May 2006

Analysis of the Upstream Regulatory Region of the Enhancer of Split m7 gene in Drosophila

Bryanne E. Robson
Connecticut College, berob@conncoll.edu

Recommended Citation

Robson, Bryanne E., "Analysis of the Upstream Regulatory Region of the Enhancer of Split m7 gene in Drosophila " (2006). *Biology Honors Papers*. 1.
http://digitalcommons.conncoll.edu/biohp/1

Analysis of the Upstream Regulatory Region of the *Enhancer of Split m7* gene in *Drosophila*

by

Bryanne E. Robson

A thesis submitted in partial fulfillment
of the requirements for the degree of

Bachelor of Arts

in Biology, Cell and Molecular Concentration

Connecticut College

2006

Approved by ____
Chairperson of Thesis Committee

Date____

Thesis Committee:
Professor Deborah Eastman, Ph.D., Advisor
Department of Biology, Connecticut College

Professor Stephen H. Loomis, Ph.D.
Department of Biology, Connecticut College

Professor T. Page Owen, Ph.D.
Department of Botany, Connecticut College

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

Without Deborah Eastman’s exceptional guidance and encouragement, this research would not have been possible. I would like to thank her for the unparalleled experiences and mentoring, especially outside of lab, she has afforded me since we began working together in 2004. Seeing all she has accomplished makes me proud to be a woman in science.

Beto Zuniga is one of the most patient and benevolent individuals I have had the pleasure to work with. Without his expertise and Dove chocolates, some days in lab would have felt truly impossible.

I would like to acknowledge Stephen Loomis and T. Page Owen for taking the time to read and offer suggestions and corrections on my work. I greatly appreciate their insight, as they have been such admirable individuals throughout my time at Connecticut College.

The support and friendship from Ben Polansky and Morgan Maeder have been incredible. Had these two not shared with me the pain and excitement that comes with a thesis, I would have been lost. Although we might spend all day in lab without speaking, Ben is such a wonderful friend and his realistic perspective has kept me level-headed and saved me from myself many times. I truly admire Morgan and her relentless determination. She is also an amazing friend who will find you in the library buried under research articles and get you the deli ham that you just can’t live without.

I’d like to acknowledge my three best friends Laura Deming, Lindsey Watt, and Merrill Swig for always supporting me and my ambitions, whether crazy or practical, and for picking me up and putting me back on my feet. These three are the backbone of my successes.

ABSTRACT

The Notch pathway is one of the vital signaling pathways used during *Drosophila* development. Present in many organisms and extensively studied in *D. melanogaster*, this pathway serves to transduce signals between neighboring developing cells and inhibits neuronal differentiation by lateral inhibition. The primary targets of Notch are the *Enhancer of split (E(spl))* genes. Although the upstream regulatory regions of the *E(spl)* genes contain biding sites for Suppressor of Hairless, Proneural, and E(spl) proteins, their expression patterns are not identical. There is a hidden complexity in the regulatory regions of these genes that may help explain the conservation of the overall organization of the *E(spl)* complex between different species of *Drosophila*. We are interested in determining what is responsible for this variation in expression, and have investigated these genes using different approaches. First, using the upstream regulatory sequence of *D. melanogaster* and *D. pseudoobscura* as reference sequences, we compared the upstream regions of the *E(spl) m7* gene to *D. simulans, D. sechellia, D. yakuba, D. erecta, D. ananassae, D. persimilis, D. grimshawi, D. virilis, and D. mojavensis* using BLAT (Kent 2002) and EvoPrinter (Odenwald et al. 2005) applications. Second, we isolated and sequenced part of the regulatory region of *D. pseudoobscura m7* to confirm previous published results and gain more insight on the functionality of the region. The differences and similarities in upstream sequences of the *E(spl)* genes are being used as a tool to help further determine if functions of these genes are conserved. This approach will give insight into which regulatory sites are essential to proper *Drosophila* development after millions of years of species divergence.

INTRODUCTION

During development cells undergo changes that make them become more specialized in form and function through a process called differentiation. Differentiation is one of the most important processes that occurs during development because it results in a multitude of different cell types with different functions in the organism. This course of events must be highly regulated and defined so that the cells adopt the correct fate. One of the key signaling pathways used during *Drosophila* development is the Notch pathway. This pathway is present in many organisms from sea urchins to humans (Artavanis-Tsakonas et al. 1999). Notch pathway mutations are associated with different human pathologies. Joutel et al. (1996) found that mutations in the human *Notch3* gene could be the defective protein in patients suffering from cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL). Stroke and dementia come as a result of the hereditary adult-onset disorder CADASIL. Early onset Alzheimer's disease is associated with mutations in the Notch pathway gene, *Presenilin* (for a review see Brunkan and Goate 2005). Although identified in many organisms, the Notch pathway has been most studied in *D. melanogaster*.

The Notch pathway serves to transduce signals between neighboring developing cells and affects differentiation through lateral inhibition and induction. Notch laterally inhibits proneural genes during neurogenesis, so that only certain cells develop into neurons and express the proneural genes and the other cells surrounding the neuronal cell do not express these genes due to Notch pathway inhibition. Type 1 transmembrane ligands, Delta and Serrate, activate the Notch receptor on a neighboring cell (reviewed by

Artavanis-Tsakonas et al. 1999). Signals transmitted through the Notch receptor (along with other factors) influence differentiation, proliferation, and apoptotic actions at all steps of development; however, the *Enhancer of split (E(spl))* genes are the primary Notch targets (reviewed by Artavanis-Tsakonas et al. 1999).

Components of the Notch signaling pathway are illustrated in Figure 1. When the Notch receptor binds its ligand, a conformational change occurs and the proteinase Presenilin is able to cleave the intracellular domain of Notch (ICN), which contains a transcriptional activation domain (TAD). ICN is transported to the nucleus via its nuclear localization signal (NLS). Once inside the nucleus, ICN interacts with the protein suppressor of hairless (Su(H)) (Bailey and Posakony 1995). Su(H) is an upstream repressor DNA binding protein of the *E(spl)* bHLH genes. The mammalian homologue of Su(H) is CBF-1 (Bray and Furriols 2001). However, when ICN is present, these two proteins interact to activate transcription of the *E(spl)* genes (Bailey and Posakony 1995).

Cell-specific activation of certain Notch target genes requires interaction between the ICN and proneural bHLH activator proteins (Cooper et al. 2000). Transcriptional collaboration between the ICN and proneural activator proteins requires specific upstream regulatory region organization of the target genes. Cave et al. (2005) have shown that a specifically oriented pair of Su(H) sites are necessary for Notch-proneural cooperation, because this DNA construction allows for cooperative interactions between Su(H) and proneural proteins.

The *E(spl)* genes are responsible for the inhibition of proneural gene expression and thus promoting these cells to adopt an epidermal fate. The E(spl) proteins contain a basic domain that binds DNA and a Helix-Loop-Helix domain that allows for homodimerization and heterodimerization with other bHLH proteins (Delidakis 1992). These proteins act as repressors to turn off the expression of proneural genes: *achaete* (ac), *atonal (ato)*, *scute (sc)* and *lethal of scute (lsc)* (Brody 1996). These genes code for proneural proteins and are essential for proper specification of the peripheral nervous system sensory organ precursors (SOPs). Different proneural genes specify different SOPs. The *achaete-scute* (*ac/sc)* complex (*achaete, asense, lethal of scute, and scute*) specify precursors of the external sense organs (Campuzano and Modolell 1992), whereas *atonal* is involved with chordotonal organ precursors and R8 photoreceptors (Jarman et al. 1993). To achieve different developmental functions, different proneural proteins have to activate common neural target genes and specific neuronal subtype specific target genes (Powell et al. 2004).

Daughterless (*da*) is also involved in neural differentiation (Brody 1996). The expression of this protein is necessary for the expression of neuron specific genes. Proneural proteins: ato, sc and da bind to *E(spl)* and *achaete* promoters and activate their transcription. The E(spl) proteins function to antagonize proneural proteins consequently turning off expression of the genes that become activated by proneural genes (Brody 1996). *Daughterless* is essential in controlling the expression of genes needed in the peripheral nervous system SOPs; without Da, proper development would fail to occur (Brody 1996). The proneural proteins atonal and scute heterodimerize with daughterless to then interact with E-box DNA binding sequences. These heterodimers are necessary for promoting transcription (Brody 1996).

Proneural proteins prefer the E-box DNA binding site with the variant CAGSTG, where S is either C or G (Powell et al. 2004). Different DNA binding proteins prefer E-boxes containing certain bases flanking either side of and contained in the E-box sequence __CANNTG__ (Jennings et al. 1999). The Sc/Da heterodimers prefer the sequences **G**CAGSTG**G** and **G**CAGSTG**T**. Powell et al. (2004) have investigated the preferences for Ato/Da heterodimers. Chien et al. (1996) proposed the DNA contact residues in the bHLH domains of Sc and Ato are completely conserved, suggesting that distinct binding sequences for Ato and Sc might not exist. However, Powell et al. (2004) determined that the difference in the E-box sequence for atonal and scute is due to a difference in flanking bases; this research was conducted by monitoring the regulation of the gene *Bearded (Brd).* Chien et al. (1996) determined that different interactions with "specificity cofactors" could be a major determinant of proneural specificity as well.

In determining the difference in E-box consensus sites for Ato/Da and Sc/Da, Powell et al. (2004) found that the major change arises in the 5' flanking base of the E-box sequence when a G changes to an A or T. They determined the Ato/Da binding site to contain the sequence AWCAKGTGK (W=A/T, K=G/T) whereas the Sc/Da consensus sequence is GCAGSTGK. For correct enhancer function, proneural proteins must interact with subtype specific DNA binding factors, as well as bind to the correct enhancer sequence with the right flanking bases.

Regulation involving protein interaction with N-box binding sequences is also significant to the upstream regulatory region. E(spl) proteins have shown an affinity for binding the N-box consensus sequences CACNAG and CACGYG (high-affinity) (Kramatschek and Campos-Ortega 1994). Kramatschek and Campos-Ortega (1994) have shown that *E(spl)* transcription is activated in the neuroectoderm, and repressed in neuroblasts because of distinct interactions of proximal regions in the promoters that contain N-boxes and E-boxes. Nellesen et al. (1999) have noted that in both *mδ* and *m7* N-boxes with the sequence CACAAG[1] overlap Su(H) sites that are members of the paired Suppressor of Hairless sites. Only five base pairs separate a Su(H) site and a CACGAG N-box in *mγ*. Since bHLH repressor and Su(H) binding sites are frequently concurrent upstream of bHLH repressor-encoding genes, it was hypothesized they may possess the same transcriptional regulation (Nellesen et al. 1999). It was proposed that Su(H) could displace a bound bHLH repressor complex as the gene is transcriptionally activated by the Su(H) protein (Nellesen et al. 1999).

The conservation of the *E(spl)* locus has been investigated to determine the gene functionality throughout *Drosophila* species divergence. Maier et al. (1993) found that the *E(spl)* region is well conserved between *D. melanogaster* and *D. hydei* and that the seven *E(spl)bHLH* genes were derived from a single ancestral gene before the divergence of the two subgenre *Drosophila* and *Sophophora*. Figure 2 shows the *Drosophila* species tree. Neither arrangement nor number of *E(spl)* genes has changed during approximately 60 million years of evolution. Due to this high degree of conservation, Maier et al. (1993) concluded that the *E(spl)bHLH* gene products may have different specificities and are not redundant. More results (Ligoxygakis et al. 1999) show that the *E(spl)* proteins have partially redundant functions, yet have evolved distinct preferences in implementing diverse cell fate decisions.

There are seven *E(spl) bHLH* genes: *m5, m7, m8, mβ, mγ, mδ,* and *m3*. Wech et al. (1999) show the unique expression patterns of the *E(spl)* genes during embryogenesis, and Cooper et al. (2000) have shown the distinctive expression patterns of the *E(spl)*

---

[1] In the case of *D. melanogaster m7,* the N-box sequence overlapping the paired Su(H) site is the sequence CTTGTG (the reverse complement of CACAAG).

genes in the eye/antenna, wing and leg imaginal discs. Ectopic expression of these genes also illustrated the fact that these genes have their own independent functions (Ligoxygakis et al. 1999). Specifically, Ligoxygakis et al. (1999) found the E(spl) m7 and E(spl) m8 proteins are most successful at suppressing sensory organ fates, E(spl) mβ protein is the best at antagonizing vein fate, in compliance with its expression in intervein areas, E(spl) mγ protein is active at the wing DV boundary to repress the *vg* quadrant enhancer and to activate *cut,* and E(spl) mδ protein is most successful at suppressing the photoreceptor R8 fate and is strongly expressed in the morphogenetic furrow of the eye. The specialization of E(spl) m3 and E(spl) m5 proteins is not clear, because research has primarily focused on the wing and eye discs where these genes are not usually expressed.

To gain a better understanding of the function and regulation of the *E(spl)* genes, we are interested in comparing their expression in various *Drosophila* species - mainly *D. melanogaster* and *D. pseudoobscura*. *D. pseudoobscura* was selected because it was the first species to be sequenced from a distant subgroup from *D. melanogaster.* For genes illustrating a high level of similarity between species, it is expected that their expression patterns will be similar as well; conversely, for species that differ significantly, it is predicted they will have different expression patterns of the same gene.

Comparison of the E(spl) protein sequences has shown strong similarities that suggest that this gene sequence has been highly conserved throughout the evolutionary divergence of these two *Drosophila* species. Protein sequence alignments between *D. melanogaster* and *D. pseudoobscura* were constructed for E(spl) mβ, mγ, mδ and m7, and these comparisons led to the investigation of the upstream regulatory regions of the *E(spl)* genes.

Upstream regulatory regions are important components in the regulation of transcription of the corresponding genes. Without this specific control of the genes, incorrect cell fates could occur. The upstream regulatory regions of the *E(spl)* genes have been analyzed and all the genes contain binding sites for Su(H), Proneural proteins, and E(spl) proteins, yet each gene shows a different expression pattern. This suggests that other regulatory DNA binding proteins must also play a role.

The upstream regulatory regions of eleven species of *Drosophila* were analyzed using ClustalW (EMBL-EBI 2005) alignment tools as well as BLAT (Kent 2002) and EvoPrinter (Odenwald et al. 2005) alignment technology. *D. melanogaster* and *D. pseudoobscura* served as reference sequences when aligning the upstream regulatory regions against the species *D. simulans, D. sechellia, D. yakuba, D. erecta, D. ananassae, D. persimilis, D. grimshawi, D. virilis, and D. mojavensis.*

The Blast-Like Alignment Tool (BLAT) for DNA (Kent 2002) serves to align a reference sequence with another species genome which is stored in its database. It is able to locate sequences of 95% or more similarity; it finds areas of possible homology, loads the genome into memory and creates a meticulous alignment.

The EvoPrinter program is used in assessing evolutionary divergence of both coding and non-coding (upstream regulatory) regions of genes. One is able to use the program when comparing a number of species to discover very well-conserved sections in a group of orthologous DNA sequences in a process known as phylogenetic footprinting (Odenwald et al. 2005). Once a sequence is selected to be analyzed, it is entered into the BLAT search engine and aligned with the desired species. The BLAT outputs are inserted into the EvoPrinter input areas and an EvoPrint is generated. A maximum of eight other BLATs can be selected in the generation of an EvoPrint. Instead of producing an alignment with nucleotides interspersed with gaps, the EvoPrinter output only lists the reference sequence with highlighted areas of conservation.

Comparisons of the upstream regulatory sequences of each gene allows further analysis of the similarities and differences among *Drosophila* species and their mechanisms of transcriptional regulation. These comparisons will be used as a starting point in analyzing future *in situ* hybridization and cell culture reporter assays involving *D. pseudoobscura* to determine whether significant differences in upstream binding regions among species still produce the same expression patterns among all *Drosophila* species. The upstream sequence differences are being used as a tool to determine if expression patterns of the *E(spl)* genes will yield different patterns among the different species of *Drosophila.*

MATERIALS AND METHODS

Alignments of E(spl) proteins from *D. melanogaster* and *D. pseudoobscura*

The VISTA Genome browser (Couronne et al. 2003) was used to determine the DNA sequence of *D. pseudoobscura E(spl) mγ,* and *m7* genes. The gene sequences were translated into protein sequences using the ExPASy translate tool (ELG 1999), and the *D. pseudoobscura* protein sequences were aligned with the *D. melanogaster* protein sequences (Deilidakis 1992).

Comparison of *E(spl) m7* and *mγ* regulatory regions from multiple species of *Drosophila*

Upstream regulatory sequences for the *E(spl)* gene *m7* were obtained from: the VISTA Genome browser (Couronne et al. 2003) for *D. pseudoobscura*, Macdonald and Long (2005) for *D. melanogaster* and *D. simulans,* and Nellesen et al. (1999) for *D. hydei* (*mγ* only). Using the ClustalW (EMBL-EBI 2005) sequence alignment program, *E(spl) mγ* sequences for *D. pseudoobscura*, *D. melanogaster, D. simulans,* and *D. hydei* were aligned and analyzed. Binding sites (Table 1) for E-boxes, N-boxes and Su(H) were identified for all four of the species based on the previous research by Nellesen et al. (1999) and Powell et al. (2004). Figure 3 illustrates the upstream regulatory region components of the *E(spl) mγ* and *m7* genes. The same was completed for *E(spl) m7,* except no sequence was available for *D. hydei* at the time of alignment.

EvoPrinter analysis of *m7* upstream regulatory sequence in multiple *Drosophila* species

***D. melanogaster* reference**

The *D. melanogaster* sequence used in the BLAT (Kent 2002) program was obtained from MacDonald and Long (2005) and entered as the reference sequence. This sequence was BLATed individually against *D. simulans, D. yakuba, D. erecta, D. ananassae, D. pseudoobscura, D. grimshawi, D. virilis, and D. mojavensis.* The results of the BLAT (Kent 2002) gave a relative percentage of agreement for each species BLATed against the *D. melanogaster* sequence. These relative percentages of identity were used to develop a system to judge the relative relationships of the divergences of the *Drosophila* species. The sequence used was (upstream sites annotated):

AAAATGGTTTCTTTTTTTAGGAATTTTTTAAACATCATAACCAGACAAGAACATCTTTTTCAAATATTTCTTTATTTCAC
CCATCTGGGTACAGTAATCCCATACATAGATATTCTGTAAACTGCAACCTTCACGAACCTTAATGTTTTCCAAGACTCAG
G**TTCCCAC**TCACCGTCGCAGCTAACAAAATTATCCGTAGCAAGTAGAAAACACCTTCCACAAAAGTCCCTATACTCGGAC
TATATACGCCTAGAAAAACCTCTAGGAACAAATGTACCTTTCTAACATAGCAGATAGTTTAGGAAAAATTCTTTACAAGT
AGAAATACATCGAGAAATCCATAAACACTTACTCTAAAACCTGCAAGTCGTAAAATTGCACGCATGTTCCGTTTGGTAAA
ACCCCAAAAATCACATTCGAAAACTCACA**TTCCCACG**ACTTTGAATAGAGGTTCTCAAATTAACCGTAGCAAGTTGAGAA
CAGGCAACAAAAAATCGGCAAGATACTGATTTCCAAACCTGCGAGTCCCAAAATTGCACACATCTTCGCCTTGGAAAACC
CCATTTCCGACATCCCAAAACTCGCA**TTCCCACG**ATCCCTACAGAAATTGCTAAAACTTTGTCTAGCAAGTAGATAACAG
TATGCGTAAATTCCGTAAACCTTAAATTTCACGTAGTTGGAAAAGCCTTGAAGAATTCATTCAAGCAAGGCACCCTCAAA
ATTCCCGCATGTGTTTCATCGGTTTGGCCCACTGTGCCATGTGCATAAGGGCTCGGAACTCGGAGCTCGGACTCGGGACT
CGCACACGGGCTACCTCCGATTATAACTTATAACACCAACGAGCGAGAAAAT**CTTGTGGGAA**ACTTGAGGCAAAGTGT**T**
TCCCACGATTCGAATGTCAGTGGAAAGGAAAACAGAGAAGCCAGCATGTGCAACCACCGTCGCACCACCACCACTGCACT
ATCGGCATCTTCATCCCCTGGCCGAGACTCACTTAACACCGAAG**CACCAG**TGGCCATAGCAATAGTACTGAGAACATATT
TAAAGTTTTCTAGATAATGTTTATATTTTCAAAATCACTACTATTAAGAAGTGTGCTAATTTAATTTTTTAGGTGTAAAC
GTATAACATTATGTTAATTTCTTTTTGTTTAATAAAACGACATTTAAGTGCTTTAAGGGAAACCCATTTGAAGTAATTGT
TAAAAACGTAACGTACTTCTTAGGTATATAACAACAAATTTGTATATTCGCTCGAAGAACTATAGCTGCTTTAAACACAT
TTCTCATACTATTTTTACCGCAGGTGTGTTCCAAACTTGTTTGTGGAGAAAAACCCACCACCGAAAATGCAGGCGAGAGG
ATCGTGTCCCAGTGACTGAGTAAAAGTAGTTTGCTCCGCAGGTGGTGGTTCTTCGGAGGCTCCGCAGCTCTCTCTTCTCT
CTTTCTCCGAGCAGCGTGCGCTGAGCACAGACGCTCTCGCGGCGG**CACGTG**CAGCTATAAAAGCAGCGGTAACCGGAGAC
GAATGCAACATTTCGAACGCAATCGTCGTGCAGTCAACATCTCAGGATAACCTCCTCCGAAAGATCGAAGAAGTTCTTCA
ACAAGTCGTTCTTCCATTTACCTAAACGCAACAAATAAACAAAACAAACACACAATGGCCACCAAATACGAGATGTCGAA
AACCTATCAGTACCGCAAG

For both reference species **orange** sequences indicate Su(H) binding sites, **blue** indicates E-box binding sites and **pink** indicates N-box binding sites. The TATA box is highlighted in yellow to help show relative binding site locations.

The BLAT result from each species was entered into the EvoPrinter (Odenwald 2005) program in the following order: *D. simulans, D. yakuba, D. erecta, D. ananassae,* *D. pseudoobscura, D. grimshawi, D. virilis, and D. mojavensis.* The first two species (*D. simulans* and *D. yakuba*) were selected to generate an EvoPrint. Next, *D. erecta* was added to the list of selected species, and another EvoPrint was generated. Each of the species was added in order until all were used.

A second, longer reference sequence for *D. melanogaster* was also obtained from the FlyBase database and individually BLATed against *D. melanogaster, D. simulans, D. sechellia, D. yakuba, D. erecta, D. ananassae, D. pseudoobscura, D. persimilis, D. grimshawi, D. virilis, and D. mojavensis.* The sequence was (including annotated sites; capital letters indicate coding region):

```
tcaagcttcaggcctgaaaaaattaaactctttttatttatttatttttttttttttttgtttttcttttttgcatttcatgcgatttt
attgcgcatgcgtggcgcgcagcaagcttctgttggggccctttggcttttgtttttagctgcccgcgtctgtcgcgcgtttggcgtgc
caagcatgcgcacacatattcaattacacggccagcaacgtcggttgcccagcagcagcagcgattatagcgatgcattgacaataagc
caaccccgatgcgtgcccaaatggttgcccctacttaataaaaagtggaaaaaacaacaacaaaagcccggagctggccatgtgcagct
cgagaaatttctcacagtcgccagaaaaattgtaacggcccgttgttcagcacactctcagcacgcagtgcactgcacacacacactt
cctgccgcagccagcgttcctcaaaccggccacgtatattgtttatatagttaacgaatgcgatacatttaatctttagtgtgtgtcaa
tttgccgtacagtcatttagtttgcatgtccctcgtcgatccatcgaacaatcatccctttttctgggacctgggtcccacgatgatga
tgtcccgacgtgagcaggaataccatacgcgtatatgtatgccacttagcgatggacgatcccaaccgatgccatccgaaacgaaaccc
agggcccaagtgttcaacagaagctacaattcagcattcataattgacaagagcaggagcgagggagagggcgatatgtgtagggccca
cgatggacaagaaaaaattaaataaaataaagggcaggtggtagtttttcagcactacccgtgactcgcactagcagctcggtcagat
ttacagtggttcgaatggaacgaaaattggttgaaaacaatttagggtcttgtcacaatctttagatgcttttagctacaaactacaaa
atggtttctttttttagaaatttttttaaacatcataacctgacaagaacatcttttccaaatatttctttatttcacccatctgggtac
agtaatcccatacatagatattctgtaaactgcaaccttcacgaaccttaatgtttttcaagactcaggttcccactcaccgtcgcagc
taacaaaattatccgtagcaagtagaaaacacctttcacaaaagtccctatactcggactatatatgcctagaaaaacctctaggaaca
aatgtacctttccaacatagcagatagtttagtaaaaattctttacaagtagaaacacatcgagaaatccataaacacttactctaaaa
cctgcaagtcgtaaaattgcacgcatgttccgtttggtaaaaccccaaaaatcacattcgaaaactcacattcccacgactttgaatag
aggttctcaaattaaccgtagcaagttgagaacaggcaacaaaaaatccgtaagatactgatttccaaacctgcgagtcccaaaattgc
acacatcttcgccttggaaaaccccatttccgacatcccaaaactcgcattcccacgatccctacagaaattgctaaaactttgtctag
caagtagataacagtatgcgtaaattccgtaaccttaaatttcacgtagttggaaaagccttgaagaattcattcaagcaaggcaccc
tcaaaattcccgcatgtgtttcatcggtttggcccactgtgccatgtgcataagggctcggaactcggagctcggactcgggactcgca
cacgggctacctccgattataacttataacaccaacgagcgagaaaaatcttgtgggaaacttgagggcaaagtgtttcccacgattcg
aatgtcagtggaaaggaaaacagagaagcagcatgtgcaaccaccgtcgcaccaccaccactgcactatcgcatcttcatcccctgccg
agactcacttacaccgaacaccagtggccatagcaatagtactgagaacatatttaaagttctagataatgtttatattttcaaaatca
ctactattaagaagtgtgctaatttaattttttaggtgtaaacgtataacattatgttaatttctttttgtttaataaaacgacattta
agtgctttaagggaaacccatttgaagtaattgttaaaaacgtaacgtacttcttaggtatataacaacaaatttgtatattcgctcga
agaactatagctgctttaaacacatttctcatactatttttaccgcaggtgtgttccaaacttgtttgtggagaaaaacccaccaccga
aaatgcaggcgagaggatcgtgtcccagtgactgagtaaaagtagtttgctccgcaggtggtggttcttcggaggctccgcagctctct
cttctctctttctccgagcagcgtgcgctgagcacgagacgctctcgcggcggcacgtgcagctataaaagcagcggtaaccggagacg
aatgcaacATTTCGAACGCAATCGTCGTGCAGTCAACATCTCAGGATAACCTCCTCCGAAAGATCGAAGAAGTTCTTCAACAAGTCGTT
CTTCCATTTACCTAAACGCAACAAATAAACAAAACAAACACACAATGGCCACCAAATACGAGATGTCGAAAACCTATCAGTACCGCAAG
GTGATGAAGCCCTTGCTGGAGCGGAAGCGTCGTGCCAGGATCAACAAGTGCCTGGACGAGCTCAAGGATCTGATGGCCGAGTGCGTTGC
TCAGACTGGCGATGCCAAATTCGAGAAGGCCGACATCCTGGAAGTCACCGTGCAGCATCTCCGCAAGCTGAAAGAGTCTAAGAAGCATG
TTCCGGCCAATCCCGAACAGAGTTTCCGTGCTGGATACATCCGGGCAGCCAACGAAGTGTCCCGCGCCCTGGCCTCCCTTCCCCGAGTG
GATGTGGCTTTTGGAACCACACTGATGACCCATCTGGGCATGCGTCTCAACCAACTGGAGCAGCCCATGGAACAACCGCAAGCCGTCAA
CACTCCACTCAGTATCGTCTGTGGATCCAGCTCTAGCTCCAGCACCTACTCCAGTGCCAGCTCCTGCTCCTCCATCAGTCCCGTTTCCA
GTGGATATGCCAGCGACAACGAGTCTCTGCTCCAGATCAGCAGCCCCGGACAGGTGTGGCGTCCCTGGTAAACAAAGATCTACTTGGAA
GGACTGAATTCCCGACGGG
```

This longer sequence was used to generate EvoPrints with individual *Drosophila* species by entering one BLAT alignment at a time. *D. melanogaster*'s BLAT was entered into the EvoPrinter program and used in the generation of each EvoPrint. Selecting the BLAT result from one of the species: *D. simulans, D. sechellia, D. yakuba, D. erecta, D. ananassae, D. pseudoobscura, D. persimilis, D. grimshawi, D. virilis, and D. mojavensis* at a time, ten EvoPrints were generated that aligned *D. melanogaster* with one other species.

***D. pseudoobscura* reference**

The *D. pseudoobscura* sequence used in the BLAT program was obtained from the VISTA genome browser (Couronne et al. 2003) and entered as the reference sequence and BLATed individually against *D. melanogaster, D. simulans, D. sechellia, D. yakuba, D. erecta, D. ananassae, D. persimilis, D. grimshawi, D. virilis, and D. mojavensis.* The sequence used was (including annotated sites):

TGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCATGTCCCTCGACCCATCCGTATCGTATCCCTGT
ACCTATACATATCCCTCACTGGGTCCCACGATGATGATGTCCCGACGTGCCCGGGCATGCCACTTAGCGTTCCCTTCTCTGTTGCCCCT
GGCCCTGCCTTGGCCCAAGTGTTCAACAGAAGCTACAATTCCAGCATTCATAATTGACAAAACAGACCTGCAGGAGGGAAGGAGACGGA
TAGCGGGAGCGGGACAGAGAACCTATTGGGCCCACGACTGACAAGAAAAAAAATTAAGTAAAATAAAGG**GCAGGTG**GTAGTTTTTTTC
ACCACTCTCCATGACTCGCACTCGGAGCACGGTCAGATTTACAGTGGTACGAGTGACTGAAAAACTAGTGGCGAATTGTTTGTGAAGGA
TATTTATAGAATTATAAATTTAGTACTATTAAATATGTACTTCCTTCCTATAGAAATATGAAAAAGAAATAAAAAATGCACTGATATTG
ATACCATCCTGACCAATAAAAAAACATTCCCATTAACTTCCCTTAGCAACATCTCGTACGGAAATACCAGAATTCCTTAACGTTTCCAC
CCAAAGAAAACCTGCAATGTCAAACCTAGCACAAAACTCCCTACAGTTCACACACTACGCCTAGAAAGTCCCTAAATCAGTGTCCCTGG
CAAATTACAACGCATCAACCTTACAAGACCCACCTTATTTCCAAAACTCGCA**TTCCCACG**CACCGTTCCAGAGTCCGGAAATTAACCAA
AGCAAGTAGAACACAGGCCACCTTTAAATCCGTAGAATACTGAAATATTCACAGGTCCCAAACAACACGTACATAGATGGAAAACCTTC
ATTCAGTCGGTTCAGTAGGTATCCAGAAATACCCACATGTGTGCAAATTGCACCGCATACACCTTCCCAGACTCACCTTA**TTCCCACG**C
ACCGTCACAGAGTCCGGAAATCAAGTAGAAATCCGTATTCGCAGGTCCAAAAAAACACGTACATACATCGGTAGGTGGACTGCCTTCAT
TCAGTCGGTCGCAGAGATTCGGAAATTCCCGCATGTGTGCAATGGTTTGACCCACTGTGCCATGTGCATAAGGGCTCGGACCTCGGGGA
CTCGGATCAGACACGGACTCGGGGTTCGTTCGCTCCCTACCTCCGATTATAACTTATAACACCAACGAGCGAGAAAAAT**CTTGTGGGAA**
ACTTGAGGGCAAAGTGT**TTCCCACG**ATTCGAATGTCAGGCTAAAAGAAAACAGAGGAGAGCACACACAAGAAAACGCACA**CACGTG**CAT
CGGCCGCATGTGCAACCAACCACCACCGCCAAGAGCAGCAGCAGCAAGCAGCAACAACAACAGCTATCCACCAACACAGGAATGGAGGA
GACACACCATCACATCATCTTCATCCCCTGCCAAGACGACTTGTTTGTCGAGAAAAACCCACCACCGAAAGCAGGCGAGGATCGTGGCT
GTCGGCCCGAGTAACGATTCGTAACGACGACACGAACGACAATCGGCTCGACCCGACCCAGTGACTGACG**GCAGGTG**GTCGCAATGGTG
GTGCTCGCTCCGGTGGTGCCTCTCTCGTTTCCTGCGCATGCGCGCTCTCTCTGCGCGCGCTGAGCACGAGACGACTCTGCTGCGG**CACG**
TGCAGCTATAAAAGCGGCGGTAACTGGAGACGAGAGCAACATTCTCAACAAATTCGACGTGCAGTGAACACACACAAGACAACCGGATC
CCAGCCAGATCTCGCTCAACTTCTTCACAATTTTACGTCAAGAAAACACACATCTAAGCTACAATGGCCACCAAATACGAAGTTTCCAA
GACCTACCAGTACCGCAAGGTGATGAAGCCTCTGCTGGAACGCAGGCGTCGGGCCCGCATCAACATGTGCCTGGACGAGCTCAAGGATC
TGATGGTCCAGTGCACGGTCCAGAGCGGTGATGGCAAATTCGACAGGGCCGACATACTGGAGGTCACTGTGGACCATTTGCGCAAACTG
AAGCAGGCCAGGATCGAGGCCACCGCTGCCGCAGCAAAAGCCACCACCAACACTACCCCAGAGCAGAGTTTCCGCGATGGCTTCATCAG
GGCCGCCGACGAAGTCTCCCGGGCACTGGCTTCCCTACCCAATGTGGATGTCGTCTTTGGCACCCATCTGATGACACACCTCGGCTTGC
GCCTCAACCAATTGGAAACGCCAACTGCAGCACCAAAGCCAATGAATGCCCCGCTCTCCATCATCTGCGGCAGCGGCGGCAGCAGCAGC
AGCAGCAGCAACTGCAGCTCCAACAGCCGGGAGGCCTGCAGTCCCGTGTCCAGCGGCTACTGCAGCGACAGCGATTGCAGTGTCTCCTC
CACTCAAGCGGCCCAGAATCTGCTCCAGATCAGCACTGGTCAGGTTTGGCGCCCCTGGTGAATCTATAGATTGGATTGGCCAACACCCT
CATCGGGCAATCG

The BLAT result from each species was entered into the EvoPrinter program in the following order: *D. melanogaster, D. simulans, D. yakuba, D. erecta, D. grimshawi, D. virilis, D. ananassae,* and *D. mojavensis.* The first two species (*D. melanogaster* and

D. simulans) were selected to generate an EvoPrint. Next, *D. yakuba* was added to the list of selected species, and another EvoPrint was generated. Each of the species was added in order until all were used.

A second set of EvoPrints was generated, and the BLAT results from the *D. pseudoobscura* sequence from the VISTA genome browser were used again. *D. pseudoobscura*'s BLAT was entered into the EvoPrinter program and used in the generation of each EvoPrint. Selecting the BLAT result from one of the species: *D. melanogaster, D. simulans, D. sechellia, D. yakuba, D. erecta, D. ananassae, D. persimilis, D. grimshawi, D. virilis, and D. mojavensis* at a time, ten EvoPrints were generated that aligned *D. pseudoobscura* with one other species.

The flanking bases of the E-box sites in all species were further examined following the EvoPrinter alignments.

Confirmation of sequence in D. pseudoobscura upstream regulatory region

The *D. melanogaster E(spl) m7* upstream regulatory sequence was obtained from MacDonald and Long (2005). This region, from approximately 38514 – 42495, was entered into the Primer3 program (Whitehead Institute 2004) to obtain primer sequences that would contain the region from the TATA box upstream to the paired suppressor of hairless site. The enclosed region in *D. pseudoobscura* contains 456 base pairs, and 632 base pairs in *D. melanogaster*. The primers were selected from regions that have shown a high level of conservation between *D. melanogaster* and *D. pseudoobscura* ensuring that they would bind to sequences in both species. Right and left primers (Table 2) were ordered from Invitrogen Corporation (2006).

D. pseudoobscura genomic DNA was isolated (Ish-Horowicz 1989). Polymerase chain reaction (PCR) (Sambrook et al. 1989, 14.5-6) was run using two sets of genomic *D. pseudoobscura* DNA, two corresponding positive controls, and a negative control. PCR was run for 35 cycles with a denaturation temperature of 95°C, annealing temperature of 59.5°C, and an elongation temperature of 75°C. The products were then separated on a 1.8% low melt agarose gel. A 500 base pair PCR product was purified using phenol chloroform and precipitated with ethanol (Sambrook et al. 1989, 6.30-1). The purified *D. pseudoobscura* product was sent to the Dana-Farber Cancer Center DNA Resource Core (DNA Tools 2006) to be sequenced.

The results of the sequenced *D. pseudoobscura m7* product was blasted (NCBI 2006) against the *D. pseudoobscura* sequence obtained from the VISTA Genome Browser.

RESULTS

Alignments of E(spl) proteins from *D. melanogaster* and *D. pseudoobscura*

Figure 4 shows the protein alignment of the E(spl) mγ, and m7 proteins between *D. pseudoobscura* and *D. melanogaster.* The comparison for m7 yielded the most number of differences between *D. melanogaster* and *D. pseudoobscura,* while mγ resulted in a higher level of similarity. The similarities between the two species were in the Helix-Loop-Helix regions of the protein sequence, and the greatest amount of diversity between the two species was found in the carboxy terminus.

Comparison of *E(spl) m7* and *mγ* regulatory regions from multiple species of *Drosophila*

Upstream regulatory region alignments were created to investigate the similarities and differences of binding sites among evolutionarily divergent species of *Drosophila.* The ClustalW alignment of *m7* regulatory regions can be seen in Figure 5, and mγ in Figure 6. Table 3 illustrates in detail the similarities in binding sites between the species.

E-box sites identified by Nellesen et al. (1999) for *D. melanogaster* in *mγ* were conserved in *D. pseudoobscura, D. hydei* and *D. simulans*, and in *m7* the two sites were conserved among *D. melanogaster* and *D. simulans*, but *D. pseudoobscura* showed only one E-box sequence in between the aforementioned two.

As for N-boxes, in *m7*, all three sites, identified by Nellesen et al. (1999), were conserved among *D. melanogaster* and *D. simulans*, and *D. pseudoobscura* except for site V for which there was no corresponding sequence in *D. pseudoobscura.* In *mγ*, one N-box site (VII) was conserved among all four species. *D. hydei* was one base pair different from the other three species in site VIII. *D. hydei* had an extra N-box binding site at about -257 base pairs.

The Su(H) sites in *m7* showed some conservation. In site X, *D. melanogaster* and *D. simulans* were conserved, but there was no site present for *D. pseudoobscura*. In site IX, all three species showed the same sequence, but the *D. pseudoobscura* sequence was located upstream of the other two. Site VIII was conserved among all three species, and in site IV, *D. simulans* was one base pair off of *D. melanogaster* and there was no site present for *D. pseudoobscura.* In the upstream sequences of *mγ,* one site was conserved among all four species (VI), and site V was conserved among the three species but modified in *D. hydei.* Site III was conserved among all, except *D. pseudoobscura.* Either this site is not present in this species or it is located farther upstream from the other three species. The paired Su(H) sites are conserved among all 3 and 4 species in *m7* and *mγ* respectively.

EvoPrinter analysis of *m7* upstream regulatory sequence in multiple *Drosophila* species

D. melanogaster reference

The relative percentages of identity between each *Drosophila* species and *D. melanogaster* can be seen in Table 4. These identities were used to determine relative evolutionary relatedness and the order in which species were added into alignments. These relative percentages also correlate with the *Drosophila* species phylogenetic tree (Fig. 2). As the species diverge more from the species in the reference sequence, the percent identity decreases. The BLAT alignments (Robson, unpublished results) for each reference sequence indicated similarities between species with capital letters.

A further investigation of the *m7* upstream regulatory region involved entering the BLAT alignments into the EvoPrinter program to evaluate the conservation of upstream binding sites. The EvoPrinter alignment using the *D. melanogaster* (MacDonald and Long 2005) reference sequence results are shown in Figure 7. The first two species to be compared were that of *D. simulans* and *D. yakuba*. Following the first alignment, another species of *Drosophila* was added in based on the percentages given in Table 4. This set of alignment excludes *D. sechellia* and *D. persimilis*. The upstream regulatory region sites were identified and colored in their respective colors. As seen in Figure 7, when more species are added into the EvoPrint alignment, there are fewer conserved sequences.

Table 5 shows which upstream regulatory sites are retained and which are lost as species are added into the EvoPrinter program. In the first EvoPrint alignment placing *D. simulans* and *D. yakuba* against *D. melanogaster,* Su(H) site X, Su(H) site VIII, N-box site VII, Paired Su(H) site VI, E-box site III, E-box site II, and N-box site I were all conserved. Site IX (Su(H)), site V (N-box), and site IV (Su(H)) were not conserved. (These sites remained "lost" as the rest of the alignments were constructed.) Adding *D. yakuba* into the alignment resulted in the loss of conservation of the Su(H) site VIII. When *D. ananassae* was added to the above 3 species alignment, Su(H) site X, and both E-box sites III and II were lost. Upon adding in *D. pseudoobscura, D. grimshawi, D. virilis, and D. mojavensis* the alignment did not change with respect to the conservation of upstream regulatory sites. The upstream regulatory binding sites that remain conserved throughout the addition of all 8 species are the N-box (VII) that overlaps the paired site, the paired Su(H) (VI), and the N-box furthest downstream (I).

Using a longer *D. melanogaster* sequence (FlyBase), each of the ten species was individually aligned against the reference sequence (Figure 8). The similarities and differences in the upstream regulatory binding sites are summarized in Table 6.

D. melanogaster aligned with *D. simulans* resulted in the loss of Su(H) sites IX and IV. The same result was seen when *D. melanogaster* was aligned with *D. sechellia. D. melanogaster* aligned with *D. yakuba* resulted in the loss of an N-box (V) and Su(H) site (IV). The alignment with *D. erecta* showed a loss of Su(H) sites VIII and IV and N-box site V. *D. melanogaster* alignments with *D. persimilis, D. pseudoobscura,* and *D. grimshawi* resulted in the same loss of upstream regulatory binding sites: Su(H) sites X, IX, and IV, E-box site III and N-box site V. The alignment with *D. ananassae* resulted in the highest number of upstream regulatory site losses: Su(H) X, IX, VII, IV; N-box V; and both E-boxes III and II. *D. virilis* and *D. mojavensis* resulted in the same losses when aligned with *D. melanogaster*: Su(H) X, IX, VIII, IV; N-box V and E-box III. The sites that were always conserved among all species were E-box XI, N-box VII that overlaps the paired site, Paired Su(H) VI, and N-box I.

Using the longer *D. melanogaster* sequence, a new E-box site was located further upstream from the Su(H) site X. This site was conserved between each species and *D. melanogaster*. In *D. melanogaster* this new E-box site is marked as site XI.

Both methods of aligning *Drosophila* species resulted in the same losses and conservations of sequences. The one on one comparisons were designed to further investigate each individual species upstream regulatory region. To further investigate evolutionary relatedness, two reference sequences of divergent species were used to see which *Drosophila* species are more related to each other.

***D. pseudoobscura* reference**

Figure 9, a map of the *m7* upstream regulatory region binding sites for *D. pseudoobscura*, was designed following the analysis of the EvoPrinter results. This map was created to compare *D. pseudoobscura* to that of *D. melanogaster.* Using the map of the region, a better idea of the big picture of transcriptional regulation can be gained. EvoPrinter analysis of *m7* using *D. pseudoobscura,* eight upstream sites were identified and marked with cardinal numbers 1 – 8, with site 1 being the furthest downstream. A high affinity N-box (site 1) at -47 base pairs bears the sequence CACGTG. At site 2, -151 base pairs upstream, an E-box was identified with the sequence CAGGTG. Site 3 consists of a high affinity N-box -408 base pairs upstream with the sequence CACGTG. Sites 4 and 5, -496 and -498 base pairs respectively are the paired suppressor of hairless and overlapping N-box sites. They are the same sequences as in *D. melanogaster:* CTTGTGGGAA and TTCCCACG. Sites 6 and 7 are both high-affinity suppressor of hairless binding sites at -764 and -970 base pairs respectively. They both share the sequence TTCCCACG. Site 8 is an E-box site at -1398 base pairs upstream with the sequence CAGGTG.

The results of the first EvoPrinter alignment using the *D. pseudoobscura* sequence from the VISTA genome browser are shown in Figure 10. The first two species to be compared were that of *D. melanogaster* and *D. simulans*. Following the first alignment, another species of *Drosophila* was added in based on the percentages given in Table 4. This set of alignment excludes *D. sechellia* and *D. persimilis*. The upstream regulatory region sites were identified and coded in their respective colors. As seen in Figure 10, when more species are added into the EvoPrint alignment, there is less conservation among species.

Table 7 shows which upstream regulatory sites are retained and which are lost as species are added into the EvoPrinter program. In the first EvoPrint alignment placing *D. melanogaster* and *D. simulans* against *D. pseudoobscura,* Su(H) site 6 and N-box site 3 are lost. Sites 8 (E-box), 7 (Su(H)), 5 (N-box), 4 (paired Su(H)), 2 (E-box) and 1 (N-box) remain conserved. As *D. yakuba* is added in, there is a loss of the Su(H) site 7. With *D. erecta* added into the alignment, the same loss of sites is seen as when *D. yakuba* was added. However, when adding in *D. grimshawi* there is a loss of the E-box site (2). This same alignment results are produced when *D. virilis, D. ananassae* and *D. mojavensis* are added in. The sites that remain conserved among all 8 species against *D. pseudoobscura* are E-box site 8, N-box site 5 that overlaps the paired Su(H) site 4 which also remains conserved, and N-box site 1.

The results from the EvoPrinter alignment placing one species against the *D. pseudoobscura* reference can be seen in Figure 11. The similarities and differences in the upstream regulatory binding sites are summarized in Table 8 and described below.

D. pseudoobscura aligned one hundred percent with *D. persimilis.* When *D. pseudoobscura* was aligned with *D. melanogaster, D. simulans* and *D. sechellia,* there was the same loss of upstream regulatory binding sites: Su(H) site 6 and N-box site 3. *D. yakuba, D. erecta*, *D. ananassae* and *D. grimshawi* when individually aligned with *D. pseudoobscura*, lost the above two sites as well as Su(H) site 7. Aligning *D. virilis* with the *D. pseudoobscura* reference resulted in losses similar to those found in alignments with *D. yakuba, D. erecta*, *D. ananassae* and *D. grimshawi;* however, *D. virilis* also lost E-box site 2. In the alignment between *D. mojavensis* and *D. pseudoobscura* the upstream sites that were lost were similar to those lost in alignments with *D. melanogaster, D. simulans* and *D. sechellia;* however, *D. mojavensis* also lost E-box site 2. The sites that remained conserved in each alignment were E-box site 8, N-box site 5 that overlaps the paired suppressor of hairless site, the Paired Su(H) site 4, and N-box 1.

Both EvoPrinter alignments using *D. pseudoobscura* produced the same loss and conservation of upstream binding sites.

Analysis of flanking bases on E-boxes

The flanking bases of the E-boxes were analyzed to determine if there is a preference for atonal or scute binding and if this preference is retained among species. Table 9 shows the flanking bases of the E-box binding sites in both *D. melanogaster* and *D. pseudoobscura*. In the species that an E-box is not conserved with the reference species, not conserved is written. This is because flanking bases cannot be determined. Where one letter is present instead of two, the one letter could not be determined because it wasn't conserved with the reference sequence so it merely showed up as a lowercase letter and not the actual sequence present. All of the E-boxes present are Sc/Da binding sites. Again, Sc/Da E-box sites are characterized by their preference for the sequences **G**CAGSTG**G** and **G**CAGSTG**T** (Powell et al. 2004). S is either G or C and in every *m7* E-box, the S is a G. In *D. melanogaster,* the flanking bases for sites XI and II are G and G; whereas, site III is flanked by a G and a T. *D. pseudoobscura*'s two E-boxes (8 and 2) are both flanked by two G's.

Confirmation of sequence in D. pseudoobscura upstream regulatory region

Using the primers (Table 2) in PCR reactions, followed by gel electrophoresis analysis of the PCR products, the gel that showed bands at the proper location can be seen in Figure 12. The bands were located just below the 500 base pair marker. The target region in *D. pseudoobscura* was 456 base pairs.

To confirm that there is a change in E-box arrangement between *D. melanogaster* and *D. pseudoobscura*, the experimental sequence returned from the DF-HCC Lab (DNA Tools 2006) was entered in a BLAST (NCBI 2006) alignment against the published *D. pseudoobscura* sequence. The one E-box (site 2) sequence was located in both experimental and published *D. pseudoobscura* sequences (Fig 13). The flanking bases on either side of the E-box are identical indicating scute is the preferred proneural binding protein.

DISCUSSION

Alignments of E(spl) proteins from *D. melanogaster* and *D. pseudoobscura*

The results for the *E(spl) m7* protein sequence alignment show a significant number of amino acid differences between *D. pseudoobscura* and *D. melanogaster.* Particularly in the carboxy terminus, this region is the least conserved between E(spl) proteins of *D. melanogaster* and has been predicted to be the region responsible for differential functions of the E(spl) proteins. This leads one to speculate that *m7* could have different functions in *D. melanogaster* and *D. pseudoobscura.* The opposite is predicted for *m*γ, since the protein alignment was strikingly similar in the two species.

The similarities in the m7 protein alignments were in the Helix-Loop-Helix domains which are highly conserved regions within the *D. melanogaster E(spl)* genes. It is interesting that the highly conserved regions in *D. melanogaster* are also highly conserved with *D. pseudoobscura.* The greatest number of differences in protein sequence between *D. melanogaster* and *D. pseudoobscura* were in the carboxy terminus regions. In *D. melanogaster,* this region shows the greatest diversity between the *E(spl)* bHLH genes and it has been predicted that this region could allow for different functions of the E(spl) proteins (Delidakis 1992). Since there was a great deal of diversity between species in this region of m7 it is possible that either m7 has different functions in different species or that this region does not play an important role in differential functions of the E(spl) proteins after all.

Comparison of *E(spl) m7* and *mγ* regulatory regions from multiple species of *Drosophila*

Upstream regulatory alignments were produced to further investigate the mechanism of transcriptional control that is implemented in different species of *Drosophila.* It is predicted that species with similar upstream binding sites will show similar expression patterns, since transcriptional regulation determines gene expression. The comparison of the upstream regulatory regions of these genes revealed some differences and similarities that were of particular interest to us.

Overall, *mγ* binding sites were more conserved among four *Drosophila* species than *m7. mγ* and *m7* comparisons show complete conservation in sites VII and VI (N-box and Su(H) respectively). Nellesen et al. (1999) suggested that the proximity of N-boxes to Su(H) binding sites could indicate a cooperative repressor function of Su(H) and the bHLH repressors. Our results showing the conservation of these N-box sites and their proximity to conserved Su(H) sites supports this hypothesis.

E-box Sites

Analysis of E-box sites was performed to investigate which proneural proteins prefer the sites present in the *E(spl) mγ and m7* upstream regions. *mγ* showed completely conserved E-boxes (IV and II) among the four species, but in *m7,* E-boxes III and II were conserved between *D. melanogaster* and *D. simulans,* but *D. pseudoobscura* only contained one E-box that was located between sites III and II. This begs the question; would the lack of an E-box in *D. pseudoobscura* allude to some functional difference in this species? It is also interesting that there is a significant difference in *m7,* but not in *mγ.* Based on the work by Powell et al. (2004), it is possible that different E-boxes might unravel the mechanism for different expression patterns of the *E(spl)* genes. Proneural proteins can bind any E-box to activate transcription, but the different specificities for the proteins lie in the flanking bases of the E-box sequence. The E-box sites identified from the upstream comparisons of *m7* showed a similarity to the scute consensus binding site. A cell expressing scute, becomes a sensory organ precursor cell in the proneural cluster.

Su(H) Sites

In both *mγ* and *m7,* the paired Su(H) sites were conserved among all of the species. It is not a surprise that the paired sites are completely conserved among all species and in both genes. Cave et al. (2005) showed that transcriptional cooperation between the ICN and proneural activator proteins requires specific upstream regulatory sequence organization of the target genes. A specifically oriented pair of Su(H) sites are necessary for Notch-proneural cooperation, because this DNA construction allows for cooperative interactions with proneural proteins (Cave et al. 2005).

EvoPrinter analysis of *m7* upstream regulatory sequence in multiple *Drosophila* species

As the different EvoPrint alignments were created for *D. melanogaster m7* references, it was interesting that the upstream sites I – N-box, VI – Paired Su(H), VII – N-box, and XI – E-box remained conserved among all *Drosophila* species analyzed. These findings again confirm the synergistic importance of paired Su(H) sites with N-box sites (Cave et al. 2005). The three Su(H) binding sites VIII, IX (especially) and X were not well conserved. This suggests that proper *m7* expression might not require all Su(H) sites to be functional or that expression patterns vary in different species. The N-box (V) downstream from the paired Su(H) site was also not well conserved, as only two other species showed retention of this site.

In the EvoPrints using *D. pseudoobscura* as a reference, the sites that remained conserved among all species were 1 – N-box, 4 – Paired Su(H), 5 – N-box, and 8 – E-box. The only two species that retained Su(H) site 6 and N-box site 3 were *D. persimilis* and *D. pseudoobscura*. The *D. persimilis* upstream regulatory sites matched 100% to those of *D. pseudoobscura.* Su(H) site 7 remained conserved in half of the species compared to *D. pseudoobscura*, and only two species, *D. virilis* and *D. mojavensis,* lost the E-box site 2.

The retention of the N-box (1 in *D. pseudoobscura* and I in *D. melanogaster*) furthest downstream, the paired Su(H) (4 and VI) with the overlapping N-box (5 and VII), and E-box (8 and XI) suggests these sites are the most crucial in the proper expression of the *E(spl) m7* gene. Since these sites have remained conserved throughout the divergence of the *Drosophila* species, they appear to be the sites required, at minimum, to ensure accurate development.

In determining which species are more similar to *D. melanogaster* or *D. pseudoobscura,* the percent of conserved sites was calculated for each species against each reference species. *D. melanogaster*'s upstream regulatory binding region contained eleven sites, whereas *D. pseudoobscura* had eight. Table 10 gives the percent similarity between each of the eleven *Drosophila* species and *D. melanogaster* and *D. pseudoobscura.* The percent similarity was calculated by the dividing the number of conserved sites by the total number of Su(H), E-box and N-box sites present in each of the reference species.

The species with an upstream regulatory binding region more similar to that of *D. melanogaster* are *D. simulans, D. sechellia, D. yakuba* and *D. erecta.* These four species have a higher percent similarity to *D. melanogaster* than to *D. pseudoobscura. D. persimilis, D. ananassae, D. grimshawi, D. virilis,* and *D. mojavensis* show a higher percent similarity to *D. pseudoobscura* than *D. melanogaster.* These findings are consistent with a proposed *Drosophila* phylogenetic tree (Fig. 2). This reveals that certain divergences of species occurred when the *melanogaster* subgroup split from *D. ananassae.* As seen in Table 10, *D. virilis* showed the least amount of similarity to both *D. melanogaster* and *D. pseudoobscura.* The fact that *D. mojavensis* and *D. virilis* diverged from *D. grimshawi* suggests that *D. mojavensis* and *D. virilis* could have different upstream binding sites. Both of these species showed only one E-box binding site.

Analysis of flanking bases on E-boxes

Taking into account E-boxes as a means of interpreting *Drosophila* species divergence, many conclusions can be drawn. *D. simulans, D. sechellia,* and *D. melanogaster* all have three E-boxes with the same flanking bases GG, GT, and GG. These species are the closest relatives to *D. melanogaster*. *D. yakuba* and *D. erecta* are one and two steps respectively further out on the species tree. These two species retain all three E-box binding sites (similar to *D. melanogaster*); however, the 5’ flanking base of the middle E-box site (III) is no longer a G. This difference could change the E-box site from a scute to an atonal proneural binding site. Further analysis examining the actual sequence of *D. yakuba* and *D. erecta* would have to be done to determine the kind of change taking place at the 5’ flanking base on E-box III.

When the *melanogaster* subgroup and *D. ananassae* diverged an E-box binding site (III) was gained in the *melanogaster* subgroup. *D. ananassae* and the other *Drosophila* relatives not yet addressed do not contain this third E-box site. There is a difference in *D. ananassae*'s 3' flanking base on E-box site 1. It is not a G, but could perhaps be a T and still keep its scute functionality, since the difference between atonal and scute sites lies primarily in the 5' flanking base. It's likely that when the *Drosophila* species diverged into *D. ananassae* and the *obscura* group (*D. pseudoobscura* and *D. persimilis*) that the two E-boxes maintained their scute functionalities.

In the divergence of the *Drosophila* and *Sophophora* groups, it appears that the *Hawaiian Drosophila D. grimshawi* in the *Drosophila* group is similar to the members of the *obscura* group in the *Sophophora* group (FlyBase). These species contain two E-boxes both with flanking bases that match the preference for scute binding sites. It should be noted that there have been discrepancies in the proposed *Drosophila* phylogenetic trees due to differences in research methods.

Members of the *Drosophila* group split into the *Hawaiian Drosophila* group and the *repleta* and *virilis* groups (FlyBase). *D. virilis* is a member of the *virilis* group, and *D. mojavensis* belongs in the *repleta* group. Both of these species showed a higher percentage of similarity to *D. pseudoobscura*, so they will be analyzed against *D. pseudoobscura* first. *D. mojavensis* and *D. virilis* both lost E-box site 2, and therefore only contain one E-box site (8) with respect to the *D. pseudoobscura* reference. However, if compared to *D. melanogaster* these two species retain E-box site XI as well as II. Site III is lost. *D. mojavensis'* 5' flanking base on E-box site II is no longer a G, and might be an atonal binding site instead of a scute site. Again, the actual *D.*

mojavensis upstream sequence would have to be analyzed further before this prediction can be accepted as true. Since both *D. mojavensis* and *D. virilis* lack E-box site III (with respect to *D. melanogaster*), these two species are most likely more divergently similar to *D. pseudoobscura.*

It is interesting that five *Drosophila* species have three E-box sites, four species have two E-boxes and two species only have one E-box site in the same *m7* gene. Although further experiments are needed, it appears that E-boxes are an important indication of A) divergence of species and B) the type of upstream regulatory control present on the gene.

From the above results and interpretations, it appears that using the conservation and loss of E-boxes may be a reliable way of measuring *Drosophila* species divergence.

Future experiments to determine if there is a preference for Sc/Da or Ato/Da E-box binding in the species with one, two, and three E-boxes could give insight to further transcriptional control over the *E(spl) m7* gene in *Drosophila.*

Confirmation of sequence in D. pseudoobscura upstream regulatory region

Since we identified variation in the regulatory sites of multiple species, we predict that there may be variation in expression of *m7* in different species. This could be studied using several different methods including *in situ* hybridization and tissue culture experiments. The isolated upstream sequence of *m7* from various *Drosophila* species (PCR product) could be cloned into a luciferase vector. This vector could be cotransfected into S2 tissue culture cells along with the expression vectors containing

ICN, Su(H), scute, atonal, or an E(spl) bHLH repressor. Expression of luciferase would report the level of activation by the different regulatory proteins.

Knowing which proteins are required to initiate transcription, could give rise to an explanation of whether or not differing upstream regulatory regions result in different patterns of expression. Due to the remarkable differences and similarities in the upstream regulatory regions of the 9 other *Drosophila* species compared to the upstream sequences of *D. pseudoobscura* and *D. melanogaster,* it is expected different expression pattern will be observed. It is also expected that different combinations of DNA binding proteins, namely the proneural proteins, will be required for transcription of *m7* in the different species. The notable difference in location and number of E-box biding sites from *D. melanogaster* suggests that different amounts and combinations of proneural proteins are required for transcription of the *E(spl) m7* gene in *D. pseudoobscura*.

The goal of this project was to compare the *E(spl)* genes of *D. melanogaster* to those of *D. pseudoobscura*. This was accomplished through the use of a number bioinformatics alignments and comparisons along with the beginnings of expression vector and cloning experiments. By revealing the important components of transcriptional regulation, Notch activated gene expression patterns can give insight into the evolutionary functionality of the Notch pathway and its interaction with the *Enhancer of split* genes. Using bioinformatics, similarities and differences were observed in the protein and upstream regulatory sequences for the *E(spl)* genes focusing mainly on *m7* because of its remarkable amount of differences in the upstream regulatory binding site region. Although the *Drosophila* genus has diverged throughout evolution, the *E(spl)* sequences as a whole are conserved; indicating these genes are necessary for development in *Drosophila*.

More experiments will be performed to hopefully achieve conclusive results about what is important in *m7* expression. It is expected that our experimental *D. pseudoobscura* expression patterns will be different than those of the documented *D. melanogaster* results, and will be seen from the embryo and larval imaginal disc staining patterns as well as tissue culture cell assays.

REFERENCES

Artavanis-Tsakonas, S., M.D. Rand, and R.J. Lake. (30 April 1999). Notch Signaling: Cell Fate Control and Signal Integration in Development. Science **284**: 770-776.

Bailey, A.M., and J.W. Posakony. (1995). Suppressor of Hairless directly activates transcription of *Enhancer of split* Complex genes in response to Notch receptor activity. Genes and Development **9**: 2609-2622.

Bray, S., and M. Furriols. (2001). Notch pathway: Making sense of Suppressor of Hairless. Current Biology **11**: R217-R221.

Brody, T.B. (1996) The Inter*active* Fly. < http://flybase.bio.indiana.edu/allied-data/lk/interactive-fly/aimain/1aahome.htm>.

Brunkan, A.L. and A.M. Goate. (2005) Presenilin function and gamma-secretase activity. Journal of Neurochemistry **93**(4):769-92.

Campuzano, S. and J. Modolell. (1992). Patterning of the *Drosophila* nervous system: the *achaete-scute* gene complex. Trends in Genetics **8**: 202-208.

Cave, J.W., F. Loh, J. Surpris, L. Xia and M. Caudy. (2005). A DNA transcription code for cell-specific gene activation by *Notch* signaling. Current Biology **15**: 94-104.

Chien, C.T., C.D. Hsiao, L.Y. Jan, and Y.N. Jan. (12 Nov 1996). Neuronal type information encoded in the basic-helix-loop-helix domain of proneural genes. Proceedings of the National Academy of Sciences USA **93**(23): 3239-44.

Cooper, M.T.D., D.M. Tyler, M. Furriols, A. Chalkiadaki, C. Delidakis, and S. Bray. 2000. Spatially restricted factors cooperate with Notch in the regulation of *Enhancer of split* genes. Developmental Biology **221**: 390-403.

Couronne, O., A. Poliakov, N. Bray, T. Ishkhanov, D. Ryaboy, E. Rubin, L. Pachter, I. Dubchak. 2003. Strategies and Tools for Whole-Genome Alignments. Genome Research **13**: 7. <http://pipeline.lbl.gov/cgi-bin/gateway2?bg=dm1>.

Delidakis, C., and S. Artavanis-Tsakonas. (September 1992). The *Enhancer of split* [*E(spl)*] locus of *Drosophila* encodes seven independent helix-loop-helix proteins. Genetics **89**: 8731-8735.

DNA Tools. (2006). DNA Sequencing Ordering and Sequence Retrieval System. < http://128.103.32.162/cgi-bin/dna/password.cgi>. Via Dana-Farber/Harvard Cancer Center Resource Core < http://dnaseq.med.harvard.edu/dna_sequencing_services.htm>.

ELG. (1999). ExPASy Translate Tool. <http://us.expasy.org/tools/dna.html>.

European Molecular Biology Institute - European Bioinformatics Institute (EMBL-EBI). (2005). ClustalW Sequence Analysis. < http://www.ebi.ac.uk/clustalw/#>.

FlyBase DroSpeGe: Drosophila Species Genomes.
<http://bugbane.bio.indiana.edu:7151/>.

Invitrogen Corporation. (2006). Online Catalog – Order Custom Primers.
< https://catalog.invitrogen.com/index.cfm?fuseaction=customerSite.primersHome>.

Ish-Horowicz, D. (1989) Isolation of DNA from adult flies. In: Ashburner, Michael, editor. *Drosophila:* A Laboratory Manual. New York: Cold Spring Harbor Laboratory Press. 106-107.

Jarman, A.P., Y. Grau, L.Y. Jan, and Y.N. Jan. (1993). *Atonal* is a proneural gene that directs chordotonal organ formation in the *Drosophila* peripheral nervous system. Cell **73**: 1307-1321.

Jennings, B.H., D.M. Tyler, and S.J. Bray. (1999). Target Specificities of *Drosophila* Enhancer of split Basic Helix-Loop-Helix Proteins. Molecular and Cellular Biology **19**(7): 4600-4610.

Joutel, A., C. Corpechot, A. Ducros, K. Vahedi, H. Chabriat, P. Mouton, S. Alamowitch, V. Domenga, M. Cecillion, E. Marechal, J. Maciazek, C. Vayssiere, C. Cruaud, E.A. Cabanis, M.M. Ruchoux, J. Weissenbach, J.F. Bach, M.G. Bousser, and E. Tournier-Lasserve. (1996). Notch3 mutations in CADASIL, a hereditary adult-onset condition causing stroke and dementia. Nature **383**(6602): 707-10.

Kent, W.J. (2002). BLAT--the BLAST-like alignment tool. Genome Research **12**: 656-64. (Maintained by UCSC Genome Bioinformatics).

Kramatschek, B. and J. Campos-Ortega. (1994). Neuroectodermal transcription of the *Drosophila* neurogenic genes *E(spl)* and *HLH-m5* is regulated by proneural genes. Development **120**: 815-826.

Ligoxygakis, P., S.J. Bray, Y. Apidianakis, and C. Delidakis. (1999). Ectopic expression of individual *E(spl)* genes has differential effects on different cell fate decisions and underscores the biphasic requirement for Notch activity in wing margin establishment in *Drosophila.* Development **126**: 2205-2214.

Macdonald, S.J. and A.D. Long. (2005). Identifying signatures of selection at the Enhancer of split neurogenic gene complex in *Drosophila*. Molecular Biology and Evolution **22**(3): 607-619.

Maier, D., B.M. Marte, W. Schafer, Y. Yu, and A. Preiss. (June 1993). *Drosophila* evolution challenges postulated redundancy in the *E(spl)* gene complex. Developmental Biology **90**: 5464-5468.

National Center for Biotechnology Information (NCBI). (2006). Basic Local Alignment Search Tool (BLAST). < http://www.ncbi.nlm.nih.gov/blast/>.

Nellesen, D.T., E.C. Lai and J.W. Posakony. (1999). Discrete enhancer elements mediate selective responsiveness of *Enhancer of split* complex genes to common transcriptional activators. Developmental Biology **213**: 33-53.

Odenwald, W.F., W. Rasband, A. Kuzin and T. Brody. (2005). EvoPrinter: A multigenomic comparative tool for rapid identification of functionally important DNA. Proceedings of the National Academy of Sciences **102**: 14700-14705.

Powell, L., P. zur Lage, D. Prentice, B. Senthinathan and A. Jarman. (2004). The proneural proteins atonal and scute regulate neural target genes through different E-box binding sites. Molecular and Cellular Biology. **24**(21): 9517-9526.

Sambrook, J., E.F. Fritsch and T. Maniatis, comps. Molecular Cloning: A Laboratory Manual. 2nd ed. New York: Cold Spring Harbor Laboratory Press, 1989. 6.30-6.31, 14.5-14.6.

Wech, I., S. Bray, C. Delidakis, and A. Preiss. (1999). Distinct expression patterns of different *Enhancer of split* bHLH genes during embryogenesis of *Drosophila melanogaster.* Development, Genes, and Evolution **209**: 370-375.

Whitehead Institute for Biomedical Research. (2004). Primer 3. < http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi>.

TABLES AND FIGURES

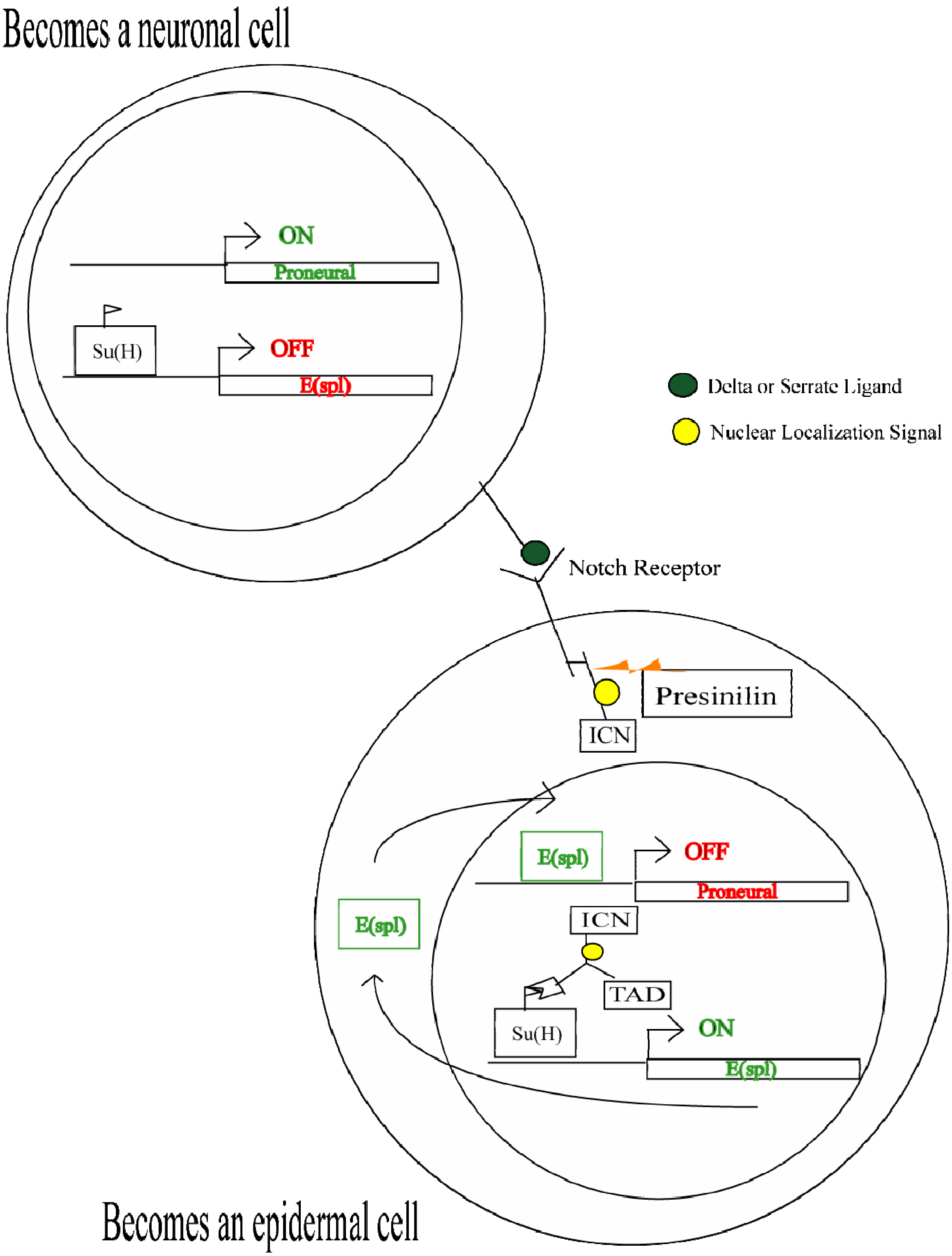


Figure 1: The Notch pathway.

The Notch pathway laterally inhibits neighboring cells from becoming neuronal cells via the regulation of the *E(spl)* genes.

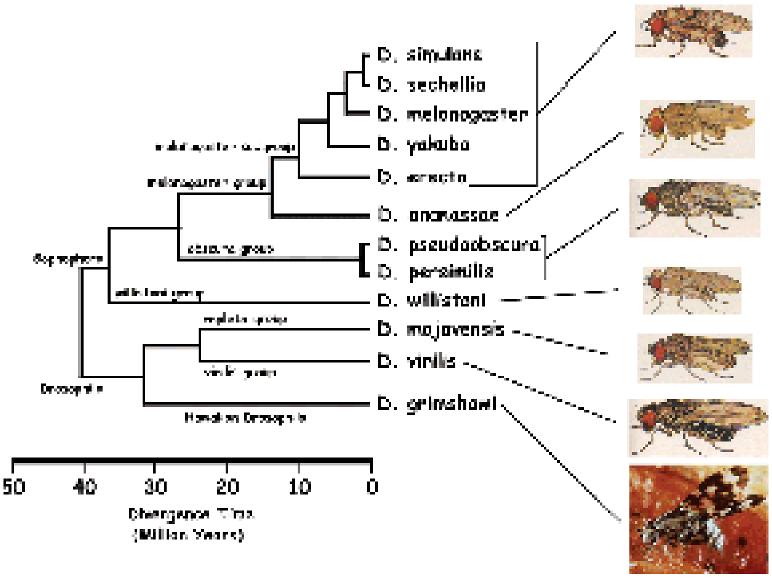


Figure 2: Divergence of *Drosophila* species.

The *Drosophila* species has diverged throughout evolution, and the relative relationships among species are shown above. This figure was obtained from FlyBase <http://bugbane.bio.indiana.edu:7151/>.

Table 1: Binding site types and sequences.

Binding Protein	Binding Sequence	Color Identifier
E-box variants	CAGGTG CAGCTG	**Blue** **Brown**
Suppressor of Hairless High Affinity Su(H)	RTGRGAR (R = A/G) YGTGRGAA (Y = C/T)	**Orange**
N-box High Affinity N-box	CACNAG (N = any base) CACGYG	**Pink**

The upstream regulatory binding site types (E-box, Su(H) and N-box) and sequences present in the *E(spl)* genes were identified in *mγ* and *m7* by using different colors to highlight the different binding sites.

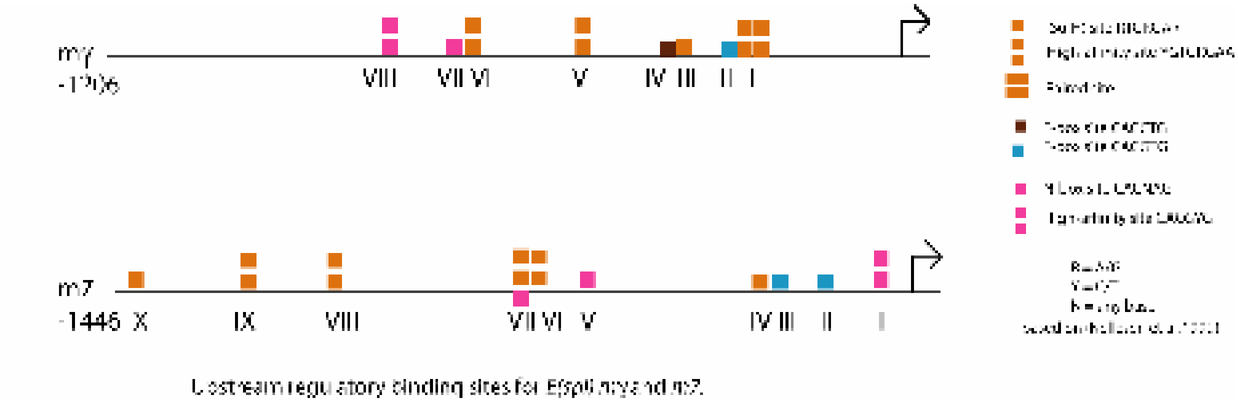


Figure 3: Upstream regulatory binding sites for *E(spl) mγ* and *m7* in *D. melanogaster.*

Based on the research by Nellesen et al. (1999), these upstream regulatory region maps were created to illustrate the relative location and type of binding sites in *D. melanogaster mγ* and *m7*.

Table 2: Primer sequences.

Species	Left Primer Sequence	Sequence Enclosed	Right Primer Sequence
D. melanogaster	GCAAAGTGTTTCCCACGATT	GCAAAGTGTTTCCCACGATTCGAATGTCAGTGGAAAGGAAAACAGAGAA GCAGCATGTGCAACCACCGTCGCACCACCACCACTGCACTATCGCATCT TCATCCCCTGCCGAGACTCACTTACACCGAA**CACCAG**TGGCCATAGCAA TAGTACTGAGAACATATTTAAAGTTCTAGATAATGTTTATATTTTCAAA ATCACTACTATTAAGAAGTGTGCTAATTTAATTTTTTAGGTGTAAACGT ATAACATTATGTTAATTTCTTTTTGTTTAATAAAACGACATTTAAGTGC TTTAAGGGAAACCCATTTGAAGTAATTGTTAAAAACGTAACGTACTTCT TAGGTATATAACAACAAATTTGTATATTCGCTCGAAGAACTATAGCTGC TTTAAACACAT**TTCTCAT**ACTATTTTTACC**GCAGGTG**TGTTCCAAACTT GTTTGTGGAGAAAAACCCACCACCGAAAATGCAGGCGAGAGGATCGTGT CCCAGTGACTGAGTAAAAGTAGTTTGCTCC**GCAGGTG**GTGGTTCTTCGG AGGCTCCGCAGCTCTCTCTTCTCTCTTTCTCCGAGCAGCGTGCGCTGAG CACGAGACGCTCTCGCGGCGG**CACGTG**CAGCTATAAAAGCAGCG Product is 632 base pairs.	GTGCAGCTATAAAAGCAGCG
D. pseudoobscura	GCAAAGTGTTTCCCACGATT	**GCAAAGTGT**TTCCCACGATTCGAATGTCAGGCTAAAAGAAAACAGAGGA GAGCACACACAAGAAAACGCACA**CACGTG**CATCGGCCGCATGTGCAACC AACCACCACCGCCAAGAGCAGCAGCAGCAAGCAGCAACAACAACAGCTA TCCACCAACACAGGAATGGAGGAGACACACCATCACATCATCTTCATCC CCTGCCAAGACGACTTGTTTGTCGAGAAAAACCCACCACCGAAAGCAGG CGAGGATCGTGGCTGTCGGCCCGAGTAACGATTCGTAACGACGACACGA ACGACAATCGGCTCGACCCGACCCAGTGACTGACG**GCAGGTG**GTCGCAA TGGTGGTGCTCGCTCCGGTGGTGCCTCTCTCGTTTCCTGCGCATGCGCG CTCTCTCTGCGCGCGCTGAGCACGAGACGACTCTGCTGCGG**CACGTG**CA GCTATAAAAGCGGCG Product is 456 base pairs.	GTGCAGCTATAAAAGCAGCG

Primers were designed using the *D. melanogaster* sequence to enclose the upstream regulatory region of the *E(spl) m7* gene. The same primers were used in experiments to isolate both *D. melanogaster* and *D. pseudoobscura* DNA. There was one base pair difference between the *D. melanogaster* primer sequence and the *D. pseudoobscura* genomic sequence; this difference is highlighted above. The sequences enclosed between the primers differ by 176 base pairs.

Gamma 5'3'

```
M S S S Q I S E M S K T Y Q Y R K V M K P M L E R K R R
M S S L Q M S E M S K T Y Q Y R K V M K P M L E R K R R

A R I N K C L D E L K D L M V A T L E S E G E H V T R L E
A R I N K C L D E L K D L M V A T L E S E G E H V T R L E

K A D I L E L T V T H L Q K M K E K R Q H K R A T G D E S L S P A
K A D I L E L T V T H L Q K M K Q Q R Q H K R A S G D E S L T P A

E G F R S G Y I H A V N E V S R S L S H L P G M N V S L G T Q L M T H L G
E G F R S G Y I H A V N E V S R S L S Q L P G M N V S L G T Q L M T H L G

Q R L N Q L Q P A V K E A L P I T A P L S V H I A S R D A Y S V P I S P V S S F A
Q R L N Q I Q P A E K E V L P V T A P L S V H I A N R D A Y S V P I S P I S S Y A

N S S L A S S E R L G S A S L L T T C S S I D V T K M E L E V D S E D E E N
N S N T S S T         S H S L L T T       I D V T K M E D     D S E D E E N

V W R P W Stop
V W R P W
```

7 5'3'

```
M A T K Y E V S K T Y Q Y R K V M K P L L E R R R R
M A T K Y E M S K T Y Q Y R K V M K P L L E R K R R

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

R A D I L E V T V D H L R K L K Q A R I E A T A A A A K A T T N T T P E
K A D I L E V T V Q H L R K L K E S K K H V P A N                   P E

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

L R L N Q L E T P T A A P K P M N A P L S I I C G S G G S S S S S S N C S S N S R E A C
M R L N Q L E Q P M E Q P Q A V N T P L S I V C G S S S S S S T Y S S A S S C S S I

S P V S S G Y C S D S D C S V S S T Q A A Q N L L Q I S T   G Q
S P V S S G Y A S D N E S                     L L Q I S S P G Q

V W R P W Stop
V W R P W
```

Figure 4: Protein Sequence Comparison between *Drosophila pseudoobscura* and *Drosophila melanogaster.*

Protein sequence comparisons constructed between *D. pseudoobscura* (in purple) and *D. melanogaster* (in blue) for *E(spl) mγ* (left) and *m7* (right). The differences were highlighted in yellow.

```
                                                        X - Su(H) (no D.pseudo)

melanogaster -1420      AACCTTCACGAACCTTAATGTTTTCCAAGACTCAGGTTCCCACTCACCGTCGCAGCTAAC
simulans     -1457      AACCTTCACGAACCTTAATTTTTTCCAAGACTCAGGTTCCCACACACCGTCGCAGCTAAC
pseudoobscura-1150      AACTTCCCTTAGCAACATCTCGTACGGAAATACCAGAATTCCTTAACGTTTCCACCCAAA
                        *** * *   * *   *     * *  * *  *  *    *    **  *  ** * **

melanogaster -1360      AAAATTATCCGTAGCAAGTAGAAAACACCTTCCACAAAAGTCCCTATACTCGGACTATAT
simulans     -1397      AAAATTATCCGTAGCAAGTAGAGAACACCTTCCACAAAAGTCCCTATATTCGGACCATAT
pseudoobscura-1090      GAAA--ACCTGCAAT--GTCAA----ACCTAGCACAAAACTCCCTACAGTTCACACAC-T
                         ***  * * * *    **  *    ****  ******* ****** * *      *  *

melanogaster -1300      ACGCCTAGAAAAACCTCTAGGAACAAATGTACCTTTCTAAC-ATAGCAGATAGTTTAGGA
simulans     -1337      CCGCCTAGAAAAACCCCTAGAAACAAATGTATCTTTCTCAC-ATAGCAGATAGTTTAGGA
pseudoobscura-1039      ACGCCTAGAAAGTCCCT---AAATCAGTGTCCCTGGCAAATTACAACGCATCAACCTTAC
                         **********  **      **  * ***  **  *  *  * * *  **
                                                       IX - Su(H) D.pseudo

melanogaster -1241      AAAATTCTTTACAAGTAGAAATACATCGAGAAATCCATAAACACTTACTCTAAAACCTGC
simulans     -1278      AAAATTCGTTACAAGTAGAAAGACATCGAGAAATCCGTAAACACTTACACTAAAACCTGC
pseudoobscura-982       AAGACCCACCTTATTTCCAAA-AC-TCG---CATTCCCACGCACC-GTTCCAGAGTCCGG
                        ** *  *     *  *  *** ** ***    ** *  *  ***     * * *  * *

melanogaster -1181      AAGTCGTAAAATTGCACGCATGTTCCGTTTGGTAAAACCCCAAAAATCACATTCGAAAAC
simulans     -1218      AAGTCGTAAAATTGCACGCATGTTCCGTTTGGTAAAACCCCAAAAATCACATTCGAAAAC
pseudoobscura-928       AAATT----AACCAAA-GCA----------AGTAGAAC---ACAGGCCACCTTTAAATCC
                        ** *     **    * ***           *** ***   * *   *** **  **  *
                                 IX - Su(H) D.m. and D.s.

melanogaster -1121      TCACATTCCCACGACTTTGAATAGAGGTTCTCAAATTAACCGTAGCAAGTTGAGAACAGG
simulans     -1158      TCACATTCCCACGACTTTGCATAGAGGTTCTCAAATTAACCGTAGCAAGTTGAGAACAGG
pseudoobscura-886       GTAGAATACTGAAATATT-CACAGG---TCCCAAACAACACGTACATAGATGGAAA--AC
                          * * * *    *  **  * **    ** ****  *  ****   ** **  **

melanogaster -1061      CAACAAAAAATCCGTAAGATACTGATTTCCAAACCTGCGAGTCCCAAAATTGCACACATC
simulans     -1098      CAACAAAAAATCCGTAAGGTACTGGTTTCCAAACCTGCGAGTCCTAAAATTGCACACATC
pseudoobscura-832       CTTCATTCAGTCGGTTCAGTA--GGTATCCAGAAATAC-----CCACATGTGTGCAAAT-
                        *  **   * ** **    **  * * **** *  * *     * * *  **  ** **
                                                                   VIII - Su(H)

melanogaster -1001      TTCGCCTTGGAAAACCCCATTTCCGACATCCCAAAACTCGCATTCCCACGATTCCTACAG
simulans     -1038      TTCCCCTTGGAAAACCCCATTTCCGACATCCCAAAACTCGCATTCCCACGATCCCTACAA
pseudoobscura-780       TGCACC---GCATACACCTTCCCAGACTCACCT-------TATTCCCACGCACCGT-CAC
                        * * **   * * ** ** *  * ***   **         *********   * * **

melanogaster -941      AAATTTCTAAAACTTTGCCTAGCAAGTAGATAACAGTATGCGTAAATTCCGTAAACCTTA
simulans     -978      AAATTTCTAAAACTTTGTCTAGCAAGTAGATAACAGGATGCGTAAAATCCGTAAACCTTA
pseudoobscura-741      AGAGTCCGGAAAT---------CAAGTAGAAATCCGTATTCGCAGGTCCAAAAAAACACG
                       * * * *  ***          ******** * * * ** ** *    *   *** *

melanogaster -881      AATTTCACGTAGTTGGAAAAGCCTTGAAGAATTCATTCAAGCAAGGCACCCTCAAAATTC
simulans     -918      AATTTCACGTAGTTGGAAAAGCCTTGAAGACTTCATTCAAGCAAGGCACCCTCAAAATTC
pseudoobscura-690      TACAT-ACATCGGTAGGTGGACTGCCTTCA-TTCAGTCGGTCGCAGAGATTCGGAAATTC
                        *  * ** * * * *     *       * **** **   *   *        ******

melanogaster -821      CCGCATGTGTTTCATCGGTTTGGCCCACTGTGCCATGTGCATAAGGGCTCGGAACTCGGA
simulans     -858      CCGCATGTGTTTCATCGGTTTGGCCCACTGTGCCATGTGCATAAGGGCTCGGAACTCGGA
pseudoobscura-632      CCGCATGTGTG-CAATGGTTTGACCCACTGTGCCATGTGCATAAGGGCTCGGACCTCGGG
                       **********  **  ****** ****************************** *****

melanogaster -761      G------------CTCGGACTCGGGACTCGCACACGGGCTACCTCCGATTATAACTTATA
simulans     -798      G------------CTCGGACTCGGGACTCGCACACGGGCTACCTCCGATTATAACTTATA
pseudoobscura-573      GACTCGGATCAGACACGGACTCGGGGTTCGTTCGCTCCCTACCTCCGATTATAACTTATA
                       *            * **********  ***  * *   **********************
                                             VII - N-box          VI - Su(H) Paired

melanogaster -713      ACACCAACGAGCGAGAAAAATCTTGTGGGAAACTTGAGG GCAAAGTGTTTCCCACGATT C
simulans     -750      ACACCAACGAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAGTGTTTCCCACGATTC
pseudoobscura-513      ACACCAACGAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAGTGTTTCCCACGATTC
                       ************************************************************

melanogaster -653      GAATGTCAGTGGAAAGGAAAACAGAGAAGCAGCATGTGCAACCACCGTCGCACCACCACC
simulans     -690      GAATGTCAGTGGAAAGGAAAACAGAGAAGCAGCATGTGCAACCACCGTCGCACCACCACC
pseudoobscura-453      GAATGTCAGGCTAAAAGAAAACAGAGGAG-AGCACACACAAGAA--AACGCAC----ACA
                       *********   *** ********** ** ****    *** *    *****    **
                                                                 V - N-box

melanogaster -593      ACTGCACCA-TCGCATCTTCATCCCCTGCCGAGACTCACTTACACCGAACACCAGTGGCC
simulans     -630      ACTGCACCA-TCGCATCTTCATCCCCTGCCGAGACTCACTTACACCGAACACCAAGGGCC
pseudoobscura-400      CGTGCATCGGCCGCATGTGCA------------ACCAACCACCACCG----CCAAGAGCA
                         **** *   ***** * **            **  **   *****    ***   **
```

```
melanogaster -534      ACGGCAATA-GTACTGAGAACATATTTAAAGTTCTAGAAAATGTTTATA---TTTTCAAA
simulans     -571      ACAGCAATA-GCACTGAGAACATATTTAAAGTTCTAGACAATTTTCAGAGTTTTTTTAAA
pseudoobscura-356      GCAGCAGCAAGCAGCAACAACA-----ACAGCTAT-------------------------
                        * ***  * * *   * ****     * ** * *

melanogaster -478      ATCACTACTA--------------------TTAAGAAGTGTGCTAATTTAATTTTTTAAG
simulans     -512      GTCACTAGTGGATTTCGCCCGCTAGTTTGCCTACAAACTGCACAAATTTAATTTTTAAGG
pseudoobscura-326      -CCACCAACA---------------------CAGGAATGGAGGAGACACAC---------
                         *** *                         *  **  *     *   *

melanogaster -438      TGTAAACGTGTAACATTATGTTAATTTCTTTTCGTTTAATAAAACTACATTTAAGTTCTT
simulans     -452      TGTA------TAACTTTATATTAACTACTTTTTGCTTGATAAAATAACATTTAAGTACTA
pseudoobscura-297      -------------CATCACATCATCTTCATCC--CCTGCCAAGACGACTT----GTTTGT
                                    * * *  * *  * * *      *   ** *  ** *    **

melanogaster -378      TAAGGGAAACCCATTTGAAGT-AATTGTTTAAAGCGTAACGAACTCCTTA-GGTATATAA
simulans     -398      TAAATTAAGAGCATTCAAGACCAATAGTAAAATATATGATGTATTCCTCAAGGTTTATAG
pseudoobscura-256      CGAGAAAAACCCACC-----------ACCGAAAGCAGG----------CGAGGATCGTGG
                         *   **   **                 **                   **    *

melanogaster -320      CAA------------------------------------------CAAATTTGTATATTC
simulans     -338      AAAATACTTTTGATCGTCTCTTTAATATGTCTTAAAAACGTTCTACAAAAATTTGTATTC
pseudoobscura-217      C-------------------------------------------------------TGTCG
                                                                  * *

                                                       IV - Su(H)     III - E-box

melanogaster -278      ACTCGAAGAACTATAACTGCTTTAAACACATTTCTCATACTATTTTTACCGCAGGTGTGT
simulans     -278      ACTCAAAGAACTATAGCTGCATTAATCACATTGCTCATACTATTTTTACCGCAGGTGTGC
pseudoobscura-211      GCCCGAGTAACGAT--------------------TCGTAACG-----ACGACACGAACGA
                        * * *  *** **                    ** **        **  ** *   *


melanogaster -218      TCCAAACTTGTTTGTGGAGAAAAACCCACCACCGAAAATGCAGGCGAGAGGATCGTGTCC
simulans     -218      TCCAAACTTGTTTGTGGAGAAAAACCCACCACCGAAAATGCAGGCGAGAGGATCGTGTCC
pseudoobscura-176      C---AATCGGCTCG---------ACCCGACCCAGTGACTGACGGCAGGTGG------TCG
                           **   * * *         ****  * * *  * **  ***  * **      **
                                                 II - E-box

melanogaster -158      CAGTGACTGAGTAAAAGTAGTTTGCTCCGCAGGTGGTGGTTCT-TCGGAGGCTCCGCAGC
simulans     -158      CAGTGACTGAGTAAAAGTAGTTTGCTCCGCAGGTGGTGGTTCT-TCGGAGGCTCCGCAGC
pseudoobscura-134      CAATGGT---------GGTGCTCGCTCCG---GTGGTGCCTCTCTCGTTTCCTGCGCA--
                       ** **           *  * * ******   ******  *** ***    ** ****
                                                       I - N-box

melanogaster -99      TCTCTCTTCGCTCTTTCTCCGAGCAGCGTGCGCTGAGCACGAGACGCTCTCGCGGCGGCA
simulans     -99      TCTCACTTCTCTCTTTCTCCGAGCAGCGTGCGCTGAGCACGAGACGCTCTCGCGGCGGCA
pseudoobscura-88      ------TGCGCGCTCTCTCTG-----CGCGCGCTGAGCACGAGACGACTCTGCTGCGGCA
                            * * * ** **** *     ** *****************     ** ******

melanogaster -39      CGTGCAGCTATAAAAGCAGCGGTAACCGGAGACGAATGC~AACATTTCGAACGCAATCGTC
simulans     -39      CGTGCAGCTATAAAAGCAGCGGTAACCGGAGACGAATGC~AACATTTCGAACGCAATCGTC
pseudoobscura-39      CGTGCAGCTATAAAAGCGGCGGTAACTGGAGACGAGAGC~AACATTCTCAACAAATTCGAC
                      ***************** ******** ********  **~******   ***  * *** *
```

Figure 5: ClustalW alignment for *E(spl) m7* upstream regulatory region.

The ClustalW program aligned the *E(spl) m7* upstream regulatory sequences for *D. melanogaster, D. simulans* and *D. pseudoobscura*. Regulatory binding sites were highlighted in their respective colors, and the sites were identified using the roman numerals to differentiate between sties.

```
melanogaster -872       ATCCATCCATTACAAAGGTCCTGGAACGCGAGCAACTTACAGCCGAAAGTGTGTCCACAG
simulans     -878       ATCCATCCATTACAAAGGTCCTGGAACGCGAGCAACTTACAGCCGAAAGTGTGTCCACAG
hydei        -932       AAAAGTC------GAAAGTGTGTTACCGTGCGCCACATGCGAGT-AACGAGCAGCAAGAG
pseudoobscura-1068      A-CAATCGATTACAAAAGTCCTGGAACGCGTACAACTTACAGCCGAAAGTGTGTCCACCG
                        *    **       ** **     * ** *  * ** * *     ** * *   * *  *
```

VIII - High Affinity N-box (*D.hydei* 1bp dif)

```
melanogaster -812       -CGAGCCACACGCGACGAGCGCTCCAGGATCAGCATCGCT-GGCCCCTTTTACATGGCGG
simulans     -818       -CGAGCCACACGCGAGCAGCGCTCCAGGATCAGCATCGCT-GGCCCCTTTTACATGTCGG
hydei        -879       TCGAGCTGCACGCATCGAGCAGCATCGC-TTGGC--CCCT-GCCCGACCGCGACTCTCAG
pseudoobscura-1009      -CGAGCCACACGCGAGCAGCGCTCCAGGAGCAGCATCGCTTGGCCCCTTTTACAGAGCAG
                         *****  *****    ***      *     **  * ** * **            * *
```

```
melanogaster -754      ATGGAGGGAGA-GGCAGAGGTCCTGC--C-GGTCGGGGGG--------------------
simulans     -760      ATGGAGGGAGA-GGCAGAGGTCCTGC--CAGGTCGGGGGG--------------------
hydei        -823      TTACAAAAGGT-GTCCGTGGGCCGGGGGCAGAGCAGAGAC--------------------
pseudoobscura-950      AGGCAGCGCCTCAGCAGAGATCAGGGATCGGGTCAACGAGGGAAGGAGCAACGGCAGAAT
                           *         * * *  *  *   * *  *   *
```

VII - N-box   VI - Su(H)

```
melanogaster -718      ---ATCGGGAAGTCG----TACACGCATTAAGCGCACTCGACGCACACGAGCAATGTTCC
simulans     -723      ---ATCGGGAAGTCG----AACACACATTAAGCACACTCGACGCACACGAGCAATGTTCC
hydei        -784      ---GTTAGAGCAACGG--GAACGCACATTAAGCACACTCGTCGCACACGAGCAATGTTCC
pseudoobscura-890      GGAACGGGAAAGTCGAACGAACACACATTAAGCACACTCGACGCACACGAGCAATGTTCC
                              *     **     ** * ******** ****** *******************
```

```
melanogaster -665      CACAGGATC--AT-AAC--------AGCCGACCAACAAGTGCAGGATGCCTCTGTCCTTG
simulans     -670      CACAGGATC--AT-AAC--------AGCCGACCAACAAGTGCAGGATGCCTCTGTCCTTG
hydei        -729      CACAGGATC--ATCAACTGGCACAAAGCCGGCCAACAAGTGTAGGACGAAA--GGCCGAA
pseudoobscura-830      CACAGGATCGGATCGGGATCATAGCAGCCGACCAACAAGTGCAGAATGCCTCGGCCTCTG
                       *********  **            ***** ********** ** * *     * *
```

```
melanogaster -616      TCCCTTCCTCTGCCACCAAACACGAACCCGATCCAATTTGGAATGCCAACGGTAAACAGG
simulans     -621      TCCCTTCCTCTGCCACCAAACACGAACCCGATCCAATTTGGAATGCCAACGGTAAACAGG
hydei        -673      ATGGAATACCAATGCACAAACAGGAAGT-GCTTCTCTTTGCAGCTCCGCC--TAAGTGAA
pseudoobscura-770      CCTCTGCCTC--TCCCTCGGCACGAATACGATCCAAGGTGGAATGCCAACAGTAAACAGG
                                *          ** ***   * * *    ** *   **  *  ***
```

V - Modified Su(H) in *D.hydei*

```
melanogaster -556      AAGTCGTCGAGGAATCGCAACGCCTCGTCCTTATCCTTATCCCCATCCCCCGACGGAGGT
simulans     -561      AAGTCGTCGAGGAATCGCAACGCCTCGTCCTTATCCTTATCCCCATCCCCCGGCGGAGGT
hydei        -616      ACGCCCCAGCAAAATGTTGCTGCTCCTAGC--ATTCCCACGCCAACGGCCAAGCAG----
pseudoobscura-712      AAGTTGTCGACGAATCGCAACGCCCCATCC----CCCCATCCCTGTATCCCTATCGGA--
                       * *     *   ***      **  *   *     *  *  **     **     *
```

V - Su(H)

```
melanogaster -496      CCCCGCTATCGTGGGAACCCCGGAG---GACCCCGATGAGATGGCATAATGCAGGCAGTT
simulans     -501      CCCCGCTATCGTGGGAACCCCAGAG---GACCCCGATGAGATGGCATAATGCAGGCAGTT
hydei        -562      --CCTCTGTTGTTGGCACA---------------AATGAGATGGCATGAAGCGTGCAGTT
pseudoobscura-658      -------GCCGTGGGAACTCCGAGGCGAGACCTAAATGAGATGGCATAAAGCAAGCAGTT
                                 ** ** **                 ************ * **  ******
```

```
melanogaster -439      CATTATCCTTATGCCCCTCAC------CCTGCCATGCCATAGTCAAATCATAACACACAA
simulans     -444      CATTATCCTTATGCCCCTCAC------CCTGCCCTGCCATAGTCAAATCATAACACACAA
hydei        -519      CATTATCCTTTCGCTCG----------CCCATATTCTCTTAGGCAAATCATAACAC--AA
pseudoobscura-605      CATTATCCTTCTGCTCGCCTCGACATTTGTGTATGGGTATAGTCAAATCATAACACACAA
                       **********  ** *                       *** *************  **
```

IV - E-box   *D.pseudo* identical III sequence

```
melanogaster -385      ATCTAGAAACGGCAGCTGTTC------------------GCTC-----------------
simulans     -390      ATCTAGAAACGGCAGCTGTTC------------------GCTC-----------------
hydei        -471      ATCTAGAAATGGCAGCTGCGCCA-----------CAAACGCTCGCAAACACACA------
pseudoobscura-545      ATCTAGAAACGGCAGCTGTTCCTGCCCGCAAATTCCCATGCTCGAGTCCAGGTCCGAGTA
                       ********* ********  *                  ****
```

III - Su(H) [Part conserved with *D.pseudo*]

```
melanogaster -360      ----TGCAAA--TTC--CCATGCCCATGCC-GT----GGCCATGCTC--GGCTGCT----
simulans     -365      ----TGCAAA--TTC--CCATGCCCATGCCCAT----GGCCATGCTC--GGCTGCT----
hydei        -428      -AACTGCAAAAATTC--CCATGGAAAGGCTAGGC--GGAGCATGGTTCTGGCTGCG----
pseudoobscura-485      CGAGTACGAGTATGAAACCATGCCCATATCCATCCAGGGCCAAGCCCCTGACTACGAATA
                           * * *   *    *****   *           *  ** *     * ** *
```

II - E-box

```
melanogaster -319      --------TTTTGC----------GGGCAGGTGAGCGAG-----------------TCGA
simulans     -323      --------TTTTGC----------GGGCAGGTGAGCGAG-----------------TCGA
hydei        -377      --------CTTCGTTGCCCATT--GAGCAGGTGAGCCTGCTCAGA--------ATATTGG
pseudoobscura-425      TGCTCATGCTCCGTTGCTCTTTCGGGGCAGGTGAGCGAATCGAGCGAGCGAGCATATTGG
                                *  *           * **********                    * *
```

```
                                 I - Paired Su(H)
melanogaster -294      TGTGAGAAACCGAGTAGGAAAGTGTTTCCCACG--------------------------A
simulans     -298      TGTGAGAAACCGAGTAGGAAAGTGTTTCCCACG--------------------------A
hydei        -335      TGTGAGAAACCAAGGAGGAAAGTGTTTCCCACG--------------------------A
pseudoobscura-365      TGTGAGAAACCGAGGAGGAAAGTGTTTCCCACGGTTCCAAGCACACACAATGCGAATCGA
                       *********** ** ******************                          *
                                           Ttk-69                D.hydei's extra Ttk-69

melanogaster -260      TCCTGGCA-GCGATCCTGCTCCCTGGCCCGTTT-----GATAA--GGGTGCCG--GGGCC
simulans     -264      TCCTGGCA-GCGATCCTGCTCCCTGGCCCGTTT-----GATAA--GGGTGCCG--GAGCC
hydei        -301      TACTGCCAAGCAAACACACACACGCAAACGCATACAAAGGCAA--TGATCCTGCAAAGGC
pseudoobscura-305      TCCTTGCA-GCGATCCTGCTCCCTTAACCGTTGATAAGGGTGCCCTGGCCAAGCCCAGGC
                       * **  ** ** * *   * * *     **        *       *     *    * *
                                   Ttk-69 D.hydei's N-box

melanogaster -210      AACAATAAGAAATC-----GCC--CGAACA--ATAAGGA-GCGGTCTCAAGAAGCGGAAG
simulans     -214      AACAATAAGAAATC-----GCC--CGAACA--ATAAGGA-GCGGTCTCAAGAAGCGGAAG
hydei        -243      AGCGATCCTGCTCCCTGGTGCC--CGCTCATGATAAGGGTGCCACAATAAGAAAGCGAAC
pseudoobscura-246      CAGACTTCGAATCGGGGACGGTGACGGTGACGGGGACAAGGCAACAATAAGAAAGCGGAC
                            *             *    **   *     *    **      *****   * *

melanogaster -160      GGTCGGTACCG-AGCAACCCTCATGCCAGC-ACACGAAGCGTGTCCCCGTCCCTTTTGCT
simulans     -164      GGTCGATGCCG-AGCAACCCTCATGCCAGC-ACACGAAGCGTGTCCCCGTCCCTTTTGCT
hydei        -185      GAACAATAAGACAGCAGCGGAAGGGTCGACGACGTCGTCATCGTCGTCGTCGTCGTCGCA
pseudoobscura-186      AAACAATAAGA-AGCAGCCTCAAGA--AGCGGAAGGGCGACCAACCCTCATGCCAGCACA
                          *  *     **** *           *              *             *

melanogaster -102      CCTCGAATGCCAGCACACACGAC---G------ACGACG------ACGATCCCTTTGGCC
simulans     -106      CCTCGAATGCCAGCACACACGAC---G------ACGACGTCGACGACGATCCCTTTGGCC
hydei        -125      ACCCTCATGCCAGCACACGCTGC---ACACGAAATGCCAGCACAGACGAGCCCAAAGCCG
pseudoobscura-129      CGCTTCGCTCCGTCCCGCTCTGCCCGACACGAAATGCCAGCACAGG-GGCCCAGCCAGCA
                                **  * * * *  *          * * *         *  **      *

melanogaster -57      CATT-------------TTTTGGGGGTTGTTTCTGGCCA----GCTATATAAGGC--CGA
simulans     -55      CATT-------------T--TGGGGGTTGTTTCTGGCCA----GCTATATAAGGC--CGA
hydei        -68      AATGCCGAG-----GGATCATGGGGGTTGTTTTTGGCTGCTTGGCTATAAAAGGCAGCGG
pseudoobscura-70      CACACGAGATCCCTGGATCATGGGGGTTGTTTCTGGCCA----GCTATAAAAGCCA-GGG
                       *               *  ************ ****      ****** *** *   *
                                                                             -1
melanogaster -16      TCACCGAAGGTCTAGC~AACACACAACGACTTCGACTGTCGGAGCAGCAAGAGCTCAAAGA
simulans     -16      TCACCGAAGGTCTAGC~AACACACAACGTCTTCGACTGTCAGAGCAGCAAGAGCTCAAAGA
hydei        -13      CCAGCGAGG---CAGC~AGTACATAACGTCTTGCACACTCAGAGCAGCAACAGCTCAGCAA
pseudoobscura-15      CCAGGGCCGG-CAGAC~AGCACACAACGTCTTCAGCCGTCGGAGCAGCCACAGCTCAAAAG
                       **  *  *      *~*  *** **** ***   *  ** ******* * ******
```

Figure 6: ClustalW *E(spl) mγ* upstream regulatory region alignment.

The ClustalW program aligned the *E(spl) mγ* upstream regulatory sequences for *D. melanogaster, D. simulans, D. hydei* and *D. pseudoobscura*. Regulatory binding sites were highlighted in their respective colors, and the sites were identified using the Roman numerals to differentiate between sites.

Table 3: Summary of upstream binding site ClustalW alignments.

Upstream Sequence	E-Boxes	N-Boxes	Su(H)	Paired Su(H)
m7	III: M=S P same sequence located between III and II II: M=S	VII: M=S=P V: M=S no sequence for P I: M=S=P	X: M=S no sequence for P IX: M=S=P, P located upstream VIII: M=S=P IV: M, S-one bp off, no P	VI: M=S=P
mγ	IV: M=S=H=P II: M=S=H=P	VIII: M=S (H one bp off)=P VII: M=S=H=P -257 H	VI: M=S=H=P V: M=S=H(modified)=P III: M=S=H=P(either upstream or not present)	I: M=S=H=P

Similarities and differences in the upstream regulatory binding sites for *D. melanogaster, D. simulans, D. hydei* and *D. pseudoobscura* are noted using the convention M=*D. melanogaster,* S=*D. simulans,* H=*D. hydei,* and P=*D. pseudoobscura*. Upstream regulatory sites are broken into type of site and identified with their respective Roman numeral.

Table 4: Percent Identity of each *Drosophila* species with respect to the *D. melanogaster* reference (MacDonald and Long 2005).

Species	Percent Identity
D. simulans	96.2%
D. sechellia	95.7% *
D. yakuba	94.2%
D. erecta	*94.1%*
D. persimilis	91.9% *
D. pseudoobscura	*89.2%*
D. ananassae	*89.2%*
D. grimshawi	87.1%
D. virilis	87.0%
D. mojavensis	86.1%

As each species was selected in BLAT program to be aligned against *D. melanogaster*, relative percent identities were obtained. These percentages were used when creating EvoPrint alignments and assessing relative species divergence.

*sequence was not available at the time this first BLAT alignment was performed; however, these two species are used in the second BLAT alignment

Capital letters represent nucleotides in the *D. melanogaster* reference sequence that are conserved in *D. simulans* and *D. yakuba*.

aaaatggtttctttttttaggaatttttaaacatcataaccagacaagaacatcttttcaaatatttctttatttcacccATCTGGgtacagtaatcccat
acatagatatTCTGtAAAcTGCAACCTTCACGAACCtTAATgTTTTCCAAGACTCAGG**TTCCCAC**tCACCGTcGCAGCTAACAAAATTATCCGTAGCAAGTAg
AAAACACCTTCCACAAAAgTCCCTAtAcTcgGACtATATaCGCCTAGAAAAACCtCtAGgAACAAATgTAcCTTTCtaACaTAGCAgATAGTTTAGGAAAAAT
TCtTTACAAGTAGAAAtACAtCGAgAAATCCaTAAACACTTACtCTaAAACCTGCaAGTCGTAAAATTGCACGCATgTTCCGTTTGGTAAAACCCCAAaAATc
ACATTCgAAAAcTCACa**TTCcCACG**ACTTTgAatAGAGGTtCTcAAATtAACCGTAGCAAGTTGAGAACAGGCAACaaaaaatcggcaagaTACTGaTTTCcA
AACCTGCGAGTCCcAAAATTGCACACATCtTCgCCtTGGAAAaCCCCATTTCcGACATCcCAaAacTCGCA**TTCCCACG**ATCCCtacagAAATTgCTAAAACT
TTGtCTAGCAAGTAGATAaCAGtATGCGTAAAtTCCGTAAACCTTAAaTTtCACGTAGTTGGAAAAGCCTTGAAgAaTTCATTCAAGCAAGGCACCcTCaAAA
TTCCCGCATGtGTTTCATCGGTTTGGCCCACTGTGCCATGTGCATAAGGGCTCGGAACTCGGAGCTCGGACTCGGGACTCGCACACGGGCTACCTCCGATTAT
AACTTATAACACCAACGAGCGAGAAAAT**CTTGTGGAAACTTGAGGGCAAAGTGTTTCCCACG**ATTCGAATGTCAGTGGAAAGGAAAACAGAGAAGccAGCAT
GTGCAACCACCGTCGCACCACCACCACTGCACtATCgGCATCTTCATCCCCTggCCGAGACTCACTTAaCACCgaag**caccag**tggccatagcaatagtactg
agaacatatttaaagttttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaatttttttaggtgtaaacgtataacattatg
ttaatttctttttgtttaataaaacgacatttaagtgctttaagggaaacccatttgaagtaattgttaaaaacgtaacgtacttcttaggtatataacaaca
aatttgtatattcgctcgaagaactatagctgctttaaacacat**ttctCAT**ACTATTTTtaCCg**CAGGTG**TGtTCCAAACTTGTTTGTGGAGAAAAACCCACC
ACCGAAAATGCAGGCGAGAGGATCGTGTCCCAGTGACTGAGTAAAAGTAGTTTGcTCCG**CAGGTG**GTGGTTCTTCGGAGgCTCCGCAGCTCTCtCTTCTCTCT
TTCTCCGAGCAGCGTGCGCTGAGCACAGACGCTCTCGCGGCGG**CACGTG**CAGCTATAAAAGCAGCGGTAACCGGAGACGaAtGCAACATTTCgAACGCAATCG
TCGtGCAGTCAACATCTCAGGATAACCTcCTCCGAAAGATCGAAGAAGTTCTTCAACAagtcgttcttccatttacctaaacgcaacaaataAACAAAACAAA
cACACAATGGCCACCAAATACGAgATGTCgAAaACCTATCAGTACCGCAAG

Capital letters represent nucleotides in the *D. melanogaster* reference sequence that are conserved in
***D. simulans*, *D. yakuba* and *D. erecta*.**

aaaatggtttctttttttaggaatttttaaacatcataaccagacaagaacatcttttcaaatatttctttatttcacccatctgggtacagtaatcccat
acatagatattCTGtAAAcTGCAACCTTCACGAACCtTAATgTTTTCCAAGACTCAGG**TTCCCAC**tCACCGTcgCAgCTaACAAAATTATCCGTAGCAAGTAg
AAAACACCTTCCACAAAAgTCCcTAtAcTcgGACtATATaCGCCTAGAAAAACCtCtAGgAACAAATgTAcCTTTCtaAcatAGcAgATAGTTTaGGAAAAAT
TCtTTACAAGTAGaAAtACAtCGAgAAATCCaTAAACACTTACtCTaAAACCtGCaAGTCGTAAAATTGCACGCATgTTCCGTTTGGTAAAACCCCAAaAATc
aCATTCgAAAAcTCACa**TTCcCACG**ACTTTgAatAGAGGTtCTcAAATtAACCGTAGCAAGTTGAGAACAGGCAACaaaaatcggcaagaTACTGaTTTCcAA
ACCTGCGAGTCCcAAAATTGCACACATCtTCgcCtTGGAAAaCCCCATtTCcGACATCcCAaAacTCGCA**TTCcCACG**ATCcCtacagAAATTgCTAAAACTT
TGtCTAGCAAGTAGATAaCAGtATGCGTAAAtTCCGTAAACCTTaAaTTtCACgTAGTTGGAAAAGCCTTGAAgAaTTCATTCAAGcAAGGCACCcTCaAAAT
TCCCGCATGtGTTTCATCGGtTTGGCCCACTGTGCCATGTGCATAAGGGCTCGGAACTCGGAGCTCGGACTCGGGACTCGCACACGGGCTACCTCCGATTATA
ACTTATAACACCAACgAGCGAGAAAAT**CTTGTGGAAACTTGAGGGCAAAGTGTTTCCCACG**ATTCGAATGTCaGTGGAAAGGAAAACAGAGAAGccAGCATG
TGCAACCACCGTCGCACCACCACCACTGCACtATCgGCATCTTCATCCCCTggCCGAGACTCACTTAaCACCgaag**caccag**tggccatagcaatagtactga
gaacatatttaaagttttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaatttttttaggtgtaaacgtataacattatgt
taatttctttttgtttaataaaacgacatttaagtgctttaagggaaacccatttgaagtaattgttaaaaacgtaacgtacttcttaggtatataacaacaa
atttgtatattcgctcgaagaactatagctgctttaaacacat**ttctcat**actattttttaccg**CAGGTG**TGtTCCAAACTTGTTTGTGGAGAAAAACCCACCA
CCGAAAATGCAGGCGAGAGGATCGTGTCCCAGTGACTGAGTAAAAGTAGTTTGcTCCG**CAGGTG**GTGGTTCTTCGGAGgCTccgCAGCTCTCtCTTCTCTCTT
TCTCCgAgCAGCGTGCGCTGAGCACAGACGCTCTCgCggCGG**CACGTG**CAGCTATAAAAGCAGCGGTAACCGGAGACGaAtGCAACATTTCgAACGCAATCGT
CGtGCAGTCAACATCTCAGGATAaCCTcCTcCGAAAGATCGAAGAAGTTCTTCAACAagtcgttcttccatttacctaaacgcaacaaataAACAAAACAAAc
acacAATGGCCACCAAATACGAgATGTCgAAaACcTATCAGTACCGCAAG

Capital letters represent nucleotides in the *D. melanogaster* reference sequence that are conserved in *D. simulans*, *D. yakuba*, *D. erecta* and *D. ananassae*.

aaaatggtttctttttttaggaatttttaaacatcataaccagacaagaacatcttttcaaatatttctttatttcacccatctgggtacagtaatcccat
acatagatattctgtAAAcTGCAAccttcacgaaccttaatgTTTTCCAAgAcTCAGG**TTCCCac**tcaccgtcgcagctaacaaaattatccgtagcaagtag
aaaacaccttccacaaaagtccctatactcggactatatacgcctagaaaaacctctaggaacaaatgtacctttctaacatagcagatagtttaggaaaaat
tctttacaagtagaaatacatcgagaaatccataaacacttactctaaaacctgcaagtcgtaaaattgcacgcatgttccgtttggtaaaaccccaaaaatc
acattcgaaaactcaca**ttcccacg**actttgaatagaggttctcaaattaaccgtagcaagttgagaacaggcaacaaaaaatcggcaagatactgatttcca
aacctgcgagtcccaaaattgcacacatcttcgccttggaaaaccccatttccgacatcccaaaactcgca**ttcccacg**atccctacagaaattgctaaaact
ttgtctagcaagtagataacagtatgcgtaaattccgtaaaccttaaatttcacgtagttggaaaagccttgaagaattcattcaagcaaggcaccctcaaaa
ttcccgcatgtgtttcatcGGtTTGGCCCACTGTGCCATGTGCATAAGGGCTCGgAACTCGGAGCTCGGACTCGGGACTCGCACACGGgcTACCTCCGATTAT
AACTTATAACACCAACgAGCGAGAAAAT**CTTGTGGAAACTTGAGGGCAAAGTGTTTCCCACG**ATTCGAATGTCaGTGGAAAGGAAAACAGAGAAGccAGCAT
GTGCAACCACCgTCgcACCACCACCACTGCACtATCggcATCTTCATCCCCTggCCGAgactcacttaacaccgaag**caccag**tggccatagcaatagtactg
agaacatatttaaagttttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaatttttttaggtgtaaacgtataacattatg
ttaatttctttttgtttaataaaacgacatttaagtgctttaagggaaacccatttgaagtaattgttaaaaacgtaacgtacttcttaggtatataacaaca
aatttgtatattcgctcgaagaactatagctgctttaaacacat**ttctcat**actattttttaccg**caggtg**tgttCCAAACTTGTTTGTggAGAAAAACCcACC
ACCgAAAATGCAGGCGAGAGGATCGTGTcccagtgactgagtaaaagtagtttgctccg**caggtg**gtggttcttcggaggctccgcagctctctcttctctct
ttctccgagcagcgtGCGCTGAGCACAGACGCTCTCgCggcGG**CACGTG**CAGCTATAAAAGCaGcGGTAACcGGAGACgaatgCAACATTtcgAACGCAAtCg
TCGtGCAGTCAACATCTCAGGAtaacctcctccgaaagatcgaagAAGTTCTTCAacaagtcgttcttccattt

Capital letters represent nucleotides in the *D. melanogaster* reference sequence that are conserved in *D. simulans*, *D. yakuba*, *D. erecta*, *D. ananassae* and *D. pseudoobscura*.

aaaatggtttctttttttaggaatttttaaacatcataaccagacaagaacatcttttcaaatatttctttatttcacccatctgggtacagtaatcccat
acatagatattctgtaaactgcaaccttcacgaaccttaatgttttccaagactcagg**ttcccac**tcaccgtcgcagctaacaaaattatccgtagcaagtag
aaaacaccttccacaaaagtccctatactcggactatatacgcctagaaaaacctctaggaacaaatgtacctttctaacatagcagatagtttaggaaaaat
tctttacaagtagaaatacatcgagaaatccataaacacttactctaaaacctgcaagtcgtaaaattgcacgcatgttccgtttggtaaaccccaaaaatc
acattcgaaaactcaca**ttcccacg**actttgaatagaggttctcaaattaaccgtagcaagttgagaacaggcaacaaaaaatcggcaagatactgatttcca
aacctgcgagtcccaaaattgcacacatcttcgccttggaaaaccccatttccgacatcccaaaactcgca**ttcccacg**atccctacagaaattgctaaaact
ttgtctagcaagtagataacagtatgcgtaaattccgtaaaccttaaatttcacgtagttggaaaagccttgaagaattcattcaagcaaggcaccctcaaaa
ttcccgcatgtgtttcatcGGtTTGgCCCACTGTGCCATGTGCATAAGGGCTCGgAaCTCGGaGCTCGGActcgggactcgcACACGGgcTACCTCCGATTAT
AACTTATAACACCAACgAGCGAGAAAATCTTGTGGGAAACTTGAGGGCAAAGTGTTTCCCACGATTCGAATGTCaGtggAAAgGAAAACAGAGaagccaGCAT
GTGCAACCAccgtcgcACCACCACCactgcactatcggcATCTTCATCCCCTggccgagactcacttaacaccgaag**caccag**tggccatagcaatagtactg
agaacatatttaaagttttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaatttttttaggtgtaaacgtataacattatg
ttaatttcttttttgtttaataaaacgacatttaagtgctttaagggaaacccatttgaagtaattgttaaaaacgtaacgtacttcttaggtatataacaaca
aatttgtatattcgctcgaagaactatagctgctttaaacacat**ttctcat**actattttttaccg**caggtg**tgttccaaACTTGTTTGTggAGAAAAACCcACC
ACCgAAAatGCAGGCGAGagGATCGTGtcccagtgactgagtaaaagtagtttgctccg**caggtg**gtggttcttcggaggctccgcagctctctcttctctct
ttctccgagcagcgtgcgctgagcacagacgctctcgcggcGGCACGTGCAGCTATAAAAGCaGcGGTAACcGGAGACgaatgCAACATTtcgaacgcaatcg
tcgtgcagtcaacatctcaggataacctcctccgaaagatcgaagaagttcttcaacaagtcgttcttccatttacctaaacgcaacaaataaacaaaacaaa
cacacAATGGCCACCAAATACGAgaTgTCgAAaACcTAtCAGTACCGCAAG

Capital letters represent nucleotides in the *D. melanogaster* reference sequence that are conserved in *D. simulans*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura* and *D. grimshawi*.

aaaatggtttctttttttaggaatttttaaacatcataaccagacaagaacatcttttcaaatatttctttatttcacccatctgggtacagtaatcccat
acatagatattctgtaaactgcaaccttcacgaaccttaatgttttccaagactcagg**ttcccac**tcaccgtcgcagctaacaaaattatccgtagcaagtag
aaaacaccttccacaaaagtccctatactcggactatatacgcctagaaaaacctctaggaacaaatgtacctttctaacatagcagatagtttaggaaaaat
tctttacaagtagaaatacatcgagaaatccataaacacttactctaaaacctgcaagtcgtaaaattgcacgcatgttccgtttggtaaaccccaaaaatc
acattcgaaaactcaca**ttcccacg**actttgaatagaggttctcaaattaaccgtagcaagttgagaacaggcaacaaaaaatcggcaagatactgatttcca
aacctgcgagtcccaaaattgcacacatcttcgccttggaaaaccccatttccgacatcccaaaactcgca**ttcccacg**atccctacagaaattgctaaaact
ttgtctagcaagtagataacagtatgcgtaaattccgtaaaccttaaatttcacgtagttggaaaagccttgaagaattcattcaagcaaggcaccctcaaaa
ttcccgcatgtgtttcatcggtttggCCCACTGTGCcATGTGCATAAGGGCTCggAaCTCGgagctcggactcgggactcgcacacgggctacctcCGATTAT
AACTTATAACACCAACgAGCGAGAAAATCTTGTGGGAAACTTGAGGGCAAAGTGTTTCCCACGATTCGAAtGTCagtggAAAgGAAAACAGAGaagccagCAT
GTGCAACcaccgtcgcACCACCAccactgcactatcggcatCTTCATCCCCTggccgagactcacttaacaccgaag**caccag**tggccatagcaatagtactg
agaacatatttaaagttttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaatttttttaggtgtaaacgtataacattatg
ttaatttcttttttgtttaataaaacgacatttaagtgctttaagggaaacccatttgaagtaattgttaaaaacgtaacgtacttcttaggtatataacaaca
aatttgtatattcgctcgaagaactatagctgctttaaacacat**ttctcat**actattttttaccg**caggtg**tgttccaaACTTGTTTGtggAGAAAAACCcACC
ACCgaaaatgcaggcgaGagGATCGtgtcccagtgactgagtaaaagtagtttgctccg**caggtg**gtggttcttcggaggctccgcagctctctcttctctct
ttctccgagcagcgtgcgctgagcacagacgctctcgcggcGGCACGTGCAGCTATAAAAGCaGcgGTAACcGGAGACgaatgCAACATTtcgaacgcaatcg
tcgtgcagtcaacatctcaggataacctcctccgaaagatcgaagaagttcttcaacaagtcgttcttccatttacctaaacgcaacaaataaacaaaacaaa
cacacaATGGCcACCAAATACGAgaTgTCgAAaACcTAtCAGTACCGCAAG

Capital letters represent nucleotides in the *D. melanogaster* reference sequence that are conserved in *D. simulans*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. grimshawi* and *D. virilis*.

aaaatggtttctttttttaggaatttttaaacatcataaccagacaagaacatcttttcaaatatttctttatttcacccatctgggtacagtaatcccat
acatagatattctgtaaactgcaaccttcacgaaccttaatgttttccaagactcagg**ttcccac**tcaccgtcgcagctaacaaaattatccgtagcaagtag
aaaacaccttccacaaaagtccctatactcggactatatacgcctagaaaaacctctaggaacaaatgtacctttctaacatagcagatagtttaggaaaaat
tctttacaagtagaaatacatcgagaaatccataaacacttactctaaaacctgcaagtcgtaaaattgcacgcatgttccgtttggtaaaccccaaaaatc
acattcgaaaactcaca**ttcccacg**actttgaatagaggttctcaaattaaccgtagcaagttgagaacaggcaacaaaaaatcggcaagatactgatttcca
aacctgcgagtcccaaaattgcacacatcttcgccttggaaaaccccatttccgacatcccaaaactcgca**ttcccacg**atccctacagaaattgctaaaact
ttgtctagcaagtagataacagtatgcgtaaattccgtaaaccttaaatttcacgtagttggaaaagccttgaagaattcattcaagcaaggcaccctcaaaa
ttcccgcatgtgtttcatcggtttggCCCACTGTGCcATGTGCATAAGGGCTCggAaCTCGgagctcggactcgggactcgcacacgggctacctcCGATTAT
AACTTATAACACCAACgAGCGAGAAAATCTTGTGGGAAACTTGAGGGCAAAGTGTTTCCCACGATTCGAAtgTCagtggAAAgGAAAACAGAGaagccagcat
gtgcaaccaccgtcgcaccaccaccactgcactatcggcatCTTCATCCCCTggccgagactcacttaacaccgaag**caccag**tggccatagcaatagtactg
agaacatatttaaagttttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaatttttttaggtgtaaacgtataacattatg
ttaatttcttttttgtttaataaaacgacatttaagtgctttaagggaaacccatttgaagtaattgttaaaaacgtaacgtacttcttaggtatataacaaca
aatttgtatattcgctcgaagaactatagctgctttaaacacat**ttctcat**actattttttaccg**caggtg**tgttccaaacttgtttgtggagaaaaacccacc
accgaaaatgcaggcgagaggatcgtgtcccagtgactgagtaaaagtagtttgctccg**caggtg**gtggttcttcggaggctccgcagctctctcttctctct
ttctccgagcagcgtgcgctgagcacagacgctctcgcggcGGCACGTGCAGCTATAAAAGCaGcgGTAACcGGAGACgaatgCAACATTtcgaacgcaatcg
tcgtgcagtcaacatctcaggataacctcctccgaaagatcgaagaagttcttcaacaagtcgttcttccatttacctaaacgcaacaaataaacaaaacaaa
cacacaATGGCcACCAAATACGAgaTgTCgAAaACcTAtCAGTACCGCAAG

Capital letters represent nucleotides the *D. melanogaster* reference sequence that are conserved in *D. simulans*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. grimshawi*, *D. virilis* and *D. mojavensis*.

```
aaaatggtttctttttttaggaatttttttaaacatcataaccagacaagaacatctttttcaaatatttctttatttcacccatctgggtacagtaatcccat
acatagatattctgtaaactgcaaccttcacgaaccttaatgttttccaagactcaggttcccactcaccgtcgcagctaacaaaattatccgtagcaagtag
aaaacaccttccacaaaagtccctatactcggactatatacgcctagaaaaacctctaggaacaaatgtacctttctaacatagcagatagtttaggaaaaat
tctttacaagtagaaatacatcgagaaatccataaacacttactctaaaacctgcaagtcgtaaaattgcacgcatgttccgtttggtaaaaccccaaaaatc
acattcgaaaactcacattcccacgactttgaatagaggttctcaaattaaccgtagcaagttgagaacaggcaacaaaaaatcggcaagatactgatttcca
aacctgcgagtcccaaaattgcacacatcttcgccttggaaaaccccatttccgacatcccaaaactcgcattcccacgatccctacagaaattgctaaaact
ttgtctagcaagtagataacagtatgcgtaaattccgtaaaccttaaatttcacgtagttggaaaagccttgaagaattcattcaagcaaggcaccctcaaaa
ttcccgcatgtgtttcatcggtttggCCCACTGTGCcATGTGCATAAGGGCTCggaactcggagctcggactcgggactcgcacacgggctacctcCGATTAT
AACTTATAACACCAACgAGCGAGAAAATCTTGTGGGAAACTTGAGGGCAAAGTGTTTCCCACGATTCGAAtgTCagtggAAAgGAAAACAGAGaagccagcat
gtgcaaccaccgtcgcaccaccaccactgcactatcggcatcTTCATCCCCTggccgagactcacttaacaccgaagcaccagtggccatagcaatagtactg
agaacatatttaaagttttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaatttttttaggtgtaaacgtataacattatg
ttaatttctttttgtttaataaaacgacatttaagtgctttaagggaaacccatttgaagtaattgttaaaaacgtaacgtacttcttaggtatataacaaca
aatttgtatattcgctcgaagaactatagctgctttaaacacatttctcatactatttttaccgcaggtgtgttccaaacttgtttgtggagaaaaacccacc
accgaaaatgcaggcgagaggatcgtgtcccagtgactgagtaaaagtagtttgctccgcaggtggtggttcttcggaggctccgcagctctctcttctctct
ttctccgagcagcgtgcgctgagcacagacgctctcgcggcGGCACGTGCAGCTATAAAAGCaGcgGTAACcGGAGAcgaatgCAaCATTtcgaacgcaatcg
tcgtgcagtcaacatctcaggataacctcctccgaaagatcgaagaagttcttcaacaagtcgttcttccatttacctaaacgcaacaaataaacaaaacaaa
cacacaATGGCcACCAAATACGAgatgTCgAAaACcTAtCAGTACCGCAAG
```

Figure 7: EvoPrinter results: *D. melanogaster* reference (MacDonald and Long 2005).

EvoPrinter alignments were created using *D. melanogaster* as a reference. Seven other *Drosophila* species were added into the alignments according to the relative percentages of identity displayed from the corresponding BLAT alignment. The upstream regulatory binding sites were identified and highlighted accordingly.

Table 5: Summary of results for *D. melanogaster* (MacDonald and Long 2005) EvoPrint.

Species	X - Su(H)	IX - Su(H)	VIII - Su(H)	VII - N-box	VI - Paired Su(H)	V - N-box	IV - Su(H)	III - E-box	II - E-box	I - N-box
D. simulans, D. yakuba	+	lost	+	+	+	lost	lost	+	+	+
Above 2 plus D. erecta	+	lost	lost	+	+	lost	lost	+	+	+
Above 3 plus D. ananassae	lost	lost	lost	+	+	lost	lost	lost	lost	+
Above 4 plus D. pseudoobscura	lost	lost	lost	+	+	lost	lost	lost	lost	+
Above 5 plus D. grimshawi	lost	lost	lost	+	+	lost	lost	lost	lost	+
Above 6 plus D. virilis	lost	lost	lost	+	+	lost	lost	lost	lost	+
Above 7 plus D. mojavensis	lost	lost	lost	+	+	lost	lost	lost	lost	+

As the seven *Drosophila* species were added into the EvoPrint alignment against the *D. melanogaster* reference, some upstream regulatory binding sites did not remain conserved. The sites that were lost were marked with "lost," and sites that remained conserved are marked with a "+".

Capital letters represent nucleotides in the *D. melanogaster* FlyBase reference sequence that are conserved in ***D. melanogaster* and *D. simulans***.

TCAAGCTtCAGGCCTGAAAAAATTAAACTCTTTTTATTTATTTatttttttttttttttgtttttcTTTTTTGCATTTCA
TGCGATTTTATTGCGCATGCGtGGCGCGCAGCAAGCTTCTGTTGGGGCCCTTTGGCTTTTGTTTTTAGCTGCCCGCGTCT
GTCGCGCGTTTGGCGTGCCAAGCATGCGCACACATATTCAATTACACGGCcAGCAACgTCGGtTGCCCAGCAGCAGCAGC
GATTATAGCGATGCATTGACAATAAGCCAACCCCGATGCGTGCCCAAATGGTTGCCCCTACTTAATAAAAAGTGGAAAAA
ACAACAACAAaAGCCCGGAGCTGGCCATGTGCAGCTCGAGAAATTTTCTCACAGTCGCCAGAAAAaTTGTAACGGCCCGT
TGTTCAGCACACTCTCAGCACGCAGTGCActgcacaCACACACTTCCTGCCGCAGCCAGCGTTCCTCAAACCGGCCACGT
ATATTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCA
TGTCCCTCGTCGATCCATCGAACAATCATCCCTTTTTCTGGGaCCTGGGTCCCACGATGATGATGTCCCGACGTGAGCAG
GAATACCATACGCGTATATGTATGCCACTTAGCGATGGACGATCCCAACCGATGCCATCCGAAACGAAACCCAGGGCCCA
AGTGTTCAACAGAAGCTACAATTCAGCATTCATAATTGACAAGAGCaGGAGCGAGGGAGAGGGCGATATGTGTAGGGCCC
ACGATGGACAAGAAAAAATTAAATAAAATAAAGG**GCAGGTG**GTAGTTTTTTCAGCACTACCCGTGACTCGCACTAGCAGC
TCGGTCAGATTTACAGTGGTTCGAATGGAACGAAAATTGGtTGAAAAcAATTtAGGGTcTTGTCACAATCTTTAGATGCT
TTTAGCTAcAAACTACAAAATGgtttctTTTTTTAGaAATTTtTTAAAcATcATAACCTgaCAAGAAcATCTTTTCCaAA
TATTTCTTTATttcaccCATCTGGgTACAGTAATCCCATACATAGAtATTCTGtAAAcTGCAACCTTCACGAACCTTAAT
gTTTTtCAAGACTCAGG**TTCCCAC**tCACCGTCGCAGCTAACAAAATTATCCGTAGCAAGTAgAAAACACCTTtCACAAAA
gTCCCTATAcTCGGACtATATatGCCTAGAAAAACCtCTAGgAACAAATGTAcCTTTCcaACATAGCAGATAGTTTAGtA
AAAATTCtTTACAAGTAGAAAcACATCGAGAAATCCaTAAACACTTACtCTAAAACCTGCAAGTCGTAAAATTGCACGCA
TGTTCCGTTTGGTAAAACCCCAAAAATcACATTCGAAAACTCACA**TTCcCACG**ACTTTGAAtAGAGGTTCTCAAATtAAC
CGTAGCAAGTTGAGAACAGGCAACAAAAAAtCCGTAAGaTACTGaTTTCCAAACCTGCGAGTCCcAAAATTGCACACATC
TTCgCCTTGGAAAACCCCATTTCCGACATCCCAAAACTCGCA**TTCCCACG**ATCCCTACAgAAATTgCTAAAACTTTGTCT
AGCAAGTAGATAACAGtATGCGTAAAtTCCGTAAACCTTAAATTTCACGTAGTTGGAAAAGCCTTGAAGAaTTCATTCAA
GCAAGGCACCCTCAAAATTCCCGCATGTGTTTCATCGGTTTGGCCCACTGTGCCATGTGCATAAGGGCTCGGAACTCGGA
GCTCGGACTCGGGACTCGCACACGGGCTACCTCCGATTATAACTTATAACACCAACGAGCGAGAAAAAT**CTTGTGGGAA**A
CTTGAGGGCAAAGTGT**TTCCCACG**ATTCGAATGTCAGTGGAAAGGAAAACAGAGAAGCAGCATGTGCAACCACCGTCGCA
CCACCACCACTGCACtATCGCATCTTCATCCCCTGCCGAGACTCACTTACACCgAA**CACCAG**tGGCCAtAGCAATAGtAC
TGAGAACATATTTAAAGTTCTAGAtAatgtttatATTTTCAaaatcactactattaagaagtgtgctaatttaatttttt
aggtgtaaacgtataacattatgttaatttctttttgtttaataaaacgacatttaagtgctttaagggaaacccatttg
aagtaattgttaaaaacgtaacgtacttcttaggtatataacaacaaatttgtatattcgctcgAAGAACTATAGCTGCt
TTAAaCACAT**TtCtCAT**ACTATTTTTACCG**CAGGTG**TGtTCCAAACTTGTTTGTGGAGAAAAACCCACCACCGAAAATGC
AGGCGAGAGGATCGTGTCCCAGTGACTGAGTAAAAGTAGTTTGCTCCG**CAGGTG**GTGGTTCTTCGGAGGCTCCGCAGCTC
TCtCTTCTCTCTTTCTCCGAGCAGCGTGCGCTGAGCACGAGACGCTCTCGCGGCGG**CACGTG**CAGCTATAAAAGCAGCGG
TAACCGGAGACGAATGCAACATTTCGAACGCAATCGTCGTGCAGTCAACATCTCAGGATAACCTCCTCCGAAAGATCGAA
GAAGTTCTTCAACAAGTCGTTCTTCCAtTTACCTAAaCGCAACAAATAAACAAAACAAACACACAATGGCCACCAAATAC
GAGATGTCGAAAACCTATCAGTACCGCAAGGTGATGAAGCCCTTGCTGGAGCGgAAGCGTCGTGCCAGGATCAACAAGTG
CCTGGACGAGCTCAAGGATCTGATGGCCGAGTGCGTTGCtCAGACTGGCGATGCCAAATTCGAGAAGGCCGACATCCTGG
AAGTCACCGTGCAGCATCTCCGCAAGCTGAAAGAGTCTAAGAAGCAtGTTCCGGCCAATCCCGAACAGAGTTTCCGTGCT
GGATACATCCGGGCAGCCAACGAAGTGTCCCGCGCCCTGGCCTCCCTTCCCCGAGTGGATGTGGCTTTTGGAACCACACT
GATGACCCATCTGGGCATGCGTCTCAACCAACTGGAGCAGCCCATGGAACAaCCGCAaGCCGTCAACACTCCACTCAGTA
TCGTCTGTGGATCCAGCTCtAGCTCCAGCaCCTACTCCAGTGCCAGCTCCTGCTCCTCCATCAGTCCCGTTTCCAGTGGA
TATGCCAGCGACAACGAGTCTCTGCTCCAGATCAGCAGCCCCGGACAGGTGTGGCGTCCCTGGTAAACAaAGATCtACTT
GGAAGGACTGAATTCCCGACGGG

Capital letters represent nucleotides in the *D. melanogaster* FlyBase reference sequence that are conserved in ***D. melanogaster* and *D. sechellia***.

TCAAGCTtCAGGCCTGAAAAAATTAAACTCTTTTTATTTATTTatttttttttttttttgttTTTCTTTTTTGCATTTCA
TGCGATTTTATTGCGCATGCGTGGCGCGCAGCAAGCTTCTGTTGGgGCCCTTTGGCTTTTGTTTTTAGCTGCCCGCGTCT
GTCGCGCGTTTGGCGTGCCAAGCATGCGCACACATATTCAATTACACGGCcAGCAACgTCGGtTGCCCAGCAGCAGCAGC
GATTATAGCGATGCATTGACAATAAGCCAACCCCGATGCGTGCCCAAATGGTTGCCCCTACTTAATAAAAAGTGGAAAAA
ACAACAACAAaAGCCCGGAGCTGGCCATGTGCAGCTCGAGAAATTTTCTCACAGTCGCCAGAAAAaTTGTAACGGCCCGT
TGTTCAGCACACTCTCAGCACGCAGTGCActgcacaCACACACTTCCTGCCGCAGCCAGCGTTCCTCAAACCGGCCACGT
ATATTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGtCATTTAGTTTGCA
TGTCCCTCGTCGATCCATCGAACAATCATCCCTTTTTCTGGGaCCTGGGTCCCACGATGATGATGTCCCGACGTGAGCAG
GAATACCATACGCGTATATGTATGCCACTTAGCGATGGACGATCCCAACCGATGCCATCCGAAACGAAACCCAGGGCCCA
AGTGTTCAACAGAAGCTACAATTCAGCATTCATAATTGACAAgAGCAGGAGCGAGGGAGAGGGCGATATGTGTAGGGCCC
ACGATGGACAAGAAAAAATTAAATAAAATAAAGGG**CAGGTG**GTAGTTTTTTCAGCACTACCCGTGACTCGCACTAGCAGC
TCGGTCAGATTTACAGTGGTTCGAATGGAACGAAAATTGgtTGAAAAcAaTTtAGGGTcTTGTCAcaATCTtTAgATGCT
TTTAGCTAcAAACTACAAaATGgTTTcTTTTTTTAGaAATTTTTTAaacatcATAACCTGACAAGAACATCTTTTcCAAA
TATTTCTTTatTTCACCCATCTGGGtAcAGTaATCCCATACATAGAtATTCTGTAAAcTGCaACCTTCACGAACCTTAAT
gTTTTtCAAGACTCAGG**TTCCCAC**tCACCGTCgCAGCTAACAAAATTATCCGTAGCAAGTAGAAAACACCTTtCACAAAA
GTCCCTATAcTCGGACtATATAtgCCTaGAAAAACCtCTAGGAACAAATGTAcCTTTCcaaCATAGCAGATAGTTTAGtA
AAAATTCTTTACAAGTAGAAAcACATCGAGAAATCCaTAAACACTTACtCTAAAACCTGcAAGTCGTAAAATTGCACGCA
TGTTCCGTTtGGTAAAACCCCAAAAATcACATtCGAAAACTCACA**TTCcCACG**ACTTTGAAtAGAGGTTCTCAAATTAAC
CGTAGCAAGTTGAGAACAGGCAACAAAAAaTCCGTAAGaTACTGaTTTCCAAACCTGCGAGTCCcAAAATTgCACACATC
TTCgCCTtGGAAAACCCCATTTCCGACATCCCAAAACTCGCA**TTCCCACG**ATCCCTACAgAAATTgCTAAAACTTTGTCT
AGCAAGTAGATAACAGtATGCGTAAAtTCCGTaAACCTTAAAtTTCACGTAGTTGGAAAAGCCTTGAAGAaTTCATTCAA
GCAAGGCACCCTCAAAATTCCCGCATGTGTTTCATCGGTTTGGCCCACTGTGCCATGTGCATAAGGGCTCGGAACTCGGA
GCTCGGACTCGGGACTCGCACACGGGCTACCTCCGATTATAACTTATAACACCAACGAGCGAGAAAAAT**CTTGTGGGAA**A
CTTGAGGGCAAAGTGT**TTCCCACG**ATTCGAATGTCAGTGGAAAGGAAAACAGAGAAGCAGCATGTGCAACCACCGTCGCA
CCACCACCACTGCACtATCGCATCTTCATCCCCTGCCGAGACTCACTTACACCGAA**CACCAG**tGGCCAtAGCAATAGtAC
TGAGAACATATTTaAAGTTCtagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaattttttt
aggtgtaaacgtataacattatgttaatttctttttgtttaataaaacgacatttaagtgctttaagggaaacccatttg
aagtaattgttaaaaacgtaacgtacttcttaggtatataacaacaaatttgtatattcgctcgAAGAACTATAGCTGCt
TTAAaCACAT**TtCTCAT**ACTATTTTTACCG**CAGGTG**TGtTCCAAACTTGTTTGTGGAGAAAAACCCACCACCGAAAATGC
AGGCGAGAGGATCGTGTCCCAGTGACTGAGTAAAAGTAGTTTGCTCCG**CAGGTG**GTGGTTCTTCGGAGGCTCCGCAGCTC
TCtCTTCTCTCTTTCTCCGAGCaGCGTGCGCTGAGCACGAGACGCTCTCGCGGCGG**CACGTG**CAgCTATAAAAGCAGCGG
TAACCGGAGACGAATGCAACATTTCGAACGCAATCGTCgTGCAGTCAACATCTCAGGATAACCTCCTCCGAAAGATCGAA
GAAGTTCTTCAACAAGTCGTTCTTCCatTTACCTAAACGCAACAAATAAACAAAACAAACACACAATGGCCACCAAATAC
GAGATGTCGAAAACCTATCAGTACCGCAAGGTGATGAAGCCCTTGCTGGAGCGGAAGCGTCGTGCCAGGATCAACAAGTG
CCTGGACGAGCTCAAGGATCTGATGGCCGAGTGCGTTGCtCAGACTGGCGATGCCAAATTCGAGAAGGCCGACATCCTGG
AAGTCACCGTgCAGCATCTCCGCAAGCTGAAAGAGTCTAAGAAGCAtGTTCCGgCCAATCCCGAACAGAGTTTCCGTGCT
GGATACATCCGGGCAGCCAACGAAGTGTCCCGCGCCCTGGCCTCCCTTCCcCGAGTGGATGTGGCTTTTGGAACCACACT
GATGACCCATCTGGGCATGCGTCTCAACCAACTGGAGCAGCCCATGGAACAACCGCAaGCCGTCAACACtCCACTCAGTA
TCGTCTGTGGATCCAGCTCtAGCTCCAGCaCCTACTCCAGTGCcAGCTCCTGCTCCTCCATCAGTCCCGTTTCCAGTGGA
TATGCCAGCGACAACGAgTCTCTGCTCCAGATCAGCAGCCCCGGACAGGTGTGGCGTCCCTGGTAAACAaAGATCtACTT
GGAAGGACTGAATTCCCGACGGG

Capital letters represent nucleotides in the *D. melanogaster* FlyBase reference sequence that are conserved in ***D. melanogaster* and *D yakuba***.

tcaagcttcaggcctgaaaaaattaaactctttttatttatttatttttttttttTTTTgTTTTTcTTTTTTGCATTTCA
TGCGATTTTATTGCGCATGCGTGGCGcGCAGCAAGCTTCTGTTGGGGCCCTTTGGCTTTTGTtTTtAGCTGCCCGCGTCT
GTCGCGCGTTTGGCGTGCCAaGCATGCGCACACATATTCAATTACACGGCcAGCAACgTCGgtTGCCCAGCAGCAGCAGC
GATTAtAGCGATGCATTGACAATAAGCCAACCCCgATGCGTGCCCAAATGgTTGCCCCTACTTaATAAAAAGTgGAAAAA
ACAACAAcaaaAGCCCGGAGCTGGCCATGTGCAGcTCGAgAAATTTTCTCACAGTCGCCAGAAAAaTTGTAACGGCcCGT
TGTtCAGCACACTCTCAGCACGCAGTGcactgcaCACACACACTTCCTGCCGCAGCCAgCGTTCCTCAAACCGGCCACGT
ATATTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCA
TGTCCCTCGTCGATCCATCGAACAATCATCCCTTTTtCTGGGaCCtGGGTCCCACGATGATGATGTCCCGACGTGAGCAG
GAATACCATACGCGTATATGTATGCCACTTAGCGATGGaCGATCCCAaccgatgccatccgaaacgAAACCCAGGGcCCA
AGTGTTCAACAGAAGCTACAATTCAGCATTCATAATTGACAAGAGcaGGAGCGAGGGAGAGGgcgaTATGTGTAGGGCCC
ACGATGGACAAGAAAAAATTAAATAAAATAAAGGG**CAGGTG**GTAGTTTTTTCAGCACTACCCGTGACTCGCACtAGCAGC
TCGGTCAGATTTACAGTGGTTCGAATGGAAcgaaAAtTGGTTGAAAAcaatttagggtcttgtcacaatctttagatgct
tttagctacaaactacaaaatggtttctttttttagaaatttttttaaacatcataacctgacaagaacatcttttccaaa
tatttctttatttTCACCCATCTGGgtacagtaatcccatacatagatatTCTGTAAAcTGCAACCTTCACGAACCtTAAT
gTTTTtCAAGACTCAGG**TTCCCAC**tCACCGTcGCAGCTAACAAAATTATCCGTAGCAAGTAGAAAACACCTTtCACAAAA
GTCCCTAtAcTcgGACtATATAtGCCTAGAAAAACCtCtAGGAACAAATgTACCTTTCCaACaTAGCAgATAGTTTAGtA
AAAaTTCTTTACAAGTAGAAAcACAtCGAgAAATCCaTAAACACTTACTCTaAAACCTGCaAGTCGTAAAATTGCACGCA
TgTTCCGTTTGGTAAAACCCCAAaAATCACATTCgAAAAcTCACa**TTCCCACG**ACTTTgAaTAGAGGTtCTcAAATTAAC
CGTAGCAAGTTGAGAACAGGCAACaAAAAATCCGTAAGaTACTGaTTTCcAAACCTGCGAGTCCcAAAATTGCACACATC
tTCGCCtTGGAAAaCCCCATTTCcGACATCcCAaAacTCGCA**TTCCCACG**ATCCCtacagAAATTgCTAAAACTTTGtCT
AGCAAGTAGATAaCAGtATGCGTAAAtTCCGTAAACCTTAAaTTtCACGTAGTTGGAAAAGCCTTGAAgAaTTCATTCAA
GCAAGGCACCcTCaAAATTCCCGCATGtGTTTCATCGGTTTGGCCCACTGTGCCATGTGCATAAGGGCTCGGAACTCGGA
GCTCGGACTCGGGACTCGCACACGGGCTACCTCCGATTATAACTTATAACACCAACGAGCGAGAAAAAT**CTTGTGGGAA**A
CTTGAGGGCAAAGTGT**TTCCCACG**ATTCGAATGTCAGTGGAAAGGAAAACAGAGAAGCAGCATGTGCAACCACCGTCGCA
CCACCACCACTGCACtATCGCATCTTCATCCCCTGCCGAGACTCACTTACACCgaa**caccag**tggccatagcaatagtac
tgagaacatatttaaagttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaattttttt
aggtgtaaacgtataacattatgttaatttctttttgtttaataaaacgacatttaagtgctttaagggaaacccatttg
aagtaattgttaaaaacgtaacgtacttcttaggtatataacaacaaatttgtatattcgctcgaagaactatagctgct
ttaaacacat**ttcTCAT**ACTATTTTtaCCg**CAGGTG**TGTTCCAAACTTGTTTGTGGAGAAAAACCCACCACCGAAAATGC
AGGCGAGAGGATCGTGTCCCAGTGACTGAGTAAAAGTAGTTTGcTCCG**CAGGTG**GTGGTTCTTCGGAGgCTCCGCAGCTC
TCtCTTCTCTCTTTCTCCGAGCAGCGTGCGCTGAGCACGAGACGCTCTCGCGGCGG**CACGTG**CAGCTATAAAAGCAGCGG
TAACCGGAGACGaAtGCAACATTTCgAACGCAATCGTCGtGCAGTCAACATCTCAGGATAACCTcCTCCGAAAGATCGAA
GAAGTTCTTCAACAagtcgttcttccatttacctaaacgcaacaaataAACAAAACAAAcACACAATGGCCACCAAATAC
GAgATGTCgAAaACCTATCAGTACCGCAAGGTGATGAAGCCctTGCTGGAGCGGAAGCGTCGTGCCAGgATCAACAAGTG
CCTGGACGAGCTCAAGGATCTGATGGCCGAGTGCGTtGCtCAGACTGGCGATGCCAAATTCGAGAAGGCCGACATcCTGG
AAGTCACCGTGCAGCATCTCCGCAAGCTgAAAGAGTCTAAGAAGCAtGTtCCgGCCAATCCCGAaCAgAGTTTCCGTGCT
GGATACATCCGGGCAGCCAACGAAGTGTCCCGCGCCCTGGCCTCCCTTCCcCGaGTGGATGTGGCTTTTGGaACCACACT
GATGACCCAtCTGGGCATGCGTCTCAACCAACTGGAGCAGCCCATGGAACAACcgCAaGCCGTCAACACTCCaCTCAGTA
TCGTCTGtGGATCcAGCTCTAGCTCcagcaccTACTCCAGTGCCAGCTCCTGCTCCTCCATCAGTCCCGTTTCCAGTGGA
TATGCCAGCGACAACGAGTCtCTGCTCCAGATCAGCAGCCCCGGaCAGGTGTGGCGtCCCTGGTAAACaaagatctactt
ggaAGGACTGAAtTCCCGACGGG

Capital letters represent nucleotides in the *D. melanogaster* FlyBase reference sequence that are conserved in ***D. melanogaster* and *D. erecta***.

TCAAGCTtCAGGCCTGAAAAAATTaAACTCTTTTTatttatttatttttttttttttTTTGTTTTTcTTTTTTGCATTTCA
TGCGATTTTATTGCGCATGCGTGGCGCGCAGCAAGCTTCTGTTgGGGCCCTTTGGCTTTTGTtTTtAGCTGCCCGCGTCT
GTCGCGCGTTTGGCGTGCCAaGCATGCGCACACATATTCAATTACACGGCcAGCAaCgTCGGtTGCCCAGCAGCAGCAGC
GATTAtAGCGaTGCAtTGACAATAAGCCAACCCCGATGCGTGCCCAAATGgTTGCCCCTaCTTAATAAAAAGTGGAAAAA
AcAACAACAAaAGCCCGGAGCTGGCCATGTGCAGcTCGAGAAATTTTCTCACAGTCGCCAGAAAAaTTGTAACGGCCCGT
TGTtCAGCACACTCTCAGCACGCAGTGCACtgCACACACACACTTCCTGCCGCAGCCAGCGTtCCTCAAACCGGCCACGT
ATATTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCA
TGTCCCTCGTCGATCCATCGAACAATCATCCCTTTTtCTGGGaCCtGGGTCCCACGATGATGATGTCCCGACGTGAGCAG
GAATACCATACGCGTATATGTATGCCACTTAGCGATGGACGAtCCCAACCGATgCCATcCGAAaCGAAAcCCAGGGcCCA
AGTGTTCAACAGAAGCTACAATTCAGCATTCATAATTGACAAGAGCaGGAGCGAGgGAGAGGGCgATATGTGTAGGGCCC
ACGATGGACAAGAAAAAATTAAATAAAATAAAGGG**CAGGTG**GTAGTTTTTTCAGCACTACCCGTGACTCGCACTAGCAGC
TCGGTCAGATTTACAGTGGTTCGAATgGAAcgaAAATtGGTTGAAAAcaatttagggtcttgtcacaatctttagatgct
tttagctacaaactacaaaatggtttcttttttttagaaatttttttaaacatcataacctgacaagaacatcttttccaaa
tatttctttatttCACCCATCTGGgTACAGTaATCCCatacatagatattCTGTAAAcTGCAACCTTCACGAACCTTAAT
gTTTTtCAAGACTCAGG**TTCCCAC**tCACCGTCgCAgCTaACAAAATTATCCGTAGCAAGTAGAAAACACCTTtCACAAAA
GTCCcTATAcTcgGACtATATAtGCCTAGAAAAACCtCTAGGAACAAATGTACCTTTCCaAcatAGcAGATAGTTTagtA
AAAaTTCTTTACAAGTAGaAAcACATCGAgAAATCCATAAACACTTACTCTAAAACCtGCAAGTCGTAAAATTGCACGCA
TgTTCCGTTTGGTAAAACCCCAAaAATCaCATTCgAAAAcTCACA**TTCCCACG**ACTTTgAATAGAGGTTCTcAAATTAAC
CGTAGCAAGTTGAGAACAGGCAAcaAAAAATCCGTAAGaTACTGaTTTCCAAACCTGCGAGTCCCAAAATTGCACACATC
TTCGcCtTGGAAAACCCCATtTCCGACATCCCAaAACTCGCA**TTCcCACG**ATCcCTACAgAAATTgCTAAAACTTTGtCT
AGCAAGTAGATAACAGtATGCGTAAAtTCCGTAAACCTTaAATTTCACgTAGTTGGAAAAGCCTTGAAGAaTTCATTCAA
GcAAGGCACCCTCaAAATTCCCGCATGTGTTTCATCGGtTTGGCCCACTGTGCCATGTGCATAAGGGCTCGGAACTCGGA
GCTCGGACTCGGGACTCGCACACGGGCTACCTCCGATTATAACTTATAACACCAACgAGCGAGAAAAAT**CTTGTGGGAA**A
CTTGAGGGCAAAGTGT**TTCCCACG**ATTCGAATGTCaGTGGAAAGGAAAACAGAGAAGCAGCATGTGCAACCACCGTCGCA
CCACCACCACTGCACtATCGCATCTTCATCCCCTGCCGAGACTCACTTACACCgAA**CACCaG**tGGcCAtAGCAATAGtac
tgagaacatatttaaagttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaattttttt
aggtgtaaacgtataacattatgttaatttctttttgtttaataaaacgacatttaagtgctttaagggaaacccatttg
aagtaattgttaaaaacgtaacgtacttcttaggtatataacaacaaatttgtatattcgctcgaagaactatagctgct
ttaaacacat**ttctcat**actatttttaccg**CAGGTG**TGTTCCAAACTTGTTTGTGGAGAAAAACCCACCACCGAAAATGC
AGGCGAGAGGATCGTGTCCCAGTGACTGAGTAAAAGTAGTTTGCTCCG**CAGGTG**GTGGTTCTTCGGAGgCTccgCAGCTC
TCTCTTCTCTCTTTCTCCgAgCAGCGTGCGCTGAGCACGAGACGCTCTCgCggCGG**CACGTG**CAGCTATAAAAGCAGCGG
TAACCGGAGACGAATGCAACATTTCgAACGCAATCGTCGTGCAGTCAACATCTCAGGATAaCCTCCTcCGAAAGATCGAA
GAAGTTCTTCAACAagtcgttcttccatttacctaaacgcaACAAATAAACAAAACAAACacacAATGGCCACCAAATAC
GAGATGTCgAAAACcTATCAGTACCGCAAGGTGATGAAGCCcTTGCTGGAGCGGAAGCGTCGTGCCAGGATCAACAAGTG
CCTGGACGAGCTCAAGGATCTGATGGCCGAGTGCGTTGCtCAGACTGGCGATGCCAAATTCGAGAAGGCCGACATCCTGG
AAGTCACCGTGCAGCATCTCCGCAAGCTGAAAGAGTCTAAGAAGCAtGTtCCGGCCAATCCCGAaCAgAGTTTCCGTGCT
GGATACATCCGgGCAGCCAACGAAGTGTCCCGCGCCCTGGCCTCCCTTCCCCGaGTGGATGTgGCTTTTGGAACCACACT
GATGACCCATCTGGGCATGCGTCTCAACCAACTGGAGCAGCCCATGGAACAACcGCAaGCcGTCAACACTCCaCTCAGTA
TCGTCTGtGGATCCAGCTctAGCTCCAGCaCCTACTCCAGTGCCAGCTCCTGCTCCTCcATCAGTCCCGTTTCCAGTGGA
TATGCCAGCGAcAACGAGTCtCTGCTCCAGATCAGcAGCCCCGGACAGGTGTGGCGtCCCTGGTAAACaaagatctactt
ggaaGGACTGAATTCCCGACGGG

Capital letters represent nucleotides in the *D. melanogaster* FlyBase reference sequence
that are conserved in ***D. melanogaster*** **and** ***D. persimilis***.

tcaagcttcaggcctgaaaaaattaaactcttttatttatttatttttttttttttttgTTTTTCTTTTttgcatttca
tgcgaTTTTATtgCGCATGCGTggcgcgcagcaagcttctgttggggccctttggcttttgtttttagctgcccGCGTCT
GTCGCGCGTTTGGCGTGCCaaGCATGCGCAcACATATTCAATTACACggccagcaacgtcggttgcccagcagcagcagc
gattatagcgaTGCATTGACAATAAGCCAACCCCGaTGCGTGCCCAAATGGttGCCCCTActtaataaaaagtggaaaaa
acaacaacaaaagcccggagctGGCCATGTGCAgCTCGAgAAATTTTCTCACAgtcgccagaAAAATTGTAACGGCccGT
TGTTCAGCACACTCTCAGCACGCAgtgcactgcacacacacacttcctgccgcagccagcgttccTcAAACCGGCCACGT
ATATTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCA
TGTCCCTCGtcgatccatcgaacaatcatccctttttctgggacCTGGGTCCCACGATGATGATGTCCCGACGTGagcag
gaataccatacgcgtatatgtATGCCACTTAGCGatggacgatcccaaccgatgccatccgaaacgaaacccagGGCCCA
AGTGTTCAACAGAAGCTACAATTCAGCATTCATAATTGACAAgAGCAGGAGcgagggagagggcgatatgtgtaGGGCCC
ACGAtgGACAAGAAAAAATTAAaTAAAATAAAGGG**CAGGTG**GTAGTTTTTTcagcactacccgTGACTCGCACTaGcAGC
tCGGTCAGATTTACAGTGGTtcgaatggaacgaaaattggttgaaaacaatttagggtcttgtcacaatctttagatgct
tttagctacaaactacaaaatggtttcttttttagaaatttttaaacatcataacctgacaagaacatcttttccaaa
tatttctttatttcacccatctgggtacagtaatcccatacatagatattctgtaaactgcaaccttcacgaaccttaat
gtttttcaagactcagg**ttcccac**tcaccgtcgcagctaacaaaattatccgtagcaagtagaaaacacctttcacaaaa
gtccctatactcggactatatatgcctagaaaaacctctaggaacaaatgtacctttccaacatagcagatagtttagta
aaaattctttacaagtagaaacacatcgagaaatccataaacacttactctaaaacctgcaagtcgtaaaattgcacgca
tgttccgtttggtaaaaccccaaaaatcacattcgaaaactcaca**ttcccacg**actttgaatagaggttctcaaattaac
cgtagcaagttgagaacaggcaacaaaaaatccgtaagatactgatttccaaacctgcgagtcccaaaattgcacacatc
ttcgccttggaaaccccatttccgacatcCCAAAACTCGCA**TTCCCACG**atccctacagaaattgctaaaactttgtct
agcaagtagataacagtatgcgtaaattccgtaaaccttaaatttcacgtagttggaaaagccttgaagaattcattcaa
gcaaggcaccctcaAAATTCCCGCATGTGTttcatcGGTTTGgCCCACTGTGCCATGTGCATAAGGGCTCGGAaCTCGGa
GCTCGGActcgggactcgcACACGGgCTACCTCCGATTATAACTTATAACACCAACGAGCGAGAAAAAT**CTTGTGGGAA**A
CTTGAGGGCAAAGTGT**TTCCCACG**ATTCGAATGTCAGtggAAAgGAAAACAGAGaagcaGCATGTGCAACCAccgtcgcA
CCACCACCactgcactatcgCATCTTCATCCCCTGCCgAGACtcacttacaccgaa**caccag**tggccatagcaatagtac
tgagaacatatttaaagttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaattttt
aggtgtaaacgtataacattatgttaatttctttttgtttaataaaacgacatttaagtgctttaagggaaacccatttg
aagtaattgttaaaaacgtaacgtacttcttaggtatataacaacaaatttgtatattcgctcgaagaactatagctgct
ttaaacacat**ttctcat**actatttttaccg**caggtg**tgttccaaACTTGTTTGTgGAGAAAAACCCACCACCGAAAatGC
AGGCgaGAGGATCGTGtcccagtgactgagtaaaagtagtttgctccG**CAGGTG**GTggttcttcggaggctccgcagctc
tctcttctctctttctccgagcagcgtGCGCTGAGCACGAGACGCTCTcGCgGCGG**CACGTG**CAGCTATAAAAGCaGCGG
TAACcGGAGACGAatGCAACATTtcgaacgcaatcgtcgtgcagtcaacatctcaggataacctcctccgaaagatcgaa
gaagttcttcaacaagtcgttcttccatttacctaaacgcaacaaataaacaaaacaaacacACAATGGCCACCAAATAC
GAgaTgTCgAAaACCTAtCAGTACCGCAAGGTGATGAAGCCctTGCTGGAgCGgAaGCGTCGtGCCaGgATCAACAaGTG
CCTGGACGAGCTCAAGGATCTGATGGccgagtgcgttgctcagactggcgatgccaaattcgagaaggccgacatcctgg
aagtcaccgtgcagcatctccgcaagctgaaagagtctaagaagcatgttccggccaatcccgaacagagtttccgtgct
ggatacatccgggcagccaacgaagtgtcccgcgccctggcctcccttccccgagtggatgtggcttttggaaccacact
gatgacccatctgggcatgcgtctcaaccaactggagcagcccatggaacaaccgcaagccgtcaacactccactcagta
tcgtctgtggatccagctctagctccagcacctactccagtgccagctcctgctcctccatcagtcccgtttccagtgga
tatgccagcgacaacgagtcTCTGCTCCAGATCAGCAgccccggacaggtgtggcgtccctggtaaacaaagatctactt
ggaaggactgaattcccgacggg

Capital letters represent nucleotides in the *D. melanogaster* FlyBase reference sequence that are conserved in ***D. melanogaster* and *D. pseudoobscura***.

tcaagcttcaggcctgaaaaaattaaactctttttatttatttatttttttttttttttgtttttctttttttgcatttca
tgcgattttattgcgcatgcgtggcgcgcagcaagcttctgttggggccctttggcttttgtttttagctgcccGCGTCT
GTCGCGCGTTTGGCGTGCCaaGCATGCGCAcACATATTCAATTACACggccagcaacgtcggttgcccagcagcagcagc
gattatagcgaTGCATTGACAAtaagCCAACCCCGaTGCGTGCCCAAATGGttGCCCCTActtaataaaaagtggaaaaa
acaacaacaaaagcccggagctGGCCATGTGCAgCTCGAgAAATTTTCTCACAgtcgccagaAAAATTGTAACGGCccGT
TGTTCAGCACACTCTCAGCACGCAgtgcactgcacacacacacttcctgccgcagccagcgttcctcAAACCGGCCACGT
ATATTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCA
TGTCCCTCGtcgatccatcgaacaatcatccctttttctgggacCTGGGTCCCACGATGATGATGTCCCGACGTGagcag
gaataccatacgcgtatatgtATGCCACTTAGCGatggacgatcccaaccgatgccatccgaaacgaaacccagGGCCCA
AGTGTTCAACAGAAGCTACAATTCAGCATTCATAATTGACAAgAGCAGGAGcgagggagagggcgatatgtgtaGGGCCC
ACGAtgGACAAGAAAAAATTAAaTAAAATAAAGGG**CAGGTG**GTAGTTTTTTcagcactacccgTGACTCGCACTaGcAGC
tCGGTCAGATTTACAGTGGTtcgaatggaacgaaaattggttgaaaacaatttagggtcttgtcacaatctttagatgct
tttagctacaaactacaaaatggtttctttttttagaaattttttaaacatcataacctgacaagaacatcttttccaaa
tatttctttatttcacccatctgggtacagtaatcccatacatagatattctgtaaactgcaaccttcacgaaccttaat
gtttttcaagactcagg**ttcccac**tcaccgtcgcagctaacaaaattatccgtagcaagtagaaaacacctttcacaaaa
gtccctatactcggactatatatgcctagaaaaacctctaggaacaaatgtacctttccaacatagcagatagtttagta
aaaattctttacaagtagaaacacatcgagaaatccataaacacttactctAAAACCTGCAAgtcgtaaaattgcacgca
tgttccgtttggtaaaaccccaaaaatcacattcgaaaactcaca**ttcccacg**actttgaatagaggttctcaaattaac
cgtagcaagttgagaacaggcaacaaaaaatccgtaagatactgatttccaaacctgcgagtcccaaaattgcacacatc
ttcgccttggaaaccccatttccgacatcCCAAAACTCGCA**TTCCCACG**atccctacagaaattgctaaaactttgtct
agcaagtagataacagtatgcgtaaattccgtaaaccttaaatttcacgtagttggaaaagccttgaagaattcattcaa
gcaaggcaccctcaAAATTCCCGCATGTGTttcatcGGTTTGgCCCACTGTGCCATGTGCATAAGGGCTCGGAaCTCGGa
GCTCGGActcgggactcgcACACGGgCTACCTCCGATTATAACTTATAACACCAACGAGCGAGAAAAAT**CTTGTGGGAA**A
CTTGAGGGCAAAGTGT**TTCCCACG**ATTCGAATGTCAGtggAAAgGAAAACAGAGaagcaGCATGTGCAACCAccgtcgcA
CCACCACCactgcactatcgCATCTTCATCCCCTGCCgAGACtcacttacaccgaa**caccag**tggccatagcaatagtac
tgagaacatatttaaagttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaatttttt
aggtgtaaacgtataacattatgttaatttctttttgtttaataaaacgacatttaagtgctttaagggaaacccatttg
aagtaattgttaaaaacgtaacgtacttcttaggtatataacaacaaatttgtatattcgctcgaagaactatagctgct
ttaaacacat**ttctcat**actatttttaccg**caggtg**tgttccaaACTTGTTTGTgGAGAAAAACCCACCACCGAAAatGC
AGGCgaGAGGATCGTGtCCCAGTGACTGAgtaaaagtagtttgctCcG**CAGGTG**GTggttcttcggaggctccgcagctc
tctcttctctctttctccgagcagcgtGCGCTGAGCACGAGACGCTCTcGCgGCGG**CACGTG**CAGCTATAAAAGCaGCGG
TAACcGGAGACGAatGCAACATTtcgaacgcaatcgtcgtgcagtcaacatctcaggataacctcctccgaaagatcgaa
gaagttcttcaacaagtcgttcttccatttacctaaacgcaacaaataaacaaaacaaacacACAATGGCCACCAAATAC
GAgaTgTCgAAaACCTAtCAGTACCGCAAGGTGATGAAGCCctTGCTGGAgCGgAaGCGTCGtGCCaGgATCAACAaGTG
CCTGGACGAGCTCAAGGATCTGATGGcCgAGTGCgttGctCAGActGGcGATGcCAAATTCGAgAaGGCCGACATcCTGG
AaGTCACcgtgcagcatctccgcaagctgaaagagtctaagaagcatgttccggccaatcccgaacagagtttccgtgct
ggatacatccgggcagccaacgaagtgtcccgcgccctggcctcccttccccgagtggatgtggcttttggaaccacact
gatgacccatctgggcatgcgtctcaaccaactggagcagcccatggaacaaccgcaagccgtcaacactccactcagta
tcgtctgtggatccagctctagctccagcacctactccagtgccagctcctgctcctccatcagtcccgtttccagtgga
tatgccagcgacaacgagtctctgctccagatcagcagccccggacaggtgtggcgtccctggtaaacaaagatctactt
ggaaggactgaattcccgacggg

Capital letters represent nucleotides in the *D. melanogaster* FlyBase reference sequence that are conserved in ***D. melanogaster* and *D. ananassae***.

tcaagcttcaggcctgaaaaaattaaactctttttatttatttatttttttttttttttgtttttctttttttgcatttca
tgcgattttattgcgcatgcgtggcgcgcagcaagcttctgttggggccctttggcttttgtttttagctgcccGCGTCt
GTCGCGCGTTTGGCGTGCCaAGCATGCGCACACATATTCAATTACACGGCcAGCAacgtcggttgcccagcagcagcagc
gattatagcgATGCATTGACAATAAGCCAaCCCCGaTGCGTGCCCAAaTGGtTGCCCCTActtaATAAAAAgtggaAAAA
ACAAcaacaaaagcccggAGCTGGCCATGTGCAGCTCGAgAAATTTTCTCACAgTCGCCAgaaaAATTGTAACGGCccGT
TGTtCAGCACACTCTCAGCACGCAgtgcactgcacaCACACACTTCCTGccgcagccagcgttcctcAAACcGGCCACGT
ATATTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCA
TGTCCCTCGtcgATCCATCGAAcaatCATCCCTTTTTCTGGGacctgggTCCCACGATGATGATGTCCCGACGTGagCAG
GAATACcatacgcgtataTGTATGCCACTTAGCGatggacgatcccaaccgatgccatccgaaacgaaacCCAGGGCCCA
AGTGTTCAACAGAAGCTACAATtCAGCATTCATAATTGACAAgAGCAGGAGCGAGGGagagggcgatatgtgtagggccc
acgatgGACAAGAAAAAATTaAATAAAATaAAGGG**CAGGTG**GTaGTTTTTTcAgCACTACCCGTGACTCGcACTaGcAGC
TCGGTCAGATTTACAGTGGttcgaatggaacgaaaattggttgaaaacaatttagggtcttgtcacaatctttagatgct
tttagctacaaactacaaaatggtttctttttttagaaattttttaaacatcataacctgacaagaacatcttttccaaa
tatttctttatttcacccatctgggtacagtaatcccatacatagatattctgtaaactgcaaccttcacgaaccttaat
gtttttcaagactcagg**ttcccac**tcaccgtcgcagctaacaaaatTATCCGTAGCAAGTagaaaacacctttcacaaaa
gtccctatactcggactatatatgcctagaaaaacctctaggaacaaatgtacctttccaacatagcagatagtttagta
aaaattctttacaagtagaaacacatcgagaaatccataaacacttactctaaaacctgcaagtCGTAAAATtgcacgca
tgttccgtttggtaaaaccccaaaaatcacattcgaaaactcaca**ttcccacg**actttgaatagaggttctcaaattaac
cgtagcaagttgagaacaggcaacaaaaaatccgtaagatactgatttccaaacctgcgagtcccaaaattgcacacatc
ttcgccttggaaaaccccatttccgacatcccaaaactcgca**ttcccacg**atccctacagaaattgctaaactttgtct
agcaagtagataacagtatgcgtaaattccgtaaaccttaaattTCACGTAGtTGGAAAagccttgaagaattcattcaa
gcaaggcaccctcaaaattcccgcatgtgtttcatcGGTTTGGCCCACTGTGCCATGTGCATAAGGGCTCGgAACTCGGA
GCTCGGACTCGGGACTCGCACACGGgcTACCTCCGATTATAACTTATAACACCAACGAGCGAGAAAAAT**CTTGTGGGAA**A
CTTGAGGGCAAAGTGT**TTCCCACG**ATTCGAATGTCaGTGGAAAGGAAAACAGAGAAGCAGCATGTGCAACCACCgTCgcA
CCACCACCACTGCACtATCGcATCTTCATCCCCTGCCGAgactcacttacaccgaa**caccag**tggccatagcaatagtac
tgagaacatatttaaagttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaatttttt
aggtgtaaacgtataacattatgttaatttctttttgtttaataaaacgacatttaagtgctttaagggaaacccatttg
aagtaattgttaaaaacgtaacgtacttcttaggtatataacaacaaatttgtatattcgctcgaagaactatagctgct
ttaaacacat**ttctcaT**ACTATTTttACCg**CaggTG**TgTtCCAAACTTGTTTGTggAGAAAAACCcACCACCgAAAATGC
AGGCGAGAGGATCGTGTcccagtgactgagtaaaagtagtttgctccg**caggtg**gtggttcttcggaggctccgcagctc
tctcttctctctttctccgagcagcgtGCGCTGAGCACGAGACGCTCTCGCGGcGG**CACGTG**CAGCTATAAAAGCaGcGG
TAACcGGAGACgaatgCAACATTtcgAACGCAAtCgTCGTGCAGTCAACATCTCAGGAtaacctcctccgaaagatcgaa
gAAGTTCTTCAacaagtcgttcttccatttacctaaacgcaacaaatAAACAAAACAAAcACACAATGGCCACCAAATAC
GAGATGTCgAAaACCTAtCAGTACCGCAAGGTGATGAAGCCCtTGCTGGAGCGgAAGCGtCGTGCCAGgATCAACAAGTG
CCTGGACGAGCTCAAgGATCTGATGGCCGAGTGCGTtGCtCAGACtGGCGATGCCAAaTTCGAGAAgGCCGACATcCTGG
AAGTcACCGTGCAGCAtCTCCGCAAgCTGAAagagtctaagaagcatgttccggccaatcccgaacagagtttccgtgct
GGATACATCCGGGCaGCCAACGAAGTGTCCCGCGCCCTGGCCTCCCTtCCccgAGTGGATGTgGCtTTtGGaACCACaCT
GATGACcCAtCTgGGcATGCGTCTCAACCAaCTGGAGCAGCCCATGGAacaaccgcaagccgtcAACACTCCACTCagtA
TCGTCTGtggaTCCAGCTCtAGCTCcagcacctactccagtgccagctcctgctcctccatcagtcccgttTCCAGTGGa
TATGCCAGCGACAACGAGTCtCTGcTcCAGATCAGCAGCCCCGGACAGGTGTGGCGtCCCTGGTAAAcaaagatctactt
ggaaggactgaattcccgacggg

Capital letters represent nucleotides in the *D. melanogaster* FlyBase reference sequence that are conserved in ***D. melanogaster* and *D. grimshawi***.

tcaagcttcaggcctgaaaaaattaaactctttttatttatttattttttttttttttgtttttctttttttgcatttca
tgcgattttattgcgcatgcgtggcgcgcagcaagcttctgttggggccctttggcttttgtttttagctgcccgcgtct
gtcgcgcgtttggcgtgccaagcatgcgcacacatattcaattacacggccagcaacgtcggttgcccagcagcagcagc
gattatagcgatgcattgacaataagCCAACCCCGATgCGTGCccaaatggttgcccctacttaataaaaagtggaaaaa
acaacaacaaaagcccggagctgGCCATGTGCAGcTcgAgAAATtTTCTCACAgTCgccaGAAAaATTGTAACGGCccGt
TGTTCAGCACACTCTCAGCACGCagtgcactgcacacacaCACTTCCTGCCgcagccagcgttcctCAAACcGGCCACGT
ATATTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCA
TGtCCCTCgtcgatccatcgaacaatcatccctttttctgggacctgggtcccacgatgatgatgtcccgacgtgagcag
gaataccatacgcgtatatgtATGCCACTTAGCGatggacgatcccaaccgatgccatccgaaacgaaacccagGGCCCA
AGTGTTCAACAGAAGCTACAATTCAGCATTCATAATTGACAAgagcaggagcgagggagagggcgatatgtgtagggccc
acgatggacaagaaaaaattaaataAAATAAAgGG**CAGGTG**GTAGTTTTTTcagcactacccgtgactcgcactagcagc
tCGGTCAGATTTACAGTGGttcgaatggaacgaaaattggttgaaaacaatttagggtcttgtcacaatctttagatgct
tttagctacaaactacaaaatggtttctttttttagaaatttttttaaacatcataacctgacaagaacatcttttccaaa
tatttctttatttcacccatctgggtacagtaatcccatacatagatattctgtaaactgcaaccttcacgaaccttaat
gtttttcaagactcagg**ttcccac**tcaccgtcgcagctaacaaaattatccgtagcaagtagaaaacacctttcacaaaa
gtccctatactcggactatatatgcctagaaaaacctctaggaacaaatgtacctttccaacatagcagatagtttagta
aaaattctttacaagtagaaacacatcgagaaatccataaacacttactctaaaacctgcaagtcgtaaaattgcacgca
tgttccgtttggtaaaaccccaaaaatcacattcgaaaactcaca**ttcccacg**actttgaatagaggttctcaaattaac
cgtagcaagttgagaacaggcaacaaaaaatccgtaagatactgatttccaaacctgcgagtcccaaaattgcacacatc
ttcgccttggaaaccccatttccgacatcccaaaactcGCA**TTCCCACG**atccctacagaaattgctaaaactttgtct
agcaagtagataacagtatgcgtaaattccgtaaaccttaaatttcacgtagttggaaaagccttgaagaattcattcaa
gcaaggcaccctcaaaattcccgCATGTGTTtcatcggtttggCCCACTGTGCcATGTGCATAAGGGCTCggAACTCGga
gctcggactcgggactcgcacacgggctacctcCGATTATAACTTATAACACCAACGAGCGAGAAAAAT**CTTGTGGGAA**A
CTTGAGGGCAAAGTGT**TTCCCACG**ATTCGAAtGTCAgTGgAAAgGAAAACAGAGAAGCAGCATGTGCAACcaccgtcgcA
CCACCAccactgcactatcgcatCTTCATCCCCTGCCgAGACtcacttacaccgaa**caccag**tggccatagcaatagtac
tgagaacatatttaaagttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaattttttt
aggtgtaaacgtataacattatgttaatttctttttgtttaataaaacgacatttaagtgctttaagggaaacccatttg
aagtaattgttaaaaacgtaacgtacttcttaggtatataacaacaaatttgtatattcgctcgaagaactatagctgct
ttaaacacat**ttctcat**actatttttaccg**caggtg**tgttccaAACTTGTTTGtgGAGAAAAACCcACCACCgaaaatgc
aggcgagaggatcgtgtcccagTGACTGAGTAAaagtagTTTGCTCcG**CAGGTG**GTGgttcttcggaggctccgcagctc
tctcttctctctttctccgagcaGCGTGCGCTGAGCACGAGACGCtctcgcggCGG**CACGTG**CAGCTATAAAAGCAGCgG
TAACcGGAGACGaatGCAACATTtcgaacgcaatcgtcgtgcagtcaacatctcaggataacctcctccgaaagatcgaa
gaagttcttcaacaagtcgttcttccatttacctaaacgcaacaaataaacaaaacaaacacacaATGGCcACCAAATAC
GAGaTgTCGAAaACcTATCAGTACCGCAAGGTGATGAAGCCctTGCTGGAGCGgAAGCGTCGtGCcaGgATCAACAAGTG
CcTGGAcGAgCTcAAGGATCTGATGGCCGAGTGcgttgctcagactggcgatgccaaattcgagaaggccgacatcctgg
aagtcaccgtgcagcatctccgcaagctgaaagagtctaagaagcatgttccggccaatcccgaacagagtttccgtgct
ggatacatccgggcagccaacgaagtgtcccgcgccctggcctcccttccccgagtggatgtggcttttggaaccacact
gatgacccatctgggcatgcgtctcaaccaactggagcagcccatggaacaaccgcaagccgtcaacactccactcagta
tcgtctgtggatccagctctagctccagcacctactccagtgccagctcctgctcctccatcagtcccgtttccagtgga
tatgccagcgacaacgagtctctgctccagatcagcagccccggacaggtgtggcgtccctggtaaacaaagatctactt
ggaaggactgaattcccgacggg

Capital letters represent nucleotides in the *D. melanogaster* FlyBase reference sequence that are conserved in ***D. melanogaster* and *D. virilis***.

tcaagcttcaggcctgaaaaaattaaactctttttatttatttatttttttttttttttgtttttctttttttgcatttca
tgcgattttattgcgcatgcgtggcgcgcagcaagcttctgttggggccctttggcttttgtttttagctgcccgcGTCT
GTCGCGCGTTtggcgtgcCAAGCATGCgCACACATAttCAATTacacggccagcaacgtcggttgcccagcagcagcagc
gattatagcgaTGCATTGACAATAaGCCAACCCCGaTGcGTgCCCAAATGgtTGCCCCTActtaataaaaagtggaaaaa
acaacaacaaaagcccggagctgGCCATGTGCAGcTcgAgAAATtTTCTCACAgtcgccaGAAAaATTGTAACGGCccGt
TGTtCAGCACACTCTCAGCACGCagtgcactgcacacacaCACTTCCTGCCgcagccagcgttcctCAAACcGGCCACGT
ATATTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCcGTACAGTCATTTAGTTTGCA
TGtccctcgtcgatccatcgaacaatcatccctttttctgggacctgggtcccacgatgatgatgtcccgacgtgagcag
gaataccatacgcgtatatgtATGCCACTTAGCGatggacgatcccaaccgatgccatccgaaacgaaacccagGGCCCA
AGTGTTCAACAGAAGCTACAATTCAGCATTCATAATTGACAAgagcaggagcgagggagagggcgatatgtgtagggccc
acgatggaCAAGAAAaaattaaatAAAATAAAgGG**CAGGTG**GTAGTTTTTTcagcactacccgtgactcgcactagcagc
tCGGTCAGATTTACaGTGGttcgaatggaacgaaaattggttgaaaacaatttagggtcttgtcacaatctttagatgct
tttagctacaaactacaaaatggtttctttttttagaaatttttttaaacatcataacctgacaagaacatcttttccaaa
tatttctttatttcacccatctgggtacagtaatcccatacatagatattctgtaaactgcaaccttcacgaaccttaat
gtttttcaagactcagg**ttcccac**tcaccgtcgcagctaacaaaattatccgtagcaagtagaaaacacctttcacaaaa
gtccctatactcggactatatatgcctagaaaaacctctaggaacaaatgtacctttccaacatagcagatagtttagta
aaaattctttacaagtagaaacacatcgagaaatccataaacacttactctaaaacctgcaagtcgtaaaattgcacgca
tgttccgtttggtaaaaccccaaaaatcacattcgaaaactcaca**ttcccacg**actttgaatagaggttctcaaattaac
cgtagcaagttgagaacaggcaacaaaaaatccgtaagatactgatttccaaacctgcgagtcccaaaattgcacacatc
ttcgccttggaaaaccccatttccgacatcccaaaactcgca**ttcccacg**atccctacagaaattgctaaaactttgtct
agcaagtagataacagtatgcgtaaattccgtaaaccttaaatttcacgtagTTGGAAAAgccttgaaGAATTCATTCAA
gcaaggcaccctCAAAATTcCCGCATGTGtttcatcggtttgGCCCACTGTGCcATGTGCATAAGGGCTCggAACTCGGa
gctcGGACTCGGgactcgcacacgggctacctCCGATTATAACTTATAACACCAACGAGCGAGAAAAAT**CTTGTGGGAA**A
CTTGAGGGCAAAGTGT**TTCCCACG**ATTCGAAtgtcagtggaaaggaaaacagagaagcagcatgtgcaaccaccgtcgca
ccaccaccactgcactatcgcatCTTCATCCCCTGCCgAGACtcacttacaccgaa**caccag**tggccatagcaatagtac
tgagaacatatttaaagttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaattttttt
aggtgtaaacgtataacattatgttaatttctttttgtttaataaaacgacatttaagtgctttaagggaaacccatttg
aagtaattgttaaaaacgtaacgtacttcttaggtatataacaacaaatttgtatattcgctcgaagaactatagctgct
ttaaacacat**ttctcat**actattttttaccg**caggtg**tgttccAAACTTGTTTGtgGAGAAAAACCcACCACCgaaaatgc
aggcgaGAGGATCGTgtCCCAgTGACTGagtaaaAGTAgTTTGCTCcG**CAGGTG**GTGgttcttcggaggctccgCAGCTC
TCTcttctctctttctccgagcaGCGTGCGCTGAGCACGAGACGCtctcgcggCGG**CACGTG**CAGCTATAAAAGCAGCGG
TAACcGGAGACGAatGCAACATTtcgAACGCAAtCGTCGTGcagtcaacatctcaggataacctcctccgaaagatcgaa
gaagttcttcaacaagtcgttcttccatttacctaaacgcaacaaataaacaaaACAAACAcacaATGGCCACCAAATAC
GAGaTgTCgAAaACCTATCAGTACCGCAAGGTGATGAAGCCctTGCTGGAGCGgAAgCGTCGtGCcaGgATCAACAAGTG
CCTGGACGAgCTcAAGGATCTGATGGCCGAGTGCGTtGCtCAgACTgGCGATGCCAAATTCGAGAAGGCCGACATccTGG
AAGTcACcGTGCAgcAtcTcCGCAAGCTGAAagagtctaagaagcatgttccggccaatcccgaacagagtttccgtgct
ggatacatccgggcagccaacgaagtgtcccgcgccctggcctcccttccccgagtggatgtggctTTTGGaACCACACT
GATGACcCATCTgGGCATGCGTctcaaccaacTGGAGCAGCCcatggaacaaccgcaagccgtcaacactccactcagta
tcgtctgtggatccagctctagctccagcacctactccagtgccagctcctgctcctccatcagtcccgtttccagtgga
tatgccagcgacaacgagtctctgctccagatcagcagccccggacaggtgtggcgtccctggtaaacaaagatctactt
ggaaggactgaattcccgacggg

Capital letters represent nucleotides in the *D. melanogaster* FlyBase reference sequence that are conserved in ***D. melanogaster* and *D. mojavensis***.

```
tcaagcttcaggcctgaaaaaattaaactctttttatttatttatttttttttttttttgttttttctttttttgcatttca
tgcgattttattgcgcatgcgtggcgcgcagcaagcttctgttggggccctttggcttttgtttttagctgcccgcGTCT
GTCGCGCGTTtggcgtgCCAAGCATGCGcaCACATATTCAATTACACGGCCAgCAACgtcggttgcccagcagcagcagc
gattatAGCGaTGCATTGACAATAAGCCAACCCCGaTGCGTGCCcAAATGgtTGCCCCTacttaataaaaagtggaaaaa
acaacaacaaaagcccggagctgGCCATGTGCAGcTcgAgAAATtTTCTCACAgtcgccaGAAAAATTGTAACGGCccGt
TGTTCAGCACACTCTCAGCACGcagtgcactgcacacacaCACTTCCTGCCgcagccagcgttcctCAAACcGGCCACGT
ATATTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCaTTTAGTTTGCA
TGtccctcgtcgatccatcgaacaatcatccctttttctgggacctgggtcccacgatgatgatgtcccgacgtgagcag
gaataccatacgcgtatatgtATGCCACTTAGCGatggacgatcccaaccgatgccatccgaaacgaaacccagGGCCCA
AGTGTTCAACAGAAGCTACAATTCAGCATTCATAATTGACAAgagcaggagcgagggagagggcgatatgtgtagggccc
acgatggacaagAAAAAATTAAATAAAAtaaagGGCAGGTGGTAGTTTTTTCagcactacccgtgactcgcactagcagc
tCGGTCAGATTTACAGTGGttcgaatggaacgaaaattggttgaaaacaatttagggtcttgtcacaatctttagatgct
tttagctacaaactacaaaatggtttcttttttttagaaatttttttaaacatcataacctgacaagaacatcttttccaaa
tatttctttatttcacccatctgggtacagtaatcccatacatagatattctgtaaactgcaaccttcacgaaccttaat
gtttttcaagactcaggttcccactcaccgtcgcagctaacaaaattatccgtagcaagtagaaaacacctttcacaaaa
gtccctatactcggactatatatgcctagaaaaacctctaggaacaaatgtacctttccaacatagcagatagtttagta
aaaattctttacaagtagaaacacatcgagaaatccataaacacttactctaaaacctgcaagtcgtaaaattgcacgca
tgttccgtttggtaaaaccccaaaaatcacattcgaaaactcacattcccacgactttgaatagaggttctcaaattaac
cgtagcaagttgagaacaggcaacaaaaaatccgtaagatactgatttccaaacctgcgagtcccaaaattgcacacatc
ttcgccttggaaaccccatttccgacatcccaaaactcgcattcccacgatccctacagaaattgctaaaactttgtct
agcaagtagataacagtatgcgtaaattccgtaaaccttaaatttcacgtagttggaaaagccttgaagaaTTCATTCAA
gcaaggcaccctCAAAATTcCCGCATGTGTTTCatcggtttgGCCCACTGTGCcATGTGCATAAGGGCTCggaactcgga
gctcggactcgggactcgcacacgggctacctcCGATTATAACTTATAACACCAACGAGCGAGAAAAATCTTGTGGGAAA
CTTGAGGGCAAAGTGTTTCCCACGATTCGAAtGTCAgTGgAAAgGAAAACAGAGAAGcagcatgtgcaaccaccgtcgca
ccaccaccactgcactatcgcatcTTCATCCCCTGCCgAGACtcacttacaccgaacaccagtggccatagcaatagtac
tgagaacatatttaaagttctagataatgtttatattttcaaaatcactactattaagaagtgtgctaatttaattttttt
aggtgtaaacgtataacattatgttaatttctttttgtttaataaaacgacatttaagtgctttaagggaaacccatttg
aagtaattgttaaaaacgtaacgtacttcttaggtatataacaacaaatttgtatattcgctcgaagaactatagctgct
ttaaacacatttctcatactattttttaccgcaggtgtgttccaaacttgtttgtggagaaaaacccaccaccgaaaatgc
aggcgagaggatcgtgtcccagtGACTGAGTAAaagtagTTTGCTCcgCAGGTGGTGgttcttcggaggctccgcagctc
tctcttCTCTCTTTcTccgagcaGCGTGCGCTGAGCACGAGACGCtctcgcggCGGCACGTGCAGCTATAAAAGCAGCgG
TAACcGGAGAcGaatGCAaCATTtcgAACGCAATCGTCGTGCAGTcaacatctcaggataacctcctccgaaagatcgaa
gaagttcttcaacaagtcgttcttccatttacctaAACGCAACaaataaacaaaacaaacacacaATGGCCACCAAATAC
GAgatgTCgAAaACcTATCAGTACCGCAAGGTGATGAAGCCCtTGCTGGAgCGgAAGCGTCGTGCcaGgATCAACAAGTG
CCTGGAcGAgCTCAAGGATCTGATGGCCGAGTGcGTtGCtCAGACTGGcGATGCCAAATTCGAgAAGGCCGACATCcTGG
AaGTCACCGTgcagcatctccgcaagctgaaagagtctaagaagcatgttccggccaatcccgaacagAGTTTCCGTGCt
GGaTACATcCGgGCaGCCAAcGAaGTgTCCCGcGCcCTGGCctCcCTtCCCcgAGTgGATGTgGCtTTtGGaACCACaCT
GATGACCCAtcTGGGCATGCGtCTcAAcCAaCTGGAGCAGCCCATGGAacaaccgcaagccgtcaacactccactcagta
tcgtctgtggatccagctctagctccagcacCTACTCCAGtgccagctcctgctcctccatCAGTCCCGTtTCCAGtGGa
TAtGCCAGCGACAacgagTCTCTGCTcCAGATCAGCAgccccggacaggtgTGGCGTCCCTGGTAAAcaaagatctactt
ggaaggactgaattcccgacggg
```

Figure 8: EvoPrinter results from one on one comparison with *D. melanogaster* (FlyBase).

EvoPrinter alignments were created using *D. melanogaster* as a reference. Ten other *Drosophila* species were aligned directly with the reference sequence. The upstream regulatory binding sites were identified and highlighted accordingly.

Table 6: Summary of results for *D. melanogaster* (FlyBase) EvoPrint alignments.

	XI - E-box	X - Su(H)	IX - Su(H)	VIII - Su(H)	VII - N-box	VI - Paired Su(H)	V - N-box	IV - Su(H)	III - E-box	II - E-box	I - N-box
D. melanogaster	+	+	+	+	+	+	+	+	+	+	+
D. simulans	+	+	not present	+	+	+	+	not present	+	+	+
D. sechellia	+	+	not present	+	+	+	+	not present	+	+	+
D. yakuba	+	+	+	+	+	+	not present	not present	+	+	+
D. erecta	+	+	+	not present	+	+	not present	not present	+	+	+
D. persimilis	+	not present	not present	+	+	+	not present	not present	not present	+	+
D. pseudoobscura	+	not present	not present	+	+	+	not present	not present	not present	+	+
D. ananassae	+	not present	not present	not present	+	+	not present	not present	not present	not present	+
D. grimshawi	+	not present	not present	+	+	+	not present	not present	not present	+	+
D. virilis	+	not present	not present	not present	+	+	not present	not present	not present	+	+
D. mojavensis	+	not present	not present	not present	+	+	not present	not present	not present	+	+

As the ten *Drosophila* species were aligned directly against the *D. melanogaster* reference, some upstream regulatory binding sites did not remain conserved. The sites that were lost were marked with "not present," and sites that remained conserved are marked with a "+".

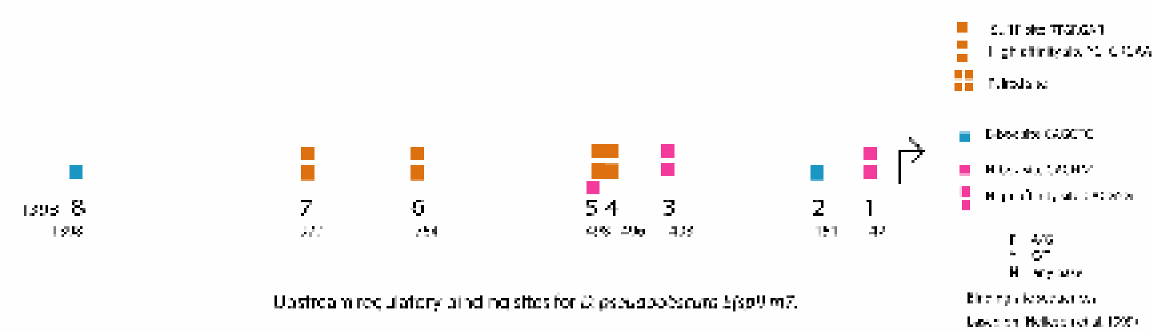


Figure 9: ***D. pseudoobscura E(spl) m7*** **upstream regulatory map.**
Following analysis of *D. pseudoobscura* EvoPrint alignments, a map illustrating relative locations and types of binding sites in the upstream regulatory region of *D. pseudoobscura E(spl) m7* was created.

Capital letters represent nucleotides or amino acids in the *D. Pseudoobscura* reference sequence that are conserved in *D. melanogaster* and *D. simulans*.

tacgttgcagcctgtctctgtctctatctctgtctttggcccagacttcctgtcaagagttttAAACCGGCCACGTATA
TTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCATGT
CCCTCGacccatccgtatcgtatccctgtacctatacatatccctcaCTGGGTCCCACGATGATGATGTCCCGACGTGcc
cgggcATGCCACTTAGCGttcccttctctgttgcccctggccctgccttGGCCCAAGTGTTCAACAGAAGCTACAATTCc
AGCATTCATAATTGACAAaacagacctgcaggagggaaggagacggatagcgggagcgggacagagaacctattGGGCCC
ACGActGACAAGaaAAAAAATTAAgTAAAATAAAGGG**CAGGTG**GTAGTTTTTTttcaccactctccaTGACTCGCACTcG
gAGCaCGGTCAGATTTACAGTGGTacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattat
aaatttagtactattaaatatgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgatacca
tcctgaccaataaaaaaacattcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttc
cacccaaagaaaacctgcaatgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaat
cagtgtccctggcaaattacaacgcatcaaccttacaagacccaccttatttCCAAAACTCGCA**TTCCCACG**caccgttc
cagagtccggaaattaaccaaagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtccc
aaacaacacgtacatagatggaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattg
caccgcatacaccttcccagactcacctta**ttcccacg**caccgtcacagagtccggaaatcaagtagaaatccgtattcg
caggtccaaaaaaacacgtacatacatcggtaggtggactgccttcattcagtcggtcgcagagattcggAAATTCCCGC
ATGTGTgcaatGGTTTGaCCCACTGTGCCATGTGCATAAGGGCTCGGAcCTCGGggactcggatcagacaCGGACTCGGG
gtTCGttCgCtccCTACCTCCGATTATAACTTATAACACCAACGAGCGAGAAAAAT**CTTGTGGGAA**ACTTGAGGGCAAAG
TGT**TTCCCACG**ATTCGAATGTCAGgctAAAaGAAAACAGAGgagagcacacacaagaaaacgcaca**cacgtg**catcggcc
GCATGTGCAACCAACCACCACCgccaagagcagcagcagcaagcagcaacaacaacagctatccaccaacacaggaatgg
aggagacacaccatcacatCATCTTCATCCCCTGCCaAGACgACTTGTTTGTcGAGAAAAACCCACCACCGAAAGCAGGC
GAGGATCGTGgctgtcggcccgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgaCCCAGTGACTG
ACgG**CAGGTG**GTcgcaatggtggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcgctctctctgcgcGC
GCT**GAGCACGAG**ACGaCTCTGCtGCGG**CACGTG**CAGCTATAAAAGCgGCGGTAACtGGAGACGAgaGCAACATTctcaac

Capital letters represent nucleotides or amino acids in the *D. Pseudoobscura* reference sequence that are conserved in *D. melanogaster*, *D. simulans* and *D. yakuba*.

tacgttgcagcctgtctctgtctctatctctgtctttggcccagacttcctgtcaagagttttAAACCGGCCACGTATA
TTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCATGT
CCCTCGacccatccgtatcgtatccctgtacctatacatatccctcactGGGTCCCACGATGATGATGTCCCGACGTGcc
cgggcATGCCACTTAGCGttcccttctctgttgcccctggccctgccttggcCCAAGTGTTCAACAGAAGCTACAATTCc
AGCATTCATAATTGACAAaacagacctgcaggagggaaggagacggatagcgggagcgggacagagaacctattGGGCCC
ACGActGACAAGaaAAAAaaTTAAgTAAAATAAAGGG**CAGGTG**GTAGTTTTTTttcaccactctccaTGACTCGCACtcG
gAGCaCGGTCAGATTTACAGTGGTacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattat
aaatttagtactattaaatatgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgatacca
tcctgaccaataaaaaaacattcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttc
cacccaaagaaaacctgcaatgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaat
cagtgtccctggcaaattacaacgcatcaaccttacaagacccaccttatttccaaaactcgca**ttcccacg**caccgttc
cagagtccggaaattaaccaaagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtccc
aaacaacacgtacatagatggaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattg
caccgcatacaccttcccagactcacctta**ttcccacg**caccgtcacagagtccggaaatcaagtagaaatccgtattcg
caggtccaaaaaaacacgtacatacatcggtaggtggactgccttcattcagtcggtcgcagagattcggAAATTCCCGC
ATGtgtgcaatGGTTTGaCCCACTGTGCCATGTGCATAAGGGCTCGGAcctcggggactcggatcagacaCGGACTCGGG
gtTCGttCgCtccCTACCTCCGATTATAACTTATAACACCAACGAGCGAGAAAAAT**CTTGTGGGAA**ACTTGAGGGCAAAG
TGT**TTCCCACG**ATTCGAATGTCAGgctAAAaGAAAACAGAGgagagcacacacaagaaaacgcaca**cacgtg**catcggcc
GCATGTGCAACCAACCACCACCgccaagagcagcagcagcaagcagcaacaacaacagctatccaccaacacaggaatgg
aggagacacaccatcacatCATCTTCATCCCCTGCCaAGACgACTTGTTTGTcGAGAAAAACCCACCACCGAAAGCAGGC
GAGGATCGTGgctgtcggcccgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgaCCCAGTGACTG
ACgG**CAGGTG**GTcgcaatggtggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcgctctctctgcgcGC
GCT**GAGCACGAG**ACGaCTCTGCtGCGG**CACGTG**CAGCTATAAAAGCgGCGGTAACtGGAGACGagaGCAACATTctcaac

Capital letters represent nucleotides or amino acids in the *D. Pseudoobscura* reference sequence that are conserved in *D. melanogaster*, *D. simulans*, *D. yakuba* and *D. erecta*.

```
tacgttgcagcctgtctctgtctctatctctgtctttggcccagacttcctgtcaagagtttttAAACCGGCCACGTATA
TTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCATGT
CCCTCGacccatccgtatcgtatccctgtacctatacatatccctcactGGGTCCCACGATGATGATGTCCCGACGTGcc
cgggcATGCCACTTAGCGttcccttctctgttgcccctggccctgccttggcCCAAGTGTTCAACAGAAGCTACAATTcc
AGCATTCATAATTGACAAaacagacctgcaggagggaaggagacggatagcgggagcgggacagagaacctattGGGCCC
ACGActGACAAGaaAAAAaaTTAAgTAAAATAAAGGGCAGGTGGTAGTTTTTTttcaccactctccaTGACTCGCACtcG
gAGCaCGGTCAGATTTACAGTGGTacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattat
aaatttagtactattaaatatgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgatacca
tcctgaccaataaaaaaacattcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttc
cacccaaagaaaacctgcaatgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaat
cagtgtccctggcaaattacaacgcatcaaccttacaagacccaccttatttccaaaactcgcattcccacgcaccgttc
cagagtccggaaattaaccaaagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtccc
aaacaacacgtacatagatggaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattg
caccgcatacaccttcccagactcaccttattcccacgcaccgtcacagagtccggaaatcaagtagaaatccgtattcg
caggtccaaaaaaacacgtacatacatcggtaggtggactgccttcattcagtcggtcgcagagattcggAAATTCCCGC
ATGtgtgcaatggtttgaCCCACTGTGCCATGTGCATAAGGGCTCGGAcctcggggactcggatcagacaCGGACTCGGG
gtTCGttCgCtccCTACCTCCGATTATAACTTATAACACCAACgAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAG
TGTTTCCCACGATTCGAATGTCaGgctAAAaGAAAACAGAGgagagcacacacaagaaaacgcacacacgtgcatcggcc
GCATGTGCAACCAACCACCACCgccaagagcagcagcagcaagcagcaacaacaacagctatccaccaacacaggaatgg
aggagacacaccatcacatCATCTTCATCCCCTGCCaAGACgACTTGTTTGTcGAGAAAAACCCACCACCGAAAGCAGGC
GAGGATCGTGgctgtcggcccgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgaCCCAGTGACTG
ACgGCAGGTGGTcgcaatggtggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcgctctctctgcgcGC
GCTGAGCACGAGACGaCTCTgctgCGGCACGTGCAGCTATAAAAGCgGCGGTAACtGGAGACGagaGCAACATTctcaac
```

Capital letters represent nucleotides or amino acids in the *D. Pseudoobscura* reference sequence that are conserved in *D. melanogaster*, *D. simulans*, *D. yakuba*, *D. erecta* and *D. grimshawi*.

```
tacgttgcagcctgtctctgtctctatctctgtctttggcccagacttcctgtcaagagtttttAAACcGGCCACGTATA
TTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCATGt
CCCTCgacccatccgtatcgtatccctgtacctatacatatccctcactgGGTCCCACGAtgatgatgtcccgaCGTGcc
cgggcATGCCACTTAGCGttcccttctctgttgcccctggccctgccttggcCCAAGTGTTCAACAGAAGCTACAATTcc
AGCATTCATAATTGACAAaacagacctgcaggagggaaggagacggatagcgggagcgggacagagaacctattgggccc
acgactgACAAgaaAAAAaattaagtaAAATAAAgGGCAGGTGGTAGTTTTTTttcaccactctccaTGACTCGcactcg
gaGCaCGGTCAGATTTACAGTGGtacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattat
aaatttagtactattaaatatgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgatacca
tcctgaccaataaaaaaacattcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttc
cacccaaagaaaacctgcaatgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaat
cagtgtccctggcaaattacaacgcatcaaccttacaagacccaccttatttccaaaactcgcattcccacgcaccgttc
cagagtccggaaattaaccaaagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtccc
aaacaacacgtacatagatggaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattg
caccgcatacaccttcccagactcaccttattcccacgcaccgtcacagagtccggaaatcaagtagaaatccgtattcg
caggtccaaaaaaacacgtacatacatcggtaggtggactgccttcattcagtcggtcgcagagattcggaaattcccgC
ATGtgtgcaatggtttgaCCCACTGTGCcATGTGCATAAGGGCTCggacctcggggactcggatcagacacggactcggg
gttcgttcgctccctacctcCGATTATAACTTATAACACCAACgAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAG
TGTTTCCCACGATTCGAAtGTCaggctAAAaGAAAACAGAGgagagcacacacaagaaaacgcacacacgtgcatcggcc
gcatgtgcaaccaaccaccaccgccaagagcagcagcagcaagcagcaacaacaacagctatccaccaacacaggaatgg
aggagacacaccatcacatCAtCTTCATCCCCTGCCaAGACgACTTGTTTGtcGAGAAAAACCcACCACCgaaaGCAGgC
GAGGATCGtggctgtcggcccgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgacccagtgactg
acggcaggtggtcgcaatggtggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcgctctctctgcgcGC
GCTGAGCACGAGACGactctgctgCGGCACGTGCAGCTATAAAAGCgGCgGTAACtGGAGACGagaGCAACATTctcaac
```

Capital letters represent nucleotides or amino acids in the *D. Pseudoobscura* reference sequence that are conserved in *D. melanogaster*, *D. simulans*, *D. yakuba*, *D. erecta*, *D. grimshawi* and *D. virilis*.

```
tacgttgcagcctgtctctgtctctatctctgtctttggcccagacttcctgtcaagagtttttAAACcGGCCACGTATA
TTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCcGTACAGTCATTTAGTTTGCATGt
ccctcgacccatccgtatcgtatccctgtacctatacatatccctcactgGGTCCCAcgatgatgatgtcccgaCGTGcc
cgggcATGCCACTTAGCGttcccttctctgttgcccctggccctgccttggcCCAAGTGTTCAACAGAAGCTACAATTcc
AGCATTCATAATTGACAAaacagacctgcaggagggaaggagacggatagcgggagcgggacagagaacctattgggccc
acgactgaCAAgaaAaaaaattaagtaAAATAAAgGGCAGGTGGTAGTTTTTTttcaccactctccaTGACTCGcactcg
gaGCaCGGTCAGATTTACaGTGGtacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattat
aaatttagtactattaaatatgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgatacca
tcctgaccaataaaaaaacattcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttc
cacccaaagaaaacctgcaatgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaat
cagtgtccctggcaaattacaacgcatcaaccttacaagacccaccttatttccaaaactcgcattcccacgcaccgttc
cagagtccggaaattaaccaaagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtccc
aaacaacacgtacatagatggaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattg
caccgcatacaccttcccagactcaccttattcccacgcaccgtcacagagtccggaaatcaagtagaaatccgtattcg
caggtccaaaaaaacacgtacatacatcggtaggtggactgccttcattcagtcggtcgcagagattcggaaattcccgC
ATGtgtgcaatggtttgaCCCACTGTGCcATGTGCATAAGGGCTCggacctcggggactcggatcagacacggactcggg
gttcgttcgctccctacctcCGATTATAACTTATAACACCAACgAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAG
TGTTTCCCACGATTCGAAtgTCaggctAAAaGAAAACAGAGgagagcacacacaagaaaacgcacacacgtgcatcggcc
gcatgtgcaaccaaccaccaccgccaagagcagcagcagcaagcagcaacaacaacagctatccaccaacacaggaatgg
aggagacacaccatcacatCAtCTTCATCCCCTGCCaAGACgACTTGTTTGtcGAGAAAAACCcACCACCgaaagcaggC
GAGGATCGtggctgtcggcccgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgacccagtgactg
acggcaggtggtcgcaatggtggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcgctctctctgcgcGC
GCTGAGCACGAGACGactctgctgCGGCACGTGCAGCTATAAAAGCgGCgGTAACtGGAGACGagaGCAACATTctcaac
```

Capital letters represent nucleotides or amino acids in the *D. Pseudoobscura* reference sequence that are conserved in *D. melanogaster, D. simulans, D. yakuba, D. erecta, D. grimshawi, D. virilis* and *D. ananassae*.

```
tacgttgcagcctgtctctgtctctatctctgtctttggcccagacttcctgtcaagagtttttAAACcGGCCACGTATA
TTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCcGTACAGTCATTTAGTTTGCATGt
ccctcgacccatccgtatcgtatccctgtacctatacatatccctcactgGGTCCCAcgatgatgatgtcccgaCGTGcc
cgggcATGCCACTTAGCGttcccttctctgttgcccctggccctgccttggcCCAAGTGTTCAACAGAAGCTACAAtTcc
AGCATTCATAATTGACAAaacagacctgcaggagggaaggagacggatagcgggagcgggacagagaacctattgggccc
acgactgaCAAgaaAaaaaattaagtaAAATaAAgGGCAGGTGGTaGTTTTTTttcaccactctccaTGACTCGcactcg
gaGCaCGGTCAGATTTACaGTGGtacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattat
aaatttagtactattaaatatgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgatacca
tcctgaccaataaaaaaacattcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttc
cacccaaagaaaacctgcaatgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaat
cagtgtccctggcaaattacaacgcatcaaccttacaagacccaccttatttccaaaactcgcattcccacgcaccgttc
cagagtccggaaattaaccaaagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtccc
aaacaacacgtacatagatggaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattg
caccgcatacaccttcccagactcaccttattcccacgcaccgtcacagagtccggaaatcaagtagaaatccgtattcg
caggtccaaaaaaacacgtacatacatcggtaggtggactgccttcattcagtcggtcgcagagattcggaaattcccgc
atgtgtgcaatggtttgaCCCACTGTGCcATGTGCATAAGGGCTCggacctcggggactcggatcagacacggactcggg
gttcgttcgctccctacctcCGATTATAACTTATAACACCAACgAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAG
TGTTTCCCACGATTCGAAtgTCaggctAAAaGAAAACAGAGgagagcacacacaagaaaacgcacacacgtgcatcggcc
gcatgtgcaaccaaccaccaccgccaagagcagcagcagcaagcagcaacaacaacagctatccaccaacacaggaatgg
aggagacacaccatcacatcAtCTTCATCCCCTGCCaagacgACTTGTTTGtcgAGAAAAACCcACCACCgaaagcaggC
GAGGATCGtggctgtcggcccgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgacccagtgactg
acggcaggtggtcgcaatggtggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcgctctctctgcgcGC
GCTGAGCACGAGACGactctgctgcGGCACGTGCAGCTATAAAAGCgGcgGTAACtGGAGACgagagCAACATTctcaac
```

Capital letters represent nucleotides or amino acids in the *D. Pseudoobscura* reference sequence that are conserved in *D. melanogaster, D. simulans, D. yakuba, D. erecta, D. grimshawi, D. virilis, D. ananassae* and *D. mojavensis*.

tacgttgcagcctgtctctgtctctatctctgtctttggcccagacttcctgtcaagagtttttAAACcGGCCACGTATA
TTGTTTATATAGTTAACGAATGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCcGTACAGTCaTTTAGTTTGCATGt
ccctcgacccatccgtatcgtatccctgtacctatacatatccctcactgGGTCCCAcgatgatgatgtcccgaCGTGcc
cgggcATGCCACTTAGCGttcccttctctgttgcccctggccctgccttggcCCAAGTGTTCAACAGAAGCTACAAtTcc
AGCATTCATAATTGACAAaacagacctgcaggagggaaggagacggatagcgggagcgggacagagaacctattgggccc
acgactgaCAAgaaAaaaaattaagtaAAAtaaagGG**CAGGTG**GTaGTTTTTTttcaccactctccaTGACTcgcactcg
gaGCaCGGTCAGATTTACaGTGGtacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattat
aaatttagtactattaaatatgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgatacca
tcctgaccaataaaaaaacattcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttc
cacccaaagaaaacctgcaatgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaat
cagtgtccctggcaaattacaacgcatcaaccttacaagacccaccttatttccaaaactcgca**ttcccacg**caccgttc
cagagtccggaaattaaccaaagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtccc
aaacaacacgtacatagatggaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattg
caccgcatacaccttcccagactcaccttaa**ttcccacg**caccgtcacagagtccggaaatcaagtagaaatccgtattcg
caggtccaaaaaaacacgtacatacatcggtaggtggactgccttcattcagtcggtcgcagagattcggaaattcccgc
atgtgtgcaatggtttgaCCCACTGTGCcATGTGCATAAGGGCTCggacctcggggactcggatcagacacggactcggg
gttcgttcgctccctacctcCGATTATAACTTATAACACCAACgAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAG
TGTTTCCCACGATTCGAAtgTCaggctAAAaGAAAACAGAGgagagcacacacaagaaaacgcaca**cacgtg**catcggcc
gcatgtgcaaccaaccaccaccgccaagagcagcagcagcaagcagcaacaacaacagctatccaccaacacaggaatgg
aggagacacaccatcacatcatcTTCATCCCCTGCCaagacgACTTGTTTGtcgAGAAAAACCcACCACCgaaagcaggC
GAGGATCGtggctgtcggcccgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgacccagtgactg
acgg**caggtg**gtcgcaatggtggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcgctctctctgcgcGC
GCT**GAGCACGAG**ACGactctgctgcGG**CACGTG**CAGCTATAAAAGCgGcgGTAACtGGAGAcgagagCAaCATTctcaac

Figure 10: EvoPrinter results: *D. pseudoobscura* reference (Couronne et al. 2003). EvoPrinter alignments were created using *D. pseudoobscura* as a reference. Seven other *Drosophila* species were added into the alignments according to the relative percentages of identity displayed from the corresponding BLAT alignment. The upstream regulatory binding sites were identified and highlighted accordingly.

Table 7: Summary of results for *D. pseudoobscura* (Couronne et al. 2003) EvoPrints.

Species	8 - E-box	7 - Su(H)	6 - Su(H)	5 - N-box	4 - Paired Su(H)	3 - N-box	2 - E-box	1 - N-box
D. melanogaster D. simulans	+	+	lost	+	+	lost	+	+
Above 2 plus D. yakuba	+	lost	lost	+	+	lost	+	+
Above 3 plus D. erecta	+	lost	lost	+	+	lost	+	+
Above 4 plus D. grimshawi	+	lost	lost	+	+	lost	lost	+
Above 5 plus D. virilis	+	lost	lost	+	+	lost	lost	+
Above 6 plus D. ananassae	+	lost	lost	+	+	lost	lost	+
Above 7 plus D. mojavensis	+	lost	lost	+	+	lost	lost	+

As the seven *Drosophila* species were added into the EvoPrint alignment against the *D. pseudoobscura* reference, some upstream regulatory binding sites did not remain conserved. The sites that were lost were marked with "lost," and sites that remained conserved are marked with a "+".

Capital letters represent nucleotides or amino acids in the *D. pseudoobscura* reference sequence that are conserved in ***D. pseudoobscura*** and ***D.melanogaster***.

TGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCATGTCCCTCGacccatccgtatcg
tatccctgtacctatacatatccctcaCTGGGTCCCACGATGATGATGTCCCGACGTGcccgggcATGCCACTTAGCGtt
cccttctctgttgcccctggccctgccttGGCCCAAGTGTTCAACAGAAGCTACAATTCcAGCATTCATAATTGACAAaA
cagacctGCAGGAGggaaggagacggatagcgggagcgggacagagaacctattGGGCCCACGActGACAAGaaAAAAAA
TTAAgTAAAATAAAGGG**CAGGTG**GTAGTTTTTTttcaccactctccaTGACTCGCACTcGgAGCaCGGTCAGATTTACAG
TGGTacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattataaatttagtactattaaata
tgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgataccatcctgaccaataaaaaaaca
ttcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttccacccaaagaaaacctgcaa
tgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaatcagtgtccctggcaaattac
aacgcatcaaccttacaagacccaccttatttCCAAAACTCGCA**TTCCCACG**caccgttccagagtccggaaattaacca
aagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtcccaaacaacacgtacatagatg
gaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattgcaccgcatacaccttcccag
actcacctta**ttcccacg**caccgtcacagagtccggaaatcaagtagaaatccgtattcgcaggtccaaaaaaacacgta
catacatcggtaggtggactgccttcattcagtcggtcgcagagattcggAAATTCCCGCATGTGTgcaatGGTTTGaCC
CACTGTGCCATGTGCATAAGGGCTCGGAcCTCGGggactcggatcagacaCGGACTCGGGgtTCGttCgCtccCTACCTC
CGATTATAACTTATAACACCAACGAGCGAGAAAAATCT**TGTGGAAACTTGAGGGCAAAGTGTTTCCCACG**ATTCGAATG
TCAGgctAAAaGAAAACAGAGgagagcacacacaagaaaacgcaca**cacgtg**catcggccGCATGTGCAACCAACCACCA
CCgccaagagcagcagcagcaagcagcaacaacaacagctatccaccaacacaggaatggaggagacacaccatcacatC
ATCTTCATCCCCTGCCaAGACgACTTGTTTGTcGAGAAAAACCCACCACCGAAAGCAGGCGAGGATCGTGgctgtcggcc
cgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgaCCCAGTGACTGACgG**CAGGTG**GTcgcaatgg
tggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcGCTCTCTCTGCGcGCGCTGAGCACGAGACGaCTCT
GCtGCGG**CACGTG**CAGCTATAAAAGCgGCGGTAACtGGAGACGAgaGCAACATTctcaacaaattcgacgtgcagtgaac
acacacaagacaaccggatcccagccagatctcgctcaacttcttcacaattttacgtcaagaaaacacacatctaagct
ACAATGGCCACCAAATACGAagTtTCcAAgACCTAcCAGTACCGCAAGGTGATGAAGCCtcTGCTGGAaCGcAgGCGTCG
gGCCcGcATCAACAtGTGCCTGGACGAGCTCAAGGATCTGATGGtccagtgcacggtccagagcggtgatggcaaattcg
acagggccgacatactggaggtcactgtggaccatttgcgcaaactgaagcaggccaggatcgaggccaccgctgccgca
gcaaaagccaccaccaacactaccccagagcagagtttccgcgatggcttcatcagggccgccgacgaagtctcccgggc
actggcttccctacccaatgtggatgtcgtctttggcacccatctgatgacacacctcggcttgcgcctcaaccaattgg
aaacgccaactgcagcaccaaagccaatgaatgccccgctctccatcatctgcggcagcggcggcagcagcagcagcagc
agcaactgcagctccaacagccgggaggcctgcagtcccgtgtccagcggctactgcagcgacagcgattgcagtgtctc
ctccactcaagcggcccagaaTCTGCTCCAGATCAGCActggtcaggtttggcgcccctggtgaatctatagattggatt

Capital letters represent nucleotides or amino acids in the *D. pseudoobscura* reference sequence that are conserved in ***D. pseudoobscura*** and ***D. simulans***.

TGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCATGTCCCTCGacccatccgtatcg
tatccctgtacctatacatatccctcaCTGGGTCCCACGATGATGATGTCCCGACGTGcccgggcATGCCACTTAGCGtt
cccttctctgttgcccctggccctgccttGGCCCAAGTGTTCAACAGAAGCTACAATTCcAGCATTCATAATTGACAAaa
cagacctgcaggagggaaggagacggatAGCGGGAGCGggacagagaacctattGGGCCCACGActGACAAGaaAAAAAA
TTAAgTAAAATAAAGGG**CAGGTG**GTAGTTTTTTttcaccactctccaTGACTCGCACTcGgAGCaCGGTCAGATTTACAG
TGGTacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattataaatttagtactattaaata
tgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgataccatcctgaccaataaaaaaaca
ttcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttccacccaaagaaaacctgcaa
tgtcaaacctagCACAAAACTCCCTAcagttcacacactacgcctagaaagtccctaaatcagtgtccctggcaaattac
aacgcatcaaccttacaagacccaccttatttCCAAAACTCGCA**TTCCCACG**caccgttccagagtccggaaattaacca
aagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtcccaaacaacacgtacatagatg
gaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattgcaccgcatacaccttcccag
actcacctta**ttcccacg**caccgtcacagagtccggaaatcaagtagaaatccgtattcgcaggtccaaaaaaacacgta
catacatcggtaggtggactgccttcattcagtcggtcgcagagattcggAAATTCCCGCATGTGTgcaatGGTTTGaCC
CACTGTGCCATGTGCATAAGGGCTCGGAcCTCGGggactcggatcagacaCGGACTCGGGgtTCGttCgCtccCTACCTC
CGATTATAACTTATAACACCAACGAGCGAGAAAAAT**CTTGTGGAAACTTGAGGGCAAAGTGTTTCCCACG**ATTCGAATG
TCAGgctAAAaGAAAACAGAGgagagcacacacaagaaaacgcaca**cacgtg**catcggccGCATGTGCAACCAACCACCA
CCgccaagagcagcagcagcaagcagcaacaacaacagctatccaccaacacaggaatggaggagacaCACCATCacatC
ATCTTCATCCCCTGCCaAGACgACTTGTTTGTcGAGAAAAACCCACCACCGAAAGCAGGCGAGGATCGTGgctgtcggcc
cgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgaCCCAGTGACTGACgG**CAGGTG**GTcgcaatgg
tggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcgctctctctgcgcGCGCTGAGCACGAGACGaCTCT
GCtGCGG**CACGTG**CAGCTATAAAAGCgGCGGTAACtGGAGACGAgaGCAACATTctcaacaaattcgacgtgcagtgaac
acacacaagacaaccggatcccagccagatctcgctcaacttcttcacaattttacgtcaagaaaacacacatctaagct
ACAATGGCCACCAAATACGAagTtTCcAAgACCTAcCAGTACCGCAAGGTGATGAAGCCtcTGCTGGAaCGCAgGCGTCG
gGCCcGcATCAACAtGTGCCTGGACGAGCTCAAGGATCTGATGGtccagtgcacggtccagagcggtgatggcaaattcg
acagggccgacatactggaggtcactgtggaccatttgcgcaaactgaagcaggccaggatcgaggccaccgctgccgca
gcaaaagccaccaccaacactaccccagagcagagtttccgcgatggcttcatcagggccgccgacgaagtctcccgggc
actggcttccctacccaatgtggatgtcgtctttggcacccatctgatgacacacctcggcttgcgcctcaaccaattgg
aaacgccaactgcagcaccaaagccaatgaatgccccgctctccatcatctgcggcagcggcggcagcagcagcagcagc
agcaactgcagctccaacagccgggaggcctgcagtcccgtgtccagcggctactgcagcgacagcgattgcagtgtctc
ctccactcaagcggcccagaatctgctccagatcagcactggtcaggtttggcgcccctggtgaatctatagattggatt

Capital letters represent nucleotides or amino acids in the *D. pseudoobscura* reference sequence that are conserved in ***D. pseudoobscura*** and ***D. sechellia***.

```
TGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGtCATTTAGTTTGCATGTCCCTCGacccatccgtatcg
tatccctgtacctatacatatccctcaCTGGGTCCCACGATGATGATGTCCCGACGTGcccgggcATGCCACTTAGCGtt
cccttctctgttgcccctggccctgccttGGCCCAAGTGTTCAACAGAAGCTACAATTCcAGCATTCATAATTGACAAaa
cagacCtGCAGGAGggaaggagacggatagcgggagcgggacagagaacctattGGGCCCACGActGACAAGaaAAAAAA
TTAAgTAAAATAAAGGGCAGGTGGTAGTTTTTTttcaccactctccaTGACTCGCACTcGgAGCaCGGTCAGATTTACAG
TGGTacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattataaatttagtactattaaata
tgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgataccatcctgaccaataaaaaaaca
ttcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttccacccaaagaaaacctgcaa
tgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaatcagtgtccctggcaaattac
aacgcatcaaccttacaagacccaccttatttCCAAAACTCGCATTCCCACGcaccgttccagagtccggaaattaacca
aagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtcccaaacaacacgtacatagatg
gaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattgcaccgcatacaccttcccag
actcaccttattcccacgcaccgtcacagagtccggaaatcaagtagaaatccgtattcgcaggtccaaaaaaacacgta
catacatcggtaggtggactgccttcattcagtcggtcgcagagattcggAAATTCCCGCATGTGTgcaatGGTTTGaCC
CACTGTGCCATGTGCATAAGGGCTCGGAcCTCGGggactcggatcagacaCGGACTCGGGgtTCGttCgCtccCTACCTC
CGATTATAACTTATAACACCAACGAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAGTGTTTCCCACGATTCGAATG
TCAGgctAAAaGAAAACAGAGgagagcacacacaagaaaacgcacacacgtgcatcggccGCATGTGCAACCAACCACCA
CCgccaagagcagcagcagcaagcagcaacaacaacagctatccaccaacacaggaatggaggagacaCACCATCacatC
ATCTTCATCCCCTGCCaAGACgACTTGTTTGTcGAGAAAAACCCACCACCGAAAGCAGGCGAGGATCGTGgctgtcggcc
cgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgaCCCAGTGACTGACgGCAGGTGGTcgcaatgg
tggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcgctctctctgcgcGCGCTGAGCACGAGACGaCTCT
GCtGCGGCACGTGCAgCTATAAAAGCgGCGGTAACtGGAGACGAgaGCAACATTctcaacaaattcgacgtgcagtgaac
acacacaagacaaccggatcccagccagatctcgctcaacttcttcacaattttacgtcaagaaaacacacatctaagct
ACAATGGCCACCAAATACGAagTtTCcAAgACCTAcCAGTACCGCAAGGTGATGAAGCCtcTGCTGGAaCGcAgGCGTCG
gGCCcGcATCAACAtGTGCCTGGACGAGCTCAAGGATCTGATGGtccagtgcacggtccagagcggtgatggcaaattcg
acagggccgacatactggaggtcactgtggaccatttgcgcaaactgaagcaggccaggatcgaggccaccgctgccgca
gcaaaagccaccaccaacactaccccagagcagagtttccgcgatggcttcatcagggccgccgacgaagtctcccgggc
actggcttccctacccaatgtggatgtcgtctttggcacccatctgatgacacacctcggcttgcgcctcaaccaattgg
aaacgccaactgcagcaccaaagccaatgaatgccccgctctccatcatctgcggcagcggcggcagcagcagcagcagc
agcaactgcagctccaacagccgggaggcctgcagtcccgtgtccagcggctactgcagcgacagcgattgcagtgtctc
ctccactcaagcggcccagaatctgctccagatcagcactggtcaggtttggcgcccctggtgaatctatagattggatt
```

Capital letters represent nucleotides or amino acids in the *D. pseudoobscura* reference sequence that are conserved in ***D. pseudoobscura*** and ***D.yakuba***.

```
TGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCATGTCCCTCGacccatccgtatcg
tatccctgtacctatacatatccctcactGGGTCCCACGATGATGATGTCCCGACGTGcccgggcATGCCACTTAGCGtt
cccttctctgttgcccctggccctgccttggcCCAAGTGTTCAACAGAAGCTACAATTCcAGCATTCATAATTGACAAaa
cagacctgcaggagggaaggagacggatagcGGGAGCGggacagagaacctattGGGCCCACGActGACAAGAAAAAAaa
TTAAgTAAAATAAAGGGCAGGTGGTAGTTTTTTttcaccactctccaTGACTCGCACtcGgAGCaCGGTCAGATTTACAG
TGGTacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattataaatttagtactattaaata
tgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgataccatcctgaccaataaaaaaaca
ttcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttccacccaaagaaaacctgcaa
tgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaatcagtgtccctggcaaattac
aacgcatcaaccttacaagacccaccttatttccaaaactcgcattcccacgcaccgttccagagtccggaaattaacca
aagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtcccaaacaacacgtacatagatg
gaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattgcaccgcatacaccttcccag
actcaccttattcccacgcaccgtcacagagtccggaaatcaagtagaaatccgtattcgcaggtccaaaaaaacacgta
catacatcggtaggtggactgccttcattcagtcggtcgcagagattcgGAAATTCCCGCATGtgtgcaatGGTTTGaCC
CACTGTGCCATGTGCATAAGGGCTCGGAcctcggggACTCGGAtcagacaCGGACTCGGGgtTCGttCgCtccCTACCTC
CGATTATAACTTATAACACCAACGAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAGTGTTTCCCACGATTCGAATG
TCAGgctAAAaGAAAACAGAGgagagcacacacaagaaaacgcacacacgtgcatcggccGCATGTGCAACCAACCACCA
CCgccaagagcagcagcagcaagcagcaacaacaacagctatccaccaacacaggaatggaggagacaCACCATCacatC
ATCTTCATCCCCTGCCaAGACgACTTGTTTGTcGAGAAAAACCCACCACCGAAAGCAGGCGAGGATCGTGgctgtcggcc
cgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgaCCCAGTGACTGACgGCAGGTGGTcgcaatgg
tggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcgctctctctgcgcGCGCTGAGCACGAGACGaCTCT
GCtGCGGCACGTGCAGCTATAAAAGCgGCGGTAACtGGAGACGagaGCAACATTctcaacaaattcgacgtgcagtgaac
acacacaagacaaccggatcccagccagatctcgctcaacttcttcacaattttacgtcaagaaaacacacatctaagct
ACAATGGCCACCAAATACGAAgTtTCCAAGACCTAcCAGTACCGCAAGGTGATGAAGCCTCTGCTGGAaCGcAgGCGTCG
gGCCcGcATCAACAtGTGCCTGGACGAGCTCAAGGATCTGATGGtccagtgcacggtccagagcggtgatggcaaattcg
acagggccgacatactggaggtcactgtggaccatttgcgcaaactgaagcaggccaggatcgaggccaccgctgccgca
gcaaaagccaccaccaacactaccccagagcagagtttccgcgatggcttcatcagggccgccgacgaagtctcccgggc
actggcttccctacccaatgtggatgtcgtctttggcacccatctgatgacacacctcggcttgcgcctcaaccaattgg
aaacgccaactgcagcaccaaagccaatgaatgccccgctctccatcatctgcggcagcggcggcagcagcagcagcagc
agcaactgcagctccaacagccgggaggcctgcagtcccgtgtccagcggctactgcagcgacagcgattgcagtgtctc
ctccactcaagcggcccagaatCTGCTCCAGATCAGCACtGGTCAGGTtTGGCGCCCCTGGTgAatctatagattggatt
```

Capital letters represent nucleotides or amino acids in the *D. pseudoobscura* reference sequence that are conserved in ***D. pseudoobscura*** and ***D. erecta***.

```
TGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCATGTCCCTCGacccatccgtatcg
tatccctgtacctatacatatccctcactGGGTCCCACGATGATGATGTCCCGACGTGcccgggcATGCCACTTAGCGtt
cccttctctgttgcccctggccctgccttggcCCAAGTGTTCAACAGAAGCTACAATTcCAGCATTCATAATTGACAAaa
cagacctgcaggagggaaggagacggatAGCGGGAGCGggacagagaacctattGGGCCCACGActGACAAGAAAAAAaa
TTAAgTAAAATAAAGGGCAGGTGGTAGTTTTTTttcaccactctccaTGACTCGCACTcGgAGCaCGGTCAGATTTACAG
TGGTacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattataaatttagtactattaaata
tgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgataccatcctgaccaataaaaaaaca
ttcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttccacccaaagaaaacctgcaa
tgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaatcagtgtccctggcaaattac
aacgcatcaaccttacaagacccaccttatttccaaaactcgcattcccacgcaccgttccagagtccggaaattaacca
aagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtcccaaacaacacgtacatagatg
gaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattgcaccgcatacaccttcccag
actcaccttattcccacgcaccgtcacagagtccggaaatcaagtagaaatccgtattcgcaggtccaaaaaaacacgta
catacatcggtaggtggactgccttcattcagtcggtcgcagagattcgGAAATTCCCGCATGTGTgcaatggtttgaCC
CACTGTGCCATGTGCATAAGGGCTCGGAcctcggggACTCGGAtcagacaCGGACTCGGGgtTCGttCgCtccCTACCTC
CGATTATAACTTATAACACCAACgAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAGTGTTTCCCACGATTCGAATG
TCaGgctAAAaGAAAACAGAGgagagcacacacaagaaaacgcacacacgtgcatcggccGCATGTGCAACCAACCACCA
CCgccaagagcagcagcagcaagcagcaacaacaacagctatccaccaacacaggaatggaggagacaCACCATCacatC
ATCTTCATCCCCTGCCaAGACgACTTGTTTGTcGAGAAAAACCCACCACCGAAAGCAGGCGAGGATCGTGgctgtcggcc
cgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgaCCCAGTGACTGACgGCAGGTGGTcgcaatgg
tggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcgctctctctgcgcGCGCTGAGCACGAGACGaCTCT
gctgCGGCACGTGCAGCTATAAAAGCgGCGGTAACtGGAGACGAgaGCAACATTctcaacaaattcgacgtgcagtgaac
acacacaagacaaccggatcccagccagatctcgctcaacttcttcacaattttacgtcaagaaaacacacatctaagct
acAATGGCCACCAAATACGAagTtTCCAAgACcTAcCAGTACCGCAAGGTGATGAAGCCTcTGCTGGAaCGcAgGCGTCG
gGCCcGcATCAACAtGTGCCTGGACGAGCTCAAGGATCTGATGGtccagtgcacggtccagagcggtgatggcaaattcg
acagggccgacatactggaggtcactgtggaccatttgcgcaaactgaagcaggccaggatcgaggccaccgctgccgca
gcaaaagccaccaccaacactaccccagagcagagtttccgcgatggcttcatcagggccgccgacgaagtctcccgggc
actggcttccctacccaatgtggatgtcgtctttggcacccatctgatgacacacctcggcttgcgcctcaaccaattgg
aaacgccaactgcagcaccaaagccaatgaatgccccgctctccatcatctgcggcagcggcggcagcagcagcagcagc
agcaactgcagctccaacagccgggaggcctgcagtcccgtgtccagcggctactgcagcgacagcgattgcagtgtctc
ctccactcaagcggcccagaatctgctccagatcagcactggtcaggtttggcgcccctggtgaatctatagattggatt
```

Capital letters represent nucleotides or amino acids in the *D. pseudoobscura* reference sequence that are conserved in ***D. pseudoobscura*** and ***D. persimilis***.

```
TGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCATGTCCCTCGACCCATCCGTATCG
TATCCCTGTACCTATACATATCCCTCaCTGGGTCCCACGATGATGATGTCCCGACGTGCCCGGGCATGCCACTTAGCGTT
CCCTTCTCTGTTGCCCCTGGCCCTGCCTTGGCCCAAGTGTTCAACAGAAGCTACAATTCCAGCATTCATAATTGACAAAA
CAGACCTGCAGGAGGGAAGGAGACGGATAGCGGGAGCGGGACAGAGAACCTATTGGGCCCACGACTGACAAGAAAAAAAA
TTAAGTAAAATAAAGGGCAGGTGGTAGTTTTTTTCACCACTCTCCATGACTCGCACTCGGAGCACGGTCAGATTTACAG
TGGTACGAGTGACTGAAAAACTAgTGGCGAATTGTTTGTGAAGgAtATTTATAGAAtTATAAaTTTAGTACTATTAAAta
tgtacttccttcctatagaaatatgaaaaagaaataaaaaatgCACTGATAtTGATACCATCCTGaccaatAAAAAAACA
TTcCCATTAACtTcCCTTAGCAACATCTcgTAcGGAAATACCAGAATTCCTTAACGTTTCCACCCAAAgaaAACCTGCAA
TGtCAaACCTAGCACAAAACTCCcTACAGTTCACAcACTACgCCTAGAAAGTCCCTAAATCAGTGTCCCTGGCAAATTAC
AACGCATCAACCTTACAAGACCCACCTTATTTCCAAAACTCGCATTCCCACGCACCGTTCCAGAGTCCGGAAATTAACCA
AAGCAAGTAGAACACAGGCCACCTTTAAATCCGTAGAATACTGAAATATTCACAgGTCcCAAACAACACGTACATAGATG
GAAAACCTTCATTCAGTCGGTtcagTAGGTATCCAGAAATACCCACATGTGTGCAAATTGCACCgCATACACCTTCCCAG
ACTCACCTTATTCCCACGCACCGTCACAGAGTCCGGAAATCAAGTAGAAATCCGTAttcgcAGGTCCAAAAAAACACGta
caTACATCGGTAGGTGGACTGCCTTCATTCAGTCGGTCGCAGAGATTCGGAAATTCCCGCATGTGTGCAATGGTTTGACC
CACTGTGCCATGTGCATAAGGGCTCGGACCTCGGGGACTCGGATCAGACACGGACTCGGGGTTCGtTCGCTCCCTACCTC
CGATTATAACTTATAACACCAACGAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAGTGTTTCCCACGATTCGAATG
TCAGGCTAAAAGAAAACAGAGGAGAGCACACACAAGAAAACGCACACACGTGCATCGGCCGCATGTGCAACCAACCACCA
CCGCCAAGAGCAGCAGCAGCAAGCAGCAAcaacaacagctatccacCAACACAGGAATGGAGGAGACACACCATCACATC
ATCTTCATCCCCTGCCAAGACGACTTGTTTGTCGAGAAAAACCCACCACCGAAAGCAGGCGAGGATCGTGGctGTCGGCC
CGAGTAACGATTCGTAACGACGACACGAACGACAATCGGCTCGACCCGACCCAGTGaCTGACGGCAGGTGGTCGCAATGG
TGGTGCTCGCTCCGGTGGTGCCTCTCTCGTTTCCTGCGCATGCGCGCTCTCTCTGCGCGCGCTGAGCACGAGACGACTCT
GCTGCGGCACGTGCAGCTATAAAAGCGGCGGTAACTGGAGACGAGAGCAACATTCTCAACAAATTCGACGTGCAGTGAAC
ACACACAAGACAACCGGATCCCAGCCAGATCTCGCTCAACTTCTTCACAaTTTTACGTCAAGAAAACACACATCTAAGCT
ACAATGGCCACCAAATACGAAGTTTCCAAGACCTACCAGTACCGCAAGGTGATGAAGCCTCTGCTGGAACGCAGGCGTCG
GGCCCGCATCAACATGTGCCTGGACGAGCTCAAGGATCTGATGGTCCAGTGCACGGTCCAGAGCGGTGATGGCAAATTCG
ACAGGGCCGACATACTGGAGGTCACTGTGGACCATTTGCGCAAACTGAAGCAGGCCAGGATCGAGGCCACCGCTGCCGCA
GCAAAAGCCACCACCAACACTACcCCAGAGCAGAGtTTCCGCGATGGCTTCATCAGGGCCGCCGACGAAGTCTCCCGGGC
ACTGGCTTCCCTACCCAATGTGGATGTCGTCTTTGGCACCCATCTGATGACACACCTCGGCTTGCGCCTCAACCAATTGG
AAAcGCCAACTGCAGCACCAAAGCCAATGAATGCCCCGCTCTCCATCATCTGCGGCAGCGGCGGCAGCAGCAGCAGCAGC
AGCAACtgCAGCTCCAACAGCCGGGAGGCCTGCAGTCCCGTGTCCAGCGGCTACTGCAGCGACAGCGATTGCAGtGTCTC
CTCCAcTCAAGCGGCCCAGAATCTGCTCCAGATCAGCACTGGTCAGGTTTGGCGCCCCTGGTGAATCTATaGATTGGATT
```

Capital letters represent nucleotides or amino acids in the *D. pseudoobscura* reference sequence that are conserved in ***D. pseudoobscura*** and ***D. ananassae***.

```
TGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCATGTCCCTCGAcCCATCcgtatcg
tatccctgtacctatacatATCCCTcaCTGGGTCCCACGATGATGATGTCCCGACGTGCCCgggcATGCCACTTAGCGTT
cccttctctgttgcccctggccctgccttgGCCCAAGTGTTCAACAGAAGCTACAAtTCCAGCATTCATAATTGACAAaa
cagacctgcaggagggaaggagacggatagcgggagcgggacagagaacctattgggcccacgaCTGACAAGAAAAAAaa
TTAAgTAAAATaAAGGGCAGGTGGTaGTTTTTTTtcACCACTctCCaTGACTCGcACTCGgAGCaCGGTCAGATTTACAG
TGGtacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattataaatttagtactattaaata
tgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgataccatcctgaccaataaaaaaaca
ttcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttccacccaaagaaaacctgcaa
tgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaatcagtgtccctggcaaattac
aacgcatcaaccttacaagacccaccttatttccaaaactcgcattcccacgcaccgttccagagtccggaaattaacca
aagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtcccaaacaacacgtacatagatg
gaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattgcaccgcatacaccttcccag
actcaccttattcccacgcaccgtcacagagtccggaaatcaagtagaaatccgtattcgcaggtccaaaaaaacacgta
catacatcggtaggtggactgccttcattcagtcggtcgcagagattcggaaattcccgcatgtgtgcaatGGTTTGaCC
CACTGTGCCATGTGCATAAGGGCTCGgacctcggggactcggatcagacacggactcggggttcgttcgctcccTACCTC
CGATTATAACTTATAACACCAACGAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAGTGTTTCCCACGATTCGAATG
TCaGgctAAAaGAAAACAGAGgagagcacacacaagaaaacgcacacacgtgcatcggccGCATGTGCAACCAACCACCA
CCgccaagagcagcagcagcaagcagcaacaacaacagctatccaccaacacaggaatggaggagacaCACCATCacatc
ATCTTCATCCCCTGCCaagacgACTTGTTTGTcgAGAAAAACCcACCACCgAAAGCAGGCGAGGATCGTGgctgtcggcc
cgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgaCCCAGTGaCTGACGGCAGGTGgtcgcaatgg
tggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcgctcTCTCTGCGCGCGCTGAGCACGAGACGaCTCT
gctgcGGCACGTGCAGCTATAAAAGCGGcGGTAACTGGAGACgagagCAACATTCTcAACaaAttCgaCGTGCAGTgAAC
AcacacaagacaaccggatcccagccagatctcgctcaacttcttcacaattttacgtcaagAAAACACAcatctaagct
ACAATGGCCACCAAATACGAagTtTCCAAGACCTACCAGTACCGCAAGGTGATGAAGCCtCTGCTGGAaCGCAgGCGtCG
gGCCcGcATCAACAtGTGCCTGGACGAGCTCAAgGATCTGATGGtCcAGTGCacGGtCCAGAgCGGtGATGgCAAaTTCG
AcAggGCCGACATaCTGGAgGTcACtGTGgAcCAttTgCGCAAACTGAAGCAgGCCAggatcgaggccaccgctgccgca
gcaaaagccaccaccaacactaccccagagcagagtttccgcgatggcttcatcagggccgccgacgaagtctcccgggc
aCTGGCtTCCCTACCcAAtGTGGATGTcGtCTTtGGCACCcaTCTGATGACACACCTCGGctTGCGcCTCAACCAattgg
aaacgccaactgcagcaccaaagccaatgaatgccccgctctccatcatctgcggcagcggcggcagcagcagcagcagc
agcaactgcagctccaacagccgggaggcctgcagtcccgtgtccagcggctactgcagcgacagcgattgcagtgtctc
ctccactcaagcggcccagaatctgctcCAGATCAGCActggtCAGGTtTGGCGCCCCTGGTgaatctatagattggatt
```

Capital letters represent nucleotides or amino acids in the *D. pseudoobscura* reference sequence that are conserved in ***D. pseudoobscura*** and ***D. grimshawi***.

```
TGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCATTTAGTTTGCATGtCCCTCgacccatccgtatcg
tatccctgtacctatacatatccctcactgGGTCCCACGAtgatgatgtcccgaCGTGCCCgGGCATGCCACTTAGCGtt
cccttctctgttgcccctggccctgcctTGGCCCAAGTGTTCAACAGAAGCTACAATTcCAGCATTCATAATTGACAAAA
CAgacctgcaggagggaaggagacggatagcgggagcgggacagagaacctattgggcccacgactgACAAgAAAAAAAA
ttaagtaAAATAAAgGGCAGGTGGTAGTTTTTTTtcaccactcTCCATGACTCGcactcggaGCACGGTCAGATTTACAG
TGGtacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattataaatttagtactattaaata
tgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgataccatcctgaccaataaaaaaaca
ttcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttccacccaaagaaaacctgcaa
tgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaatcagtgtccctggcaaattac
aacgcatcaaccttacaagacccaccttatttccaaaactcgcattcccacgcaccgttccagagtccggaaattaacca
aagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtcccaaacaacacgtacatagatg
gaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattgcaccgcatacaccttcccag
actcaccttattcccacgcaccgtcacagagtccggaaatcaagtagaaatccgtattcgcaggtccaaaaaaacacgta
catacatcggtaggtggactgccttcattcagtcggtcgcagagattcggaaattcccgCATGTGTgcaatggtttgACC
CACTGTGCcATGTGCATAAGGGCTCggacctcggggactcggatcagacacggactcggggttcgttcgctccctacctc
CGATTATAACTTATAACACCAACGAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAGTGTTTCCCACGATTCGAAtG
TCAggcTAAAAGAAAACAGAGgAGAGCacacacaagaaaacgcacacacgtgcatcggccgcatgtgcaaccaaccacca
ccgccaagagcagcagcagcaAGCAGCAACAACAACAGCtatccaccaacacaggaatggaggagacacaccatcaCATC
AtCTTCATCCCCTGCCAAGACgACTTGTTTGtCGAGAAAAACCcACCACCgaaaGCAGgCGAGGATCGtggctgtcggcc
cgagtaacgattcgtaacgacgacacgaacgacaatcggctcGACCCGAcccagTGACTGACGGCAGGTGGTcgcaatgg
tgGTGCTCGCTccggtggtgcctctctcgtttcctgcgcatgcgcgctctctcTGCGcGCGCTGAGCACGAGACGactct
gctgCGGCACGTGCAGCTATAAAAGCgGCgGTAACTGGAGACGagAGCAACATTCTCAacaaattcgacgtgcagtgaac
acacacaagacaaccggatcccagccagatctcgctcaacttcttcacaattttacgtcaagaaaacacacatCTAAGCt
acaATGGCcACCAAATACGAagTtTCcAAGACcTAcCAGTACCGCAAGGTGATGAAGCCtCTGCTGGAaCGCAgGCGTCG
gGCcCGCATCAACAtGTGCcTGGAcGAgCTcAAGGATCTGATGGtCcAGTGcacgGtCCAgAgcgGtGATGgCAAaTTCG
AcAggGCCGACATacTGGAGGTCACTGTggAccATTTGCGCaAACTgAAGCAggccaggatcgaggccaccgctgccgca
gcaaaagccaccaccaacactaccccagagcagagtttccgcgatggcttcatcagggccgccgacgaagtctcccgggc
actggcttccctacccaatgtggatgtcgtctttggcacccatctgatgacacacctcggcttgcgcctcaaccaattgg
aaacgccaactgcagcaccaaagccaatgaatgccccgctctccatcatctgcggcagcggcggcagcagcagcagcagc
agcaactgcagctccaacagccgggaggcctgcagtcccgtgtccagcggctactgcagcgacagcgattgcagtgtctc
ctccactcaagcggcccagaatctgctccagatcagcactggtcaggtttggcgcccctggtgaatctatagattggatt
```

Capital letters represent nucleotides or amino acids in the *D. pseudoobscura* reference sequence that are conserved in ***D. pseudoobscura*** and ***D. virilis***.

```
TGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCcGTACAGTCATTTAGTTTGCATGtccctcgacccatccgtatcg
tatccctgtacctatacatatccctcactgGGTCCCAcgatgatgatgtccCGACGTGCCCgGGCATGCCACTTAGCGtt
cccttctctgttgcccctggccctgcctTGGCCCAAGTGTTCAACAGAAGCTACAATTCcAGCATTCATAATTGACAAaa
cagacctgcaggagggaaggagacggatagcgggagcgggacagagaacctattgggcccacgactgaCAAGAAAaaaaa
ttaagtAAAATAAAgGGCAGGTGGTAGTTTTTTTtCACCACTCtCCATGACTCGcactcggaGCACGGTCAGATTTACaG
TGGtacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattataaatttagtactattaaata
tgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgataccatcctgaccaataaaaaaaca
ttcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttccacccaaagaaaacctgcaa
tgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaatcagtgtccctggcaaattac
aacgcatcaaccttacaagacccaccttatttccaaaactcgcattcccacgcaccgttccagagtccggaaattaacca
aagcaagtagaacacaggccacctttaaatccgtagaatactgaaatattcacaggtcccaaacaacacgtacatagatg
gaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattgcaccgcatacaccttcccag
actcaccttattcccacgcaccgtcacagagtccggaaatcaagtagaaatccgtattcgcaggtccaaaaaaacacgta
catacatcggtaggtggactgccttcattcagtcggtcgcagagattcggAAATTcCCGCATGTGtgcaatggTTTGaCC
CACTGTGCcATGTGCATAAGGGCTCggAcCTCGGGGACTCGGatcagacacggactcggggttcgttcgctccctacctC
CGATTATAACTTATAACACCAACGAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAGTGTTTCCCACGATTCGAAtg
TCAggcTAAAAGAAAACAGAGgAGAGCacacacaagaaaacgcacacacgtgcatcggccgcatgtgcaaccaaccacca
ccgccaagagcagcagcagcAAGCAGCAACAACAACAgctatCCACCAACACAggaatggaggagacacaccatcaCATC
AtCTTCATCCCCTGCCAAGACgACTTGTTTGtCGAGAAAAACCcACCACCgaaagcaggCGAGGATCGTggctgtcggcc
cgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgacccagtgactgacggcaggtggtcgcaatgg
tggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcgCTCTCTcTGCGcGCGCTGAGCACGAGACGactct
gctgCGGCACGTGCAGCTATAAAAGCgGCGGTAACTGGAGACGAgAGCAACATTCTCAACaaattcgacgtgcagtgaac
acacacaagacaaccggatcccagccagatctcgctcaacttcttcacaattttacgtcaagaaaacacacatctaagct
acaATGGCCACCAAATACGAagTtTCcAAGACCTAcCAGTACCGCAAGGTGATGAAGCCtCTGCTGGAaCGCAggCGTCG
gGCcCGCATCAACAtGTGCCTGGACGAgCTcAAGGATCTGATGGtccagtgcacggtccagagcggtgatggcaaattcg
acagggccgacatactggaggtcactgtggaccatttgcgcaaactgaagcaggccaggatcgaggccaccgctgccgca
gcaaaagccaccaccaacactaccccagagcagagtttccgcgatggcttcatcagggccgccgacgaagtctcccgggc
actggcttccctacccaatgtggatgtcgtctttggcacccatctgatgacacacctcggcttgcgcctcaaccaattgg
aaacgccaactgcagcaccaaagccaatgaatgccccgctctccatcatctgcggcagcggcggcagcagcagcagcagc
agcaactgcagctccaacagccgggaggcctgcagtcccgtgtccagcggctactgcagcgacagcgattgcagtgtctc
ctccactcaagcggcccagaatctgctccagatcagcactggtcaggtttggcgcccctggtgaatctatagattggatt
```

Capital letters represent nucleotides or amino acids in the *D. pseudoobscura* reference sequence that are conserved in ***D. pseudoobscura*** and ***D. mojavensis***.

```
TGCGATACATTTAATCTTTAGTGTGTGTCAATTTGCCGTACAGTCaTTTAGTTTGCATGtccctcgacccatccgtatcg
tatccctgtacctatacatatccctcactgGGTCCCAcgatgatgatgtccCGACGTGCCCgGGCATGCCACTTAGCGTt
cccttctctgttgcccctggccctgccttGGCCCAAGTGTTCAACAGAAGCTACAATTCcAGCATTCATAATTGACAAaa
cagacctgcaggagggaaggagacggatagcgggagcgggacagagaacctattgggcccacgactgaCAAGAAAAAAAA
TTaagtaAAATAAAgGGCAGGTGGTAGTTTTTTttCACCACTCTCCATGACTcgcactcggaGCACGGTCAGATTTACAG
TGGtacgagtgactgaaaaactagtggcgaattgtttgtgaaggatatttatagaattataaatttagtactattaaata
tgtacttccttcctatagaaatatgaaaaagaaataaaaaatgcactgatattgataccatcctgaccaataaaaaaaca
ttcccattaacttcccttagcaacatctcgtacggaaataccagaattccttaacgtttccacccaaagaaaacctgcaa
tgtcaaacctagcacaaaactccctacagttcacacactacgcctagaaagtccctaaatcagtgtccctggcaaattac
aacgcatcaaccttacaagacccaccttatttccaaaactcgCATTCCCACGCAccgttccagagtccggaaattaacca
aAGCAAGTAGAAcacaggccacctttaaatccgtagaatactgaaatattcacaggtcccaaacaacacgtacatagatg
gaaaaccttcattcagtcggttcagtaggtatccagaaatacccacatgtgtgcaaattgcaccgcatacaccttcccag
actcaccttattcccacgcaccgtcacagagtccggaaatcaagtagaaatccgtattcgcaggtccaaaaAAACACGTA
catacatcggtaggtggactgccttcATTCAGTcggtcgcagagattcggAAATTcCCGCATGTGTgcaatggtttgaCC
CACTGTGCcATGTGCATAAGGGCTCggacctcggggactcggatcagacacggactcggggttcgttcgctccctacctc
CGATTATAACTTATAACACCAACGAGCGAGAAAAATCTTGTGGGAAACTTGAGGGCAAAGTGTTTCCCACGATTCGAAtG
TCAggcTAAAAGAAAACAGAGgAGAGCacacacaagaaaacgcacacacgtgcatcggccgcatgtgcaaccaaccacca
ccgccaagagcagcagcagcaagcagcaacaacaacagctatCCACCAACACAggaatggaggAGACaCACCATCAcatc
atcTTCATCCCCTGCCAAGACgACTTGTTTGtCGAGAAAAACCcACCACCgaaAGCAGGCGAGGATCGtggctgtcggcc
cgagtaacgattcgtaacgacGACACGAACgacaatcggctcgacccgacccagtgactgacggcaggtggtcgcaatgg
tggtgctcgctccggtggtgcctctctcgtttcctgcgcatgcgcGCTCTCTcTGCGcGCGCTGAGCACGAGACGactct
gctgCGGCACGTGCAGCTATAAAAGCgGCgGTAACTGGAGAcGagAGCAaCATTCTCAACaaAtTCGaCGTGCAGTgaAC
ACAcacaagacaaccggatcccagccagatctcGCTCAACTTcttcacaattttacgtcaagaaaacacacatctaagct
acaATGGCCACCAAATACGAAgttTCcAAGACcTAcCAGTACCGCAAGGTGATGAAGCCtCTGCTGGAACGCAgGCGTCG
gGCcCGCATCAACAtGTGCCTGGAcGAgCTCAAGGATCTGATGGtCcAGTGcacGGtCCAGAgcGGTGATGgCAAATTCG
AcAgGGCCGACATAcTGGAGGTCACtGTgGAccATtTGCGCAAACTGAAGCAGgccaggatcgaggccaccgctgccgca
gcaaaagccaccaccaacactaccccagagcagagtttccgcgatggcttcatcagggccgccgacgaagtctcccgggc
actggcttccctacccaatgtggatgtcgtctttggcacccatctgatgacacacctcggcttgcgcctcaaccaattgg
aaacgccaactgcagcaccaaagccaatgaatgccccgctctccatcatctgcggcagcggcggcagcagcagcagcagc
agcaactgcagctccaacagccgggaggcctgcagtcccgtgtccagcggctactgcagcgacagcgattgcagtgtctc
ctccactcaagcggcccagaatctgctccagatcagcactggtcaggtttggcgcccctggtgaatctatagattggatt
```

Figure 11: EvoPrinter results from one on one comparison with *D. pseudoobscura* (Couronne et al. 2003).

EvoPrinter alignments were created using *D. pseudoobscura* as a reference. Ten other *Drosophila* species were aligned directly with the reference sequence. The upstream regulatory binding sites were identified and highlighted accordingly.

Table 8: Summary of results for *D. pseudoobscura* (Couronne et al. 2003) EvoPrint alignments.

	8 - E-box	7 - Su(H)	6 - Su(H)	5 - N-box	4 - Paired Su(H)	3 - N-box	2 - E-box	1 - N-box
D. melanogaster	+	+	not present	+	+	not present	+	+
D. simulans	+	+	not present	+	+	not present	+	+
D. sechellia	+	+	not present	+	+	not present	+	+
D. yakuba	+	not present	not present	+	+	not present	+	+
D. erecta	+	not present	not present	+	+	not present	+	+
D. persimilis	+	+	+	+	+	+	+	+
D. pseudoobscura	+	+	+	+	+	+	+	+
D. ananassae	+	not present	not present	+	+	not present	+	+
D. grimshawi	+	not present	not present	+	+	not present	+	+
D. virilis	+	not present	not present	+	+	not present	not present	+
D. mojavensis	+	+	not present	+	+	not present	not present	+

As the ten *Drosophila* species were aligned directly against the *D. pseudoobscura* reference, some upstream regulatory binding sites did not remain conserved. The sites that were lost were marked with "not present," and sites that remained conserved are marked with a "+".

Table 9: Flanking bases of E-box sites in *D. melanogaster* and *D. pseudoobscura.*

	D. melanogaster			*D. pseudoobscura*	
Species	**XI**	**III**	**II**	**8**	**2**
D. simulans	GG	GT	GG	GG	GG
D. sechellia	GG	GT	GG	GG	GG
D. melanogaster	GG	GT	GG	GG	GG
D. yakuba	GG	__T	GG	GG	GG
D. erecta	GG	__T	GG	GG	GG
D. persimilis	GG	not conserved	GG	GG	GG
D. pseudoobscura	GG	not conserved	GG	GG	GG
D. ananassae	GG	not conserved	not conserved	GG	G__
D. grimshawi	GG	not conserved	GG	GG	GG
D. virilis	GG	not conserved	GG	GG	not conserved
D. mojavensis	GG	not conserved	__G	GG	not conserved

To investigate the level of conservation of E-box sites with respect to the *D. pseudoobscura* and *D. melanogaster* reference sequences, this table illustrates which sites were conserved and also which bases flank either side of the E-box binding site. The sites not conserved are marked "not conserved" and the sites with a different flanking base on one side are marked with a "__".

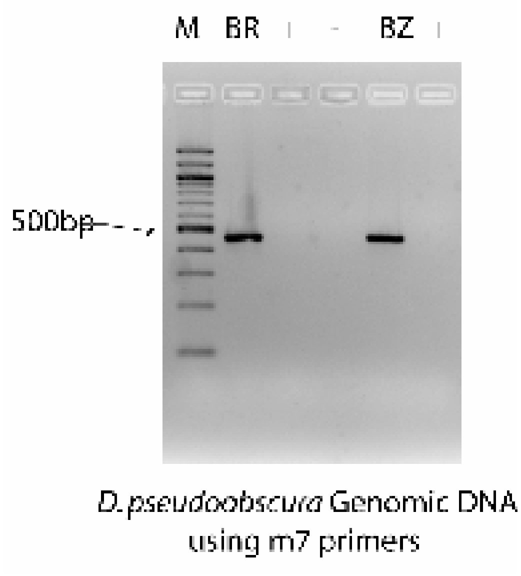
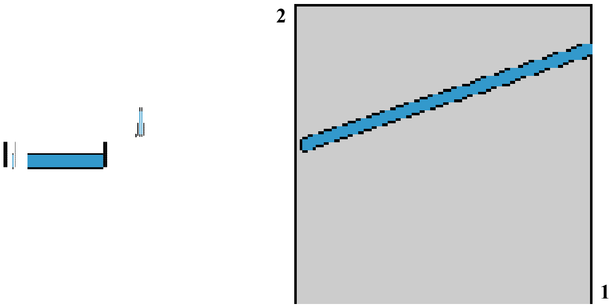


Figure 12: Gel electrophoresis of *m7* PCR products.
Using the primers designed to isolate the upstream regulatory region of *D. pseudoobscura* the PCR products were run through an agarose gel. The desired product size was 456 base pairs. Two sources of *D. pseudoobscura* genomic DNA were used (BR and BZ); "+" indicates positive control and "-" indicates negative control.

BLAST 2 SEQUENCES RESULTS VERSION BLASTN 2.2.12 [Aug-07-2005] Top of Form

Sequence 1 ***m7*** **(DNA Tools 2006)**	lcl|seq_1	**Length**	362	(1 .. 362)
Both are *D. pseudoobscura*				
Sequence 2 ***m7*** **(Couronne et al. 2003)**	lcl|seq_2	**Length**	1082	(1 .. 1082)



```
Score =  575 bits (299), Expect = e-161
Identities = 330/353 (93%), Gaps = 1/353 (0%)
 Strand = Plus / Plus
```

```
Query: 10   gccgcatgtgcaaccaaccaccaccgtccaagnnnnnnnnnnnnnnnnnnnnacaacaac 69
            |||||||||||||||||||||||||| |||||                    ||||||||
Sbjct: 577 gccgcatgtgcaaccaaccaccaccg-ccaagagcagcagcagcaagcagcaacaacaac 635


Query: 70   agctatccgccaacacaggaatggaggagacacaccatcacatcatcttcatcccctgcc 129
            |||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 636 agctatccaccaacacaggaatggaggagacacaccatcacatcatcttcatcccctgcc 695


Query: 130 aagacgacttgtttgtcgagaaaaacccaccaccgaaagcaggcgaggatcgtggccgtc 189
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||
Sbjct: 696 aagacgacttgtttgtcgagaaaaacccaccaccgaaagcaggcgaggatcgtggctgtc 755


Query: 190 ggcccgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgacccagtg 249
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 756 ggcccgagtaacgattcgtaacgacgacacgaacgacaatcggctcgacccgacccagtg 815


Query: 250 actgacggcaggtgggtcgcaatggtggtgctcgctccggtggtgcctctctcgtttcctg 309
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 816 actgacggcaggtgggtcgcaatggtggtgctcgctccggtggtgcctctctcgtttcctg 875


Query: 310 cgcatgcgcgctctctctgcgcgcgctgagcacgagacgactctgctgcggca 362
            |||||||||||||||||||||||||||||||||||||||||||||||||||||
Sbjct: 876 cgcatgcgcgctctctctgcgcgcgctgagcacgagacgactctgctgcggca 928

CPU time:     0.01 user secs.      0.00 sys. secs       0.01 total secs.
Lambda      K       H
    1.33    0.621     1.12
Gapped
Lambda      K       H
    1.33    0.621     1.12
```

Figure 13: BLAST result of two *m7 D. pseudoobscura* sequences.
A BLAST alignment was created using the *D. pseudoobscura* sequence obtained from the DF-HCC Lab and the *D. pseudoobscura m7* sequence from Couronne et al. (2003). There was a 93% identity between the two sequences; differences are indicated by yellow highlighting. *D. pseudoobscura*'s E-box (site 2) is annotated in blue.

Table 10: Percent similarity between each of the eleven species of *Drosophila* to *D. melanogaster* and *D. pseudoobscura.*

Species	Percent similarity to *D. melanogaster*	Percent similarity to *D. pseudoobscura*
D. melanogaster	**100**	75
D. simulans	**82**	75
D. sechellia	**82**	75
D. yakuba	**82**	63
D. erecta	**73**	63
D. persimilis	55	**100**
D. pseudoobscura	55	**100**
D. ananassae	36	**63**
D. grimshawi	55	**63**
D. virilis	45	**50**
D. mojavensis	45	**63**

To determine percent similarity between species, the number of conserved sites a particular species has with respect to the reference was divided by the total number of sites in the reference sequence. Bold percentages indicate a greater similarity to the reference.