGCGCCAAACTAAGCGACGCTTTGGGAAATTTCCCCTCCCTCTCATATAGCGATGGTAGCATT------AATCTTAGCTC
Dsim2  ACAGAATTAAGAGGCGCCAAACAGAGCGACGCTATGGGAAATTTCCCCTCTCTCCCACACACCGACAG---CGCAGAGAAAAATCTCAGCTC
Dsim5  ACAGAATTACGAGGCGCCAAACCGAGCGACGCTATGGGAAATTTCCCCTCTCTCCCACACAGCGACAG---CGCAGAGAAAAATCTCAGCTC
Dtei1  ACGGAT---------GACAAAAAGAGTGACGCTCTCAGAAATTTCCCCTTCCCAATACATGAAGACAACAGCATAGAGAGAAACCTCAGTAC
Dtei2  ACGGAT---------GACAAAAAGAGTGACGCTCTCAGAAATTTCCCCTTCCCAATACATGAAGACAACAGCATAGAGAGAAACCTCAGTAC
```

```
Dlu3  ATCCACCAAAATTGGACGTAACGCTTC-TTCCCCTCCTTCTTATACT-CACACACAAATAAATAAAGCCGC-GAAAC-AATTTAGAAG-CTGCTCTAA-TCTCTCGN-GCTTGG-CAATA
Dlu6  ATCCACCAAAATTGGACGTAACGCTTC-TTTCCCCTCCTTCTTATACTCACACCCAAATAAATAAAGCCGCCGNAACNAATTTANAAAGCTGCTCTAAATCTCTCGG-GCTGG--CAATA
Dma6  CTTCACCAAAATTGGACCCAACGCTTC-TTTCCCCTTTCTTCTCATGCACACACACACTANCAAATCCACTGACATAAGCTTANAAAGCCACCTCAAAATGGCCCGGCGCTTGCCCAATA
Dya4  ATCTACAAAAATTGGCTCTAAAAGCAT-TTCCCCTCACTCTCTCTCACCTACACACACAAGCAAGGGTATTAACATAAGC-ACAAACAGCCGCTCAAAAAGTCCCGC-GCTTGC-AAATG
DyaAF ATCTACAAAAATTGGCTCTAAAAGCAT-TTCCCCTCACTCTCTCTCACCTACACACACAAGCAAGGTTATTAACATAAGC-ACAAACAGCCGCTCAAAAAGTCCCGC-GCTTGC-AAATG
Dmel9 TTCCACCAAAATTGGACCAAATGCTTC-TTCCCCTCCTTCTCATGCACACACACACACTAGCAAATCCTCTAATATAAGC-TTAGAAATCCGCTCAAAATATCCCGC-GCTTGC-CAATA
DmelZ TTCCACCAAAATTGGACCAAACGCTTC-TTCCCCTCCTTCTCATGCACACACTCACACTAGCAAATCCACTGATATAAGC-TTAGAAAGCCGCTCAAAACATCCCGC-GCTTGC-CAATA
DmelB TTCTACCAAAATCGGATGCAATAAAAT-TTCCCCTCTTCCTTATGCACATACTAACACAAGCAAGGCCACAAACATAAGC-GTAGAAAGCCGCTCAAAATGTTCCGC-GCTTGT-TAATT
Dsim2 ATCCACCAAAATCGGACCTAATGCTTC-TTCCCCTCCTTCTTATACACACACGCAAATAAATAAAGCCACCGAAACAAAT-TTAGAAAGCTGCTCTAGATCTCCCGC-GCTTGC-CAATT
Dsim5 ATCCACCAAAATCGGACTTAATGCTTC-TTCCCCTCCTTCTCATACACACACGCAAGTAAATAAAGCCACCGAAAAAAAT-A-AGAAAGCTGCTCTAGATCTCCCGC-GCTTGC-CAATT
Dtei1 ATCTACCAAAATTGACTCTAAAGCCAT-TTCCCCTCACTCTCTTTCACCCACACACACACAAGCAATGTTATTAACATATGC-ACTAACAGCCGCTCTAAAAGTCCCGC-GCTTGC-AAATA
Dtei2 ATCTACCAAAATTGACACTAAAGCCAT-TTCCCCTCACTCTCTTTCACCCACACACACAAGCAATGTTATTAACATATGC-ACTAACAGCCGCTCTAAAAGTCCCGC-GCTTGC-AAATA

Dlu3  ACAAC-AGACCCATG------------CCCACAGTGGTGAATTTAANAACAGCTCGTACCCTTAGTCGGCCTGCTGCCAAGCGGGACTTATTTAAA----TCTCC---GTCCAGGAGCCCAGACGAGCAGCTCAT
Dlu6  ACAAC-AGACCCATG------------CCCACAGTGGTGAATTTAACAACAGCTCGTACCCTTAGTCGGCCTGCTGCCAAGCGGGACTTATTTAAA----TCTCC---GTCCAGGAGCCCAGACGAGCAGCTCAT
Dma6  CAATCTAGCGCCAAG------------CCCACTGTGGTGAATTTACCAGCCGCCCGCATCTTAAGCCGGCCTGCAGCCAAGCGGGATTTATTTAAAATCATCATC---CTCCGGAAGCCCAGACGAGCAGCCTAT
Dya4  -------CCACTACA------------CNCTCTATGGGTAACACATNTGCATCCAGAACCCTAAACCGGCNTGCAGCCAAGCGGGACCTTTTTAAT----TCACC---CCCCAATAGCACGGGCGCCCTGCCTAT
DyaAF -------CCACTACA------------CACTCTATGGGTAACACATCTGCATCCAGAACCCTAAACCGGCCTGCAGCCAAGCGGGACCTTTTTAAT----TCACC---CCCCAATAGCACGAACGCACTGCCTAT
Dmel9 -------ACCCCAAATCTAGCGCCAAGCCCACTGTGGTGAATTTACCAGCTGCCCGCATCCTAAGCCGGCCTGCAGCCAAGCGGGATTTATTTAAA----TCATCATCCTCACGAAGCCCAGACGAGCAGCCTAT
DmelZ -------ACCCCAAATCTAGCGCCAAGCCCACTGTGGTGAATTTACCAGCTGCCCGCATCCTAAGCCGGCCTGCAGCCAAGCGGGATTTATTTAAA----TCATCATCCTCCCGAAGCCCAGACGAGCAGCCTAT
DmelB -------AGCCCAACTCTAGCGCCATGCCCACTGTGGGGAATTTATCTGCAACCCGCACACTAAGCCGGCCTGCAGCCAAGCGGGATTTATTTAAA----TCACC---CTCCAAAAGCCCTGACGAGCAGCCTAT
Dsim2 ---------CCCAACAATAGCGCCATGCCCACTGTGGGGAACTCAACAACAGCCCGCACCCTTAGTCGGCCTGCTGCCAAACGGGACTTATTCAAC----TCACC---GTCCAGGAGCCCAGACGAGCAGCTCAT
Dsim5 ---------CCCAACAATAGCGCCTTGCCCACTGTGGGGAACTTAACAACAGCCCGCACCNTTAGGCGGCCTGCTGCCAAGCGGGACTTATTCAAC----TCACC---GTCCAGGAGCCCAGACGAGAAGCCCAT
Dtei1 ---------TCCACCA------CCACGCATACTATGGGAATCACTTCTACATCTAGAACCCTCAACCGGCCTGCAGCCAAGCGGGACCTTTTAAAT----TCACC---CCCCAATAGCACGGGCGCACAGCCTAT
Dtei2 ---------TCCACCA------CCACGCATACTATGGGAATCACTTCTACATCCAGAACCCTCAACCGGCCTGCAGCCAAGCGGGACCTTTTTAAT----TCACC---CCCCAATAGCACGGGCGCACAGCCTAT

Dlu3  GAGTTTCTCGGAAATGATTGCTGAAACTGGTCCAATCTTTGTGGCACCCACAGTTCCGGCTCCTCTAACGAAAACTCCGGCCAAGAGGACAAACGATGATTTGGACTGCTCCA------GTTTCAA----------
Dlu6  GAGTTTCTCGGAAATGATTGCTGAAACTGGTCCAATCTTTGTGGCACCCACAGTTCCGGCTCCTCTAACGAAAACTCCGGCCAAGAGGACAAACGATGATTTGGACTGCTCCA------GTTTCAA----------
Dma6  GAGTTTTTCGGAAGTGGTCGCCGGCACGGGTTCAATTTTTGCGGCACCCTGTGTCCCGGCACCTTTAACGAAAACTCCAGGCAAGCGAACAAACGACGATCTGGACTGCTCCA------ACTTTAA----------
Dya4  GAGTTTTTCGGAAGTGGTGGCCGGAACCGGTCCAGGACTAGTGGCATCCTCTGATCCGGCTCCAATTACGAAAACCCCGGGCAAGCGCACAAATACCAACATGGACATGGATA------GCTTTAGCTA------TAA
DyaAF GAGTTTTTCGGAAGTGGTGGCCGGAACCGGTCCAGGAATTGTGGCATCCTCTGATCCGGCACCAATTACGAAAACCCCGGGCAAGCGCACAAATACCAACATGGACATGGATA------GCTTTAGCTA------CAA
Dmel9 GAGTTTTTCGGAAGTGGTTGCTGGCACGGGTTCAATTTTTGCGGCACCTTGTGCCCCGGCACCTCTAACGAAAACTCCAGGCAAGCGGACAAA---CGAC---GATCTTGACT------GCTCCAACTT------TAA
DmelZ GAGTTTTTCGGAAGTGGTCGCTGGCACGGGTTCAATTTTTGCGGCACCCTGTGTCCCGGCACCTTTAACGAAAACTCCAGGCAAGCGGACAAA---CGAC---GATCTGGACT------GCTCCAACTT------TAA
DmelB GAGTTTTTCGGAAGTGGTTGCTGGAACGGATACAAATTTTGTGGCATCCCCTGATCCGACACCACAATCTAAAACTCCAGGCAAGCGGACAAA---TGAC---GAACTAGACA------GCTCAAGGTT------CAA
Dsim2 GAGTTTTTCGATAATAGTTGCTGGAACTGGTCCAATCTTTTTGGCACCCACAATTCCGGCTCCTCTAACGAAAACTCCGGCCAAGAGGACAAA---TGAC---GACTTGGACT------GCTCCAGTTT------CAA
Dsim5 GAGTTTCTCGGAAGTGGTTGCTGGAACTGGTCCAATCTTTGTGGCACCCACATTTCCGGCTCCTCTAACGAAAACTCCGGCCAAGAGAACAAA---TGAC---GACTTGGACT------GCTCCAGTCT------CAA
Dtei1 GAGATTTTCGGAAGTGGTGGCCGGAACCGGTACAGGACTAGTGGCATCCTCTGATCCGGCACCAACTTCGAAAACCCCGGGCAAGCGCAAAAT---TTCC---AACATGGACATGGATAGCTTCAGCCAAAGCTACAA
Dtei2 GAGATTTTCGGAAGTGGTGGCCGGAACCGGTACAGGACTAGTGGCATCCTCTGATCCGGCACCAACTTCGAAAACCCCGGGCAAGCGCAAAAT---TTCC---AACATGGACATGGATAGCTTCAGCCAAAGCTACAA

Dlu3  GACGCCAAATAAACGA-TTATGCGTGACCTCCAATTTCAAATCTCCCAGCATTTTTCCACCCCTCACCACACCCGTTTTCCAAAGTAAGGCAGCCAAATCTGTATATGAGGAATCCAAAAC
Dlu6  GACGCCAAATAAACGA-TTATGCGTGACCTCCAATTTCAAATCTCCCAGCATTTTTCCACCCCTCACCACACCCGTTTTCCAAAGTAAGGCAGCCAAATCTGTATATGAGGAATCCAAAAC
Dma6  GACGCCCAATAAAAAA-TTATGCGCGACTTCCAACTTTGTAACTCCCAGCATTTTTCCGCCGCTCATCACTCCCGTTTTCAAGAGCAAGGCAGCTCAATCTGTTTACGAGGAATCCAAAGC
Dya4  AACGCCCAATAAAAAA-GCATGTGTGCCCACCAATTTTGCGACCCCAAACCCTTTCCCCCCCCCTAGCCACCCCCATCTTTAAAAGTAAGGCGGCCAAAAGTATTGTCGAGGAAATAAAAGC
DyaAF AACGCCCAATAAAAAA-GCATGTGTGCCCACCAATTTTGCGACCCCAAACCCTTTCCCCCCCCCTAGCCACCCCCATCTTTAAAAGTAAGGCGGCCAAAAGTATTGTCGAGGAAATAAAAGC
Dmel9 GACGCCCAATAAAAAAATTATGCGCGACTTCCAACTTTGTAACTCCCAGCATTTTTCCGCCGCTCATCACTCCCGTTTTCAAGAGCAAGGCAGCTCAATCTGTTTACGAGGAATCCAAAGC
DmelZ GACGCCCAATAAAAAA-TTATGCGCGACTTCCAACTTTGTAACTCCCAGCATTTTTCCGCCGCTCATCACTCCCGTTTTCAAGAGCAAGGCAGCTCAATCTGTTTACGAGGAATCCAAAGC
DmelB GACACCCAACAAAAGA-ATATGTGCGACAACAAACTTTGAAACTCCTAACATTTTTCCTCCCCTCATTACACCCGTTTTTAAAAGCAAGGCAGCACAATCTGTTTATGAGGAATCCAAAGC
Dsim2 GACGCCAAATAAACGA-TTATGCGTGACCACCAATTTCAAATCTCCCAGCATTTTTCCACCCCTCACCACACCCGTTTTCCAAAGCAAGGCAGCCAAATCTGTATATGAGGAATCCAAAAC
Dsim5 GACGCCAAATAAACGA-TTATGCGCGACCACCAATTTCAAATCTCCCAGCATTTTTCCACCCGTCACCACACCCGTTTTCCATAGCATGGCAGCCAAATCTGTATATGAGGAATCCAAAAC
Dtei1 AACGCCTAATAAAAAG-GCATGCGTGCCCTCCATTTTTGCGACCCCAAACCCTCTCCCCCCCCTAGCTACCCCCATCTTTAAAAGCAAAGCGGCCAAATGTATTGCCGAGGAAATCAAAGC
Dtei2 AACGCCTAATAAAAAG-GCATGCGTGCCCTCCATTTTTGCGACCCCAAACCCTTTCCCCCCCCCTAGCTACCCCCATCTTTAAAAGCAAAGCGGCCAAATGTATTGCCGAGGAAATCAAAGC
```

```
Dlu3  CAGGAATCGACTCTCACACCAGCCGTTACCCTGCAGCACCAATGCTCCTGCACGCAGCGCAACGACGGCA----------------------CCTCCTCAAAATA--CAGAAGCTGAGCTGCCACC
Dlu6  CAGGAATCGACTCTCACACCAGCCGTTACCCTGCAGCACCAATGCTCCTGCACGCAGCGCAACGACGGCA----------------------CCTCCTCAAAATA--CAGAAGCTGAGCTGCCACC
Dma6  CAGAAATGGACCCCCCCCCGCCGGCCCTCGCCTGCAGCATCAATGCTTCTGCTCGCAGCGCAGCGGCGCCACCCGGGATTACACCCC---TACCCCCTCAAAACA--CAGATACAGAGCTGCCTCC
Dya4  ----------CCCCCGCCACC-----CAGTCGAGAGTGAAAAGGCCTCTGCACGCAGCGCACCGGCACAAACTGAAATTGCCCCCC---CTCCCCCTAAAAATT--CGGCTACCGAACTGCCCCC
DyaAF ----------CCCCCGCCACC-----CAGTCGAGAGTGAAAAGGCCTCTGCACGCAGCGTACCGGCGCAAACTGAAATTGCCCCCC---CTCCCCCTAAAAATT--CGGCTACCGAACTGCCCCC
Dmel9 CAGAAATGGACCCCCCCCGCCGGCCACCGCCTGCAGCATCAATGCCTCTGCTCGCAGCGCAGCG-CGCCACCCGGAATTGCCCCCC---TACCCCCTCAAAACA--CAGATACAGAGCTGCCTCC
DmelZ CAGAAATGGACCCCCCCCGCCGGCCCTCGCCTGCAGCATCAATGCCTCTGCTCGCAGCGCACGG-CGCCACCCGGGATCGCCCCCC---TACCCCCTCATAATA--CAGATGCAGAGCTGCCTCC
DmelB CAGGAATGGACCCCCCCCGCCAGCCGTTACCCTGCAGCATCAATGCCTCTGCTCGCAGCGCAGCG-CGCCACCCGGTTTAGCCCCCCCCCTACCCCCTCAAAACT--CAGATTCAGAACTGCCTCC
Dsim2 CAGGAATCGACACACACACCAGCCGTTACCCTGCAGCACCAATGCTCCTGCACGCAGCGCAACGTCGCCA----------------------CCCCCTCAAAATA-CAAGATGCTGAGCTGCCACC
Dsim5 CAGGAATCGACACTCCCACCAGCCGTTACCCTGCAGCACCAATGCTCCAGCACGCAGCGCAACGACGCCA----------------------CCCCCTCGAAATA--CGGATGCTGAGCTGCCACC
Dtei1 ----------CCCCCGCTACC-----CAGTCGAGAGTGTAAAGGCCACTACACGCAGCGCACCGGCGCAAACTGATATTGCCC--------CCCCCTCTAAAAATTCGGCTACTGAACTGCCCCC
Dtei2 ----------CCCCCGCTACC-----CAGTCGAGAGTGTAAAGGCCACTACACGCAGCGCACCGGCGCAAACTGATATTGCCC--------CCCCCTCTAAAAATTCGGCAACTGAACTGCCCCC

Dlu3  TTGGAAACTCGTGCCACTAAGCTGCAGAGCGCCACCCATACTCGTTGACGACGTCAAAGAAATTGTCCCGCTCCTGGAAAAGCTAAACTATACAGCAGGAGCATCCAGCTACACCACC
Dlu6  TTGGAAACTCGTGCCACTAAGCTGCAGAGCGCCACCCATACTCGTTGACGACGTCAAAGAAATTGTCCCGCTCCTGGAAAAGCTAAACTATACAGCAGGAGCATCCAGCTACACCACC
Dma6  ATGGAAAATTGTGCCCCAGAGCCGTAGAGCACCCCCTATACTCGTCAATGATGTAAAGGAAATTGTACCTCTACTGGAAAAGCTGAACTACACAGCAGGAGTCTCCAGCTATACTACC
Dya4  GTGGCAAATAGTTCCACAGAGCCGCAGGGCCCCCCCTATTCATATAGAAAATGTTAGGGAAATCGTGCCACTATTGGAAAAGCTAAATTACTCAGCAGGGGTAGACAGTTTTACAACA
DyaAF GTGGCAAATAGTTCCACAGAGCCGCAGGGCCCCCCCTATTCATATAAAAAATGTTAGGGAAATCGTGCCACTATTGGAAAAGCTAAATTACTCAGCAGGGGTAGACAGTTTTACAACA
Dmel9 ATGGAAAATCGTGCCCCAGAGCCGTAGAGCACCCCCTATACTCGTCAATGATGTAAAGGAAATTGTACCTCTACTGGAAAAGCTGAACTACACAGCAGGAGTCTCCAGCTACACTACC
DmelZ ATGGAAAATCGTGCCCCAGAGCCGTAGAGCACCTCCTATACTCGTCAATGATGTAAAGGAAATTGTACCTCTACTGGAAAAGCTGAACTACACAGCAGGAGTCTCCAGCTACACTACC
DmelB TTGGAAAATCGTGCCACAAAGTCGCAGAGCACCCCCAATAATCGTCAATAATGTAAAAGAAATAGTCCCGCTCCTGGAAAAACCAAATTACACAGCAGGAGTATCCAGTTACACCACC
Dsim2 TTGGAAACTTTTGCCACTAAGCTGCGGAGCACCACCCATACTCGTCGACGACGTCAAAGAAATTGTCCCGCTCCTGGAAAAGCTAAACTATACAGCAGGAGTCTCCAGCTACACTACC
Dsim5 TTGGAAAATTGTGCCACTAAGCTGCAGAGCACCACCCATACTCGTCAACGACGTCAAAGAAATTGTCCCGCTCCTGGAAAAGCTAAACTATACAGCAGGAGTCTCCAGCTACACTACC
Dtei1 GTGGCAAATAGTGCCACAGAGCCGCAGGGCCCCCCCCTATTCATATAGATAATGTTAGGGAAATTGTGCCACTATTGGAAAAACTAAATTACTCAGCAGGGGTGGACAGTTTCACAACA
Dtei2 GTGGCAAATAGTGCCACAGAGCCGCAGGGCCCCCCCCTATTCATATAGATAATGTTAGGGAAATTGTGCCACTATTGGAAAAACTAAATTACTCAGCAGGGGTGGACAGTTTCACAACA

Dlu3  AGGGCAACTGAAGGGAACGGGGTCAGGATCCAGGCGAAAGACATGACAGC
Dlu6  AGGGCAACTGAAGGGAACGGGGTCAGGATCCAAGCCAAGGACATGACAGC
Dma6  AGGGCTATAGAAGGAAACGGGGTCAGGATACAGGCTAAGGACATGACCGC
Dya4  AAAACGTCTATAGGAAACGGTGTAACTATACAGGCCAAAGACATGACNGN
DyaAF AAAACGTCTATAGGAAACGGTGTAACTATACAGGCTAAAGACTTGACTGC
Dmel9 AGGGCTATAGAAGGAAACGGGGTCAGGATACAGGCAAAGGACATGACCGC
DmelZ AGGGCTATAGAAGGAAACGGGGTCAGGATACAGGCAAAGGACATGACCGC
DmelB AGAGCAATAGATGGGAATGGGGTCAGGATCCAGGCCAAGGATATGACCGC
Dsim2 AGGGCAATTGAAGGGAACGGGGTCAGGATCCAGGCTAAAGACATGACCGC
Dsim5 AGGGCAATTAAAGGGAACGGGGTCAGGATCCAGGCGAAAGATATGACAGC
Dtei1 AAAACAGCTATAGGTAACGGCATAACTATTCAGGCCAAAGATATGACAGC
Dtei2 AAAACAGCTATAGGTAACGGCATAACTATTCAGGCCAAAGATATGACAGC
;
END
```

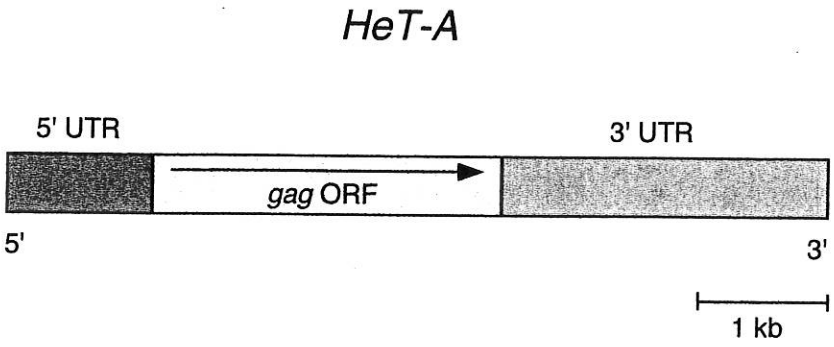


Figure 1. Schematic representation of the *HeT-A* transposable element of *D. melanogaster*. The gene includes three distinct regions: a 5' untranslated region (UTR) of unknown significance, an open reading frame (ORF) that encodes a Gag-like protein, and a 3' untranslated region that includes a promoter sequence that controls transcription of an adjacent *HeT-A* element. A complete *HeT-A* element is about 6000 bp in length.

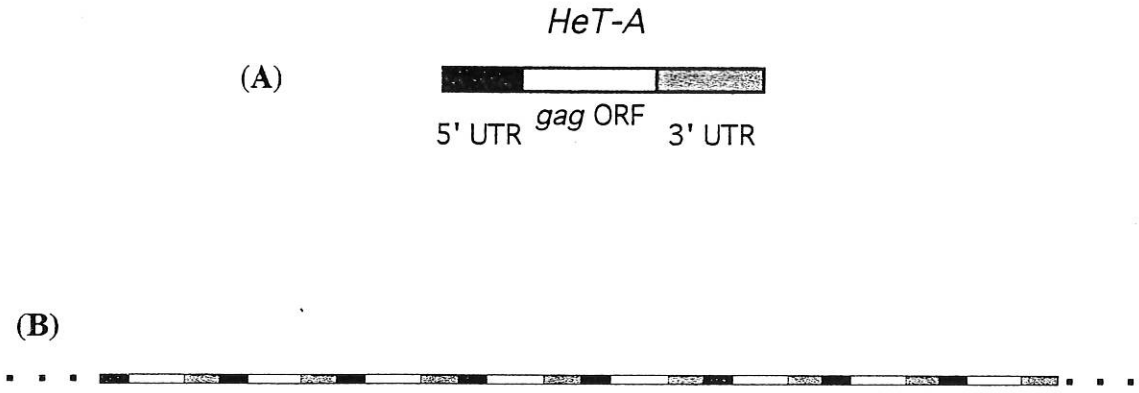


Figure 2. A schematic representation of the *HeT-A* gene. (**A**). The *HeT-A* gene includes 5' and 3' untranslated regions (UTRs) and an open reading frame (ORF) that encodes a Gag protein. (**B**). *HeT-A* is tandemly repeated at the ends of each chromosome.

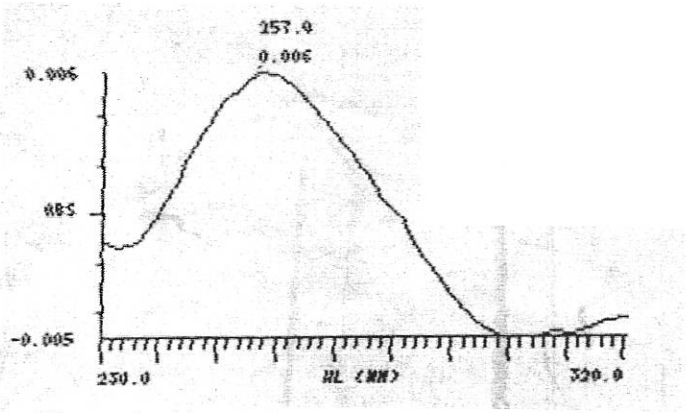


Figure 3: Absorption spectrum obtained for DNA isolated from *D. yakuba*. The x-axis shows wavelength in nm, and the y-axis the absorbance. It is seen that the absorbance maximum for this species is at 257 nm, characteristic of DNA.

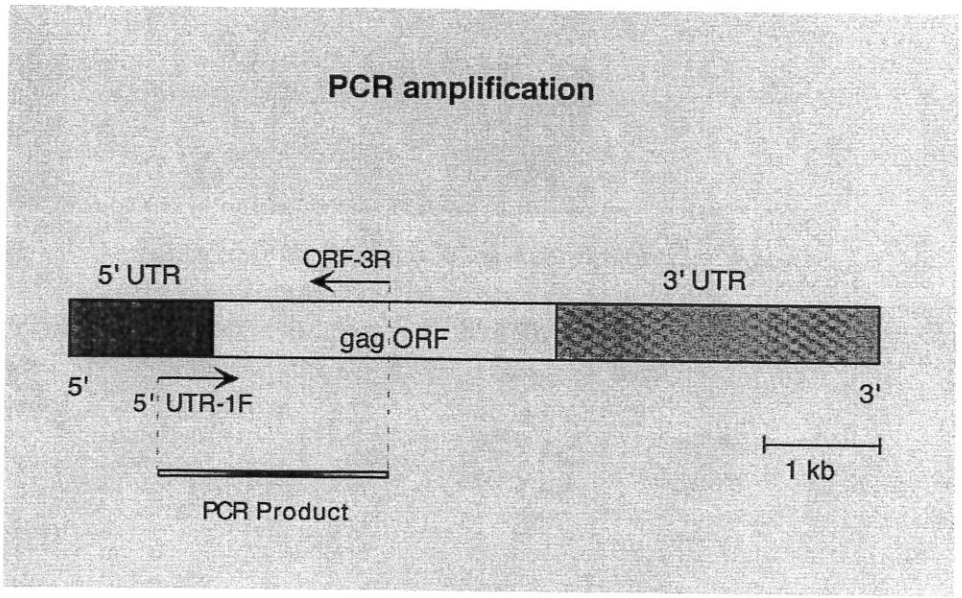


Figure 4: Location of 5'UTR-1F and ORF-3R primers used for PCR amplification of *HeT-A* fragment. Primer sequences are given in Materials and Methods. These primers amplify a fragment of about 1700 bp, corresponding to the first half of the *HeT-A* coding region.

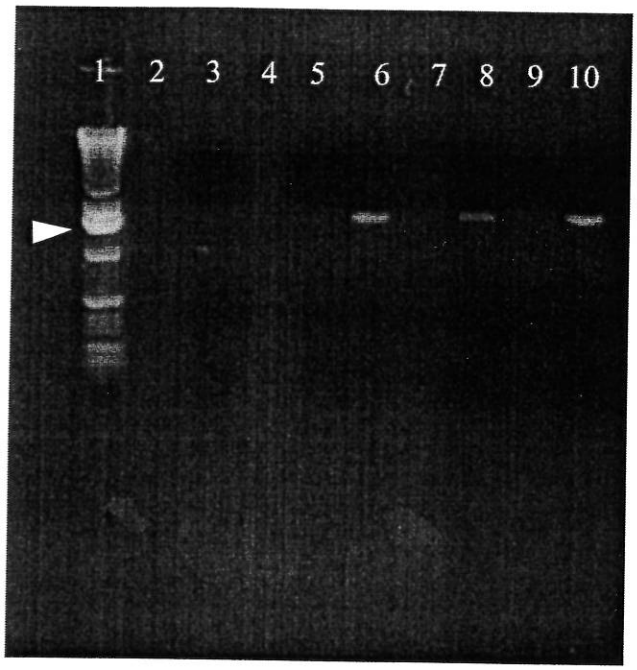
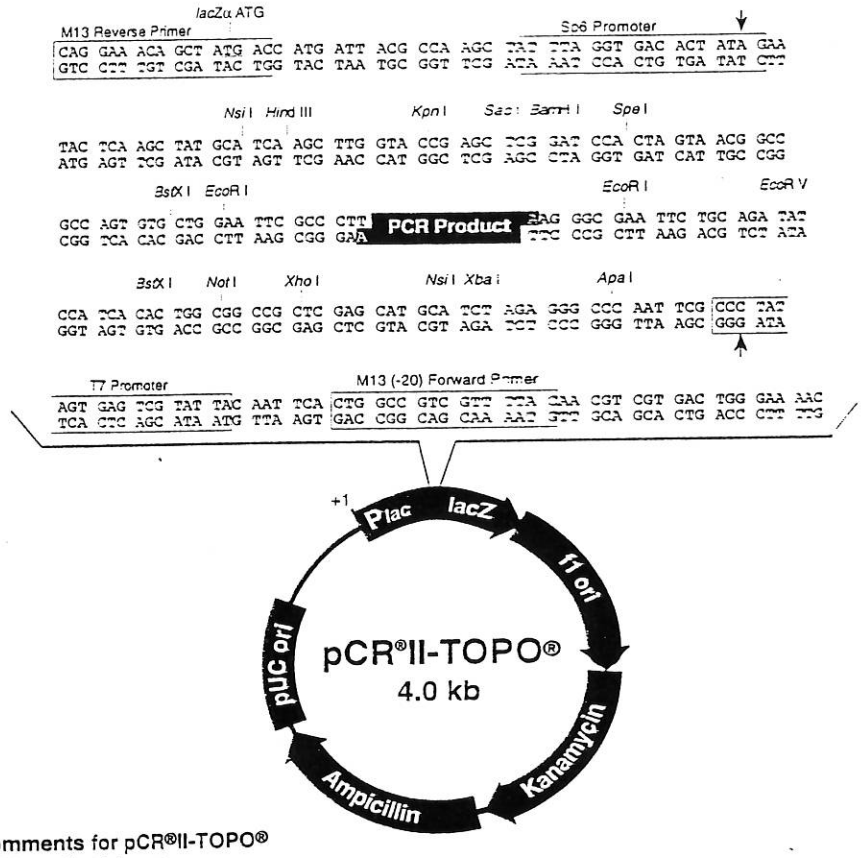


Figure 5: Visualization of PCR products after electrophoresis through 1% agarose gel. Template DNA is from *D. mauritiana*. *HeT-A* fragments are visible in lanes 6, 8, and 10. Lane 1 contains a 1 kb DNA ladder size standard. The arrow head identifies a 1.5 kb fragment.

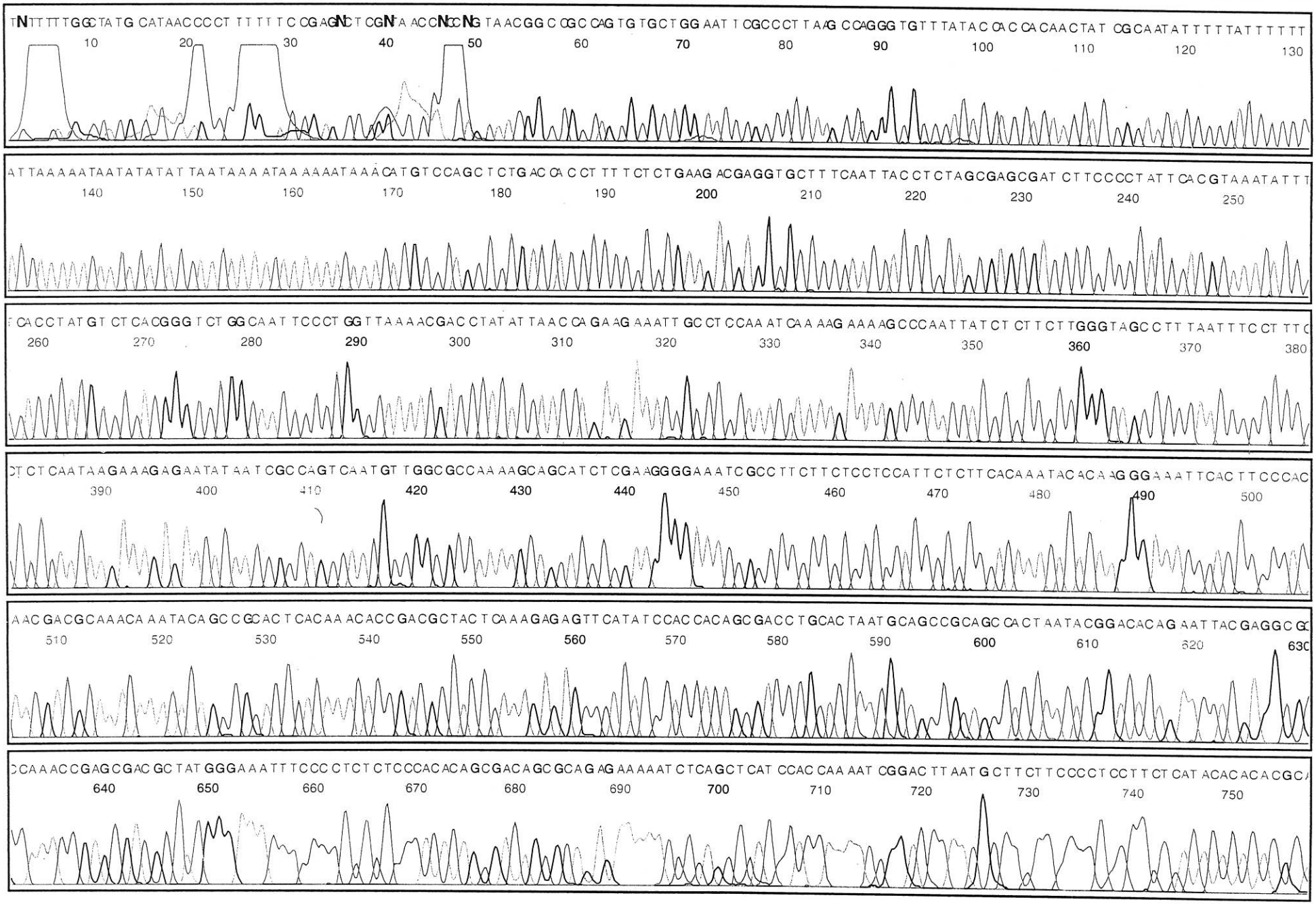


Comments for pCR®II-TOPO®
3973 nucleotides

LacZα gene: bases 1-589
M13 Reverse priming site: bases 205-221
Sp6 promoter: bases 239-256
Multiple Cloning Site: bases 269-383
T7 promoter: bases 406-425
M13 (-20) Forward priming site: bases 433-448
f1 origin: bases 590-1027
Kanamycin resistance ORF: bases 1361-2155
Ampicillin resistance ORF: bases 2173-3033
pUC origin: bases 3178-3851

Figure 6: Topo® Cloning Vector

Figure 7. Chromatogram of a DNA sequencing reaction. This sequencing reaction was primed using the SP6 promoter region of the cloning vector, pCR TOPO. The insert was the *HeT-A* PCR fragment from *D. simulans*. Sequence determination was performed on an ABI Prism automated DNA sequencer. Each nucleotide is tagged by a dye that is then recognized by a laser to record the identity. A, green; C, blue; G, black; T, red. (colors are not visible in this black and white copy).



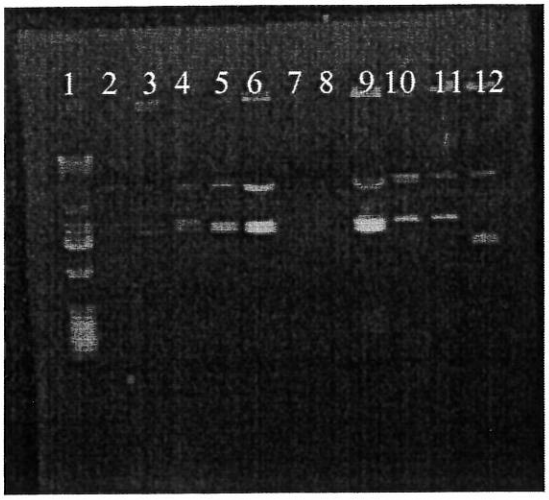


Figure 8: Electrophoresis of products from *Eco R1* restriction enzyme digest of plasmids through 1% agarose gel. *HeT-A* fragments of approximately 1700 bp are visible in lanes 3 through 6 and 9 through 12. Lane 12 appears to contain two fragment, likely due to an alternative restriction site in the *HeT-A* fragment. Plasmid vectors are visible as larger fragments in lanes 2 through 6 and 9 through 12 Lane 1 contains a 1 kb DNA ladder and lanes 7 and 8 are blank.

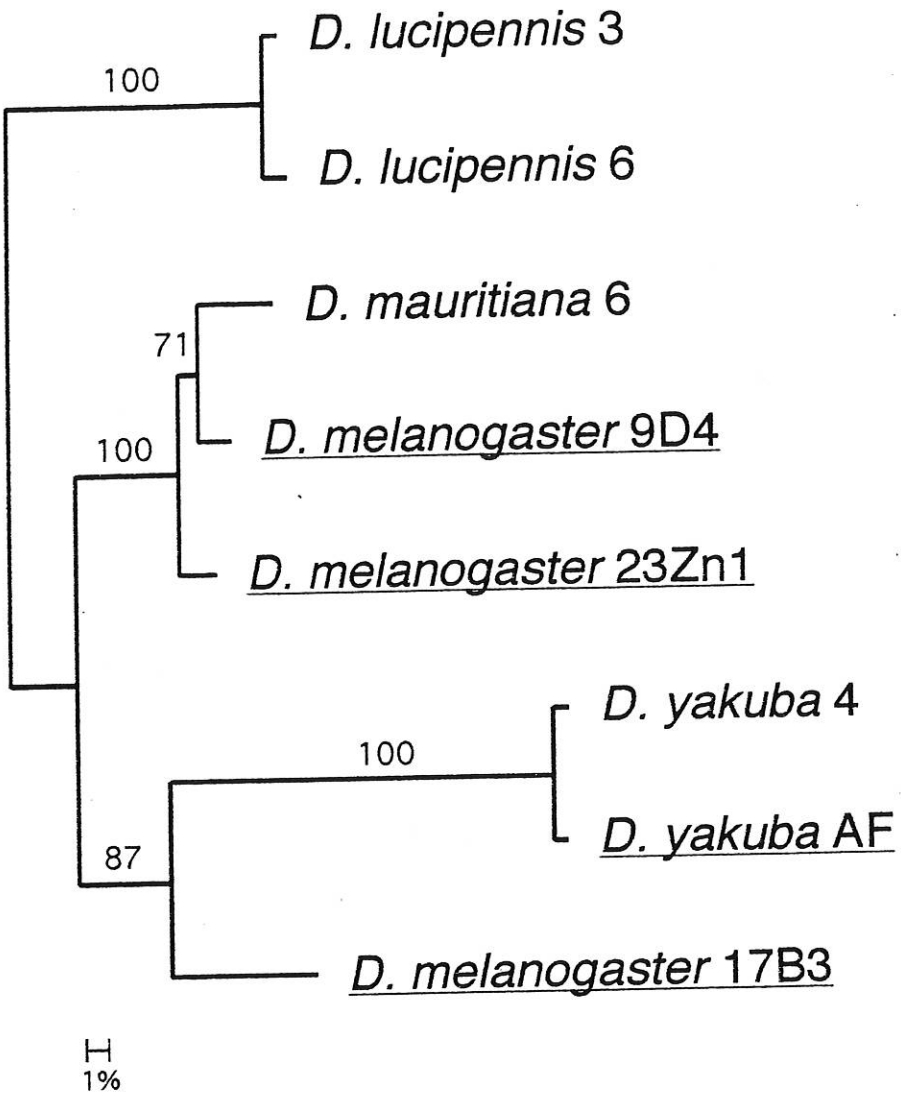


Figure 9. Phylogeny of *HeT-A* sequences as determined by parsimony analysis. Sequence that are underlined are complete *HeT-A* sequences that have been previously published. All others represent PCR fragments generated in this study. This is the single most parsimonious tree recovered, with a length of 875 steps. Numbers above branches are bootstrap percentages, out of 1000 replicates. The scale bar is proportional to the sequence divergence among sequences.

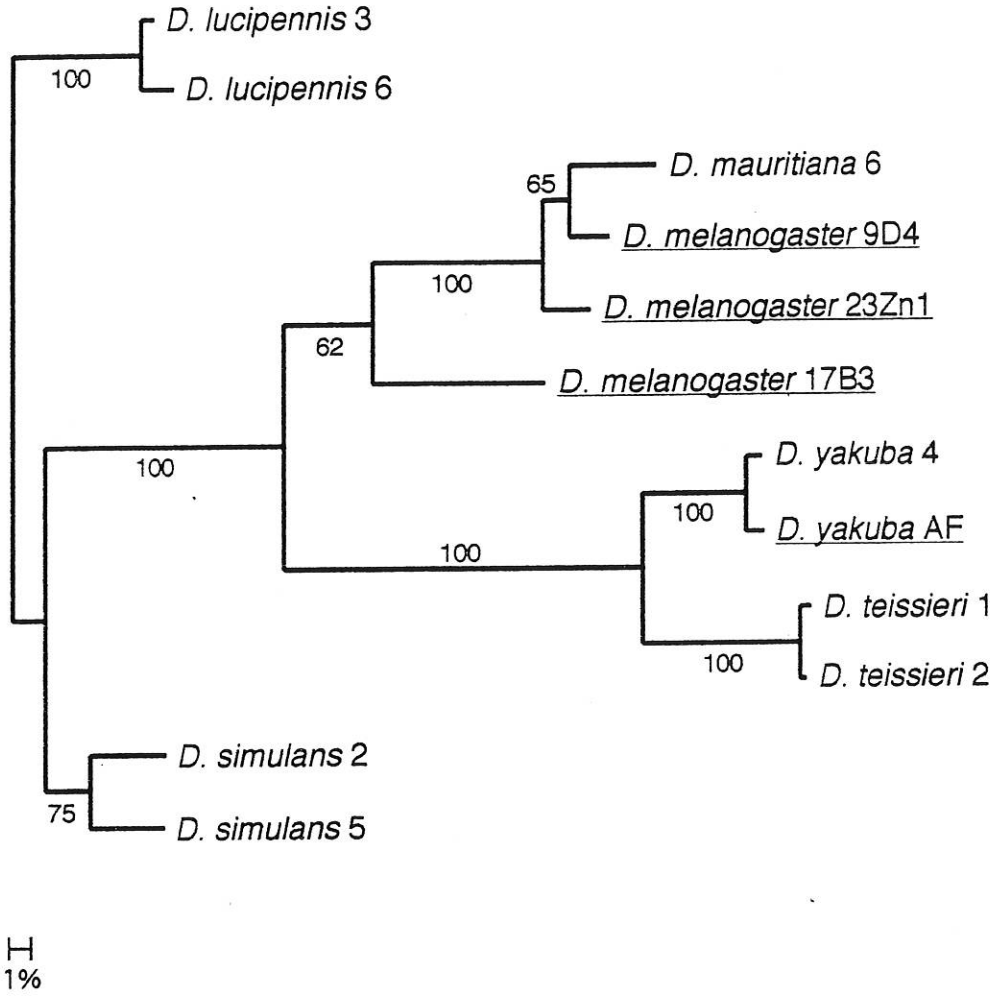


Figure 10. Phylogeny of *HeT-A* sequences as determined by parsimony analysis. Sequence that are underlined are complete *HeT-A* sequences that have been previously published. All others represent PCR fragments generated in this study. This is the single most parsimonious tree recovered, with a length of 1153 steps. Numbers above branches are bootstrap percentages, out of 1000 replicates. The scale bar is proportional to the sequence divergence among sequences.

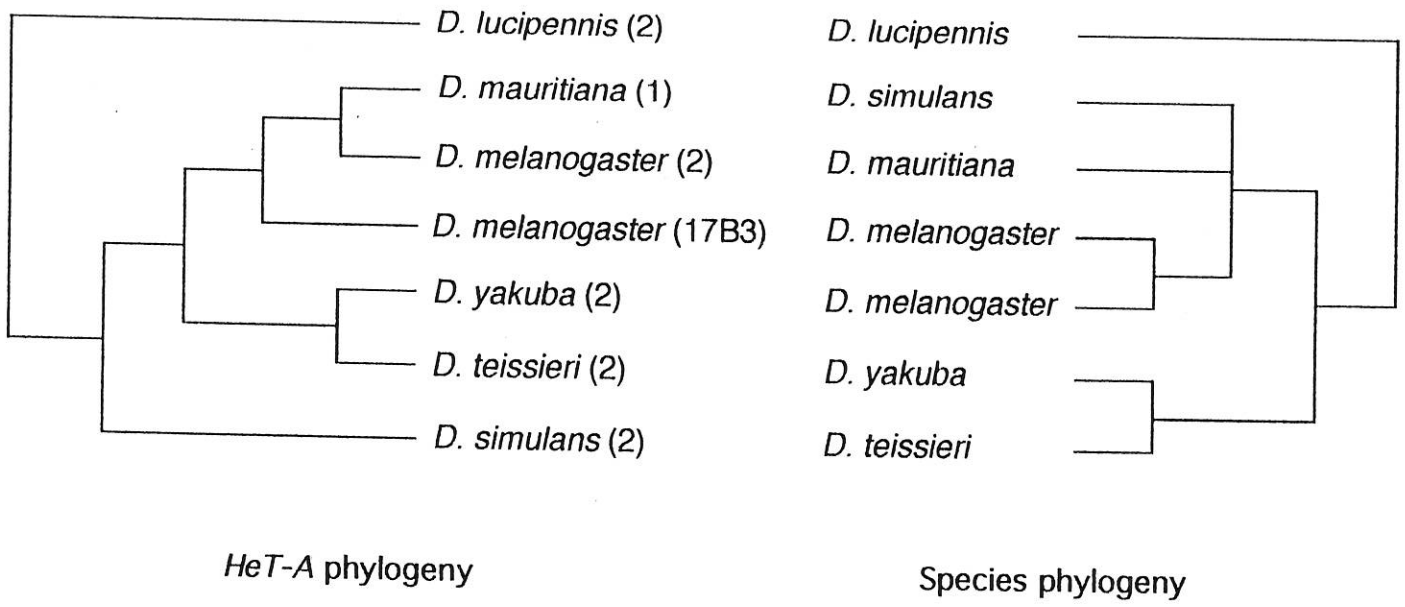


Figure 11. Comparison of *HeT-A* and species phylogenies. The tree on the left was derived from the analysis described for Figure 10. Similar sequences from the same species were collapsed into a single branch. The number of sequences represented in each taxon here is given in parentheses. The *D. melanogaster HeT-A* sequence 17B3 is clearly distinct from the other 2 sequences from this species, which are more closely related to that from *D. mauritiana*. The species phylogeny is a synthesis based on both molecular and morphological data. The relationships among *D. simulans*, *D. mauritiana*, and *D. melanogaster* are unresolved.

The author would like to express deep appreciation to Dr. Jonathan B. Clark, director of the DNA laboratory, primary investigator, and thesis advisor. It has been an honor to receive instruction, inspiration, and direction from so distinguished a scientist. In addition, thanks are due to Dr. Kent VanDeGraaf and Dr. Karen Nakaoka for willingly serving as a supportive thesis committee. Thank you to Dr. Samuel Zeveloff, chair of the department of Zoology, and to other outstanding faculty, staff, and students at Weber State University.