A STUDY OF NON-CODING ROX1 AND ROX2 RNAS IN DROSOPHILA SPECIES

by

PING HU

(Under the Direction of Russell Malmberg)

ABSTRACT

rox1 and rox2 are two non-coding RNAs found in male *Drosophila melanogaster*, which can direct the MSL complex to bind to the X-chromosome to up-regulate all X-linked genes two-fold. rox1 and rox2 RNAs lack similarity in their primary sequences, but they have functional redundancy; that is, in the presence either one of them, males survive. The mechanism of this redundancy is unclear. Here we have used comparative bioinformatics methods to predict an double internal-loop structure near 3'-end of rox RNAs shared by rox1 and rox2; this structure might contribute to the rox1 and rox2 RNA functional redundancy. rox1 and rox2 RNAs in other *Drosophila* species have not been previously reported; we identify putative rox1 and rox2 RNAs in *Drosophila simulans, sechellia, yakuba* and *erecta*.

INDEX WORDS: Non-coding RNA, rox1 RNA, rox2 RNA, dosage compasation, *Drosophila* species

A STUDY OF NON-CODING ROX1 AND ROX2 RNAS IN DROSOPHILA SPECIES

by

PING HU

B.S., China Agricultural University, P.R. 1990

M.S., Chinese Academy of Agricultural Sciences, P.R. 1996

M.S., University of Georgia, 2003

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment

of the Requirements for the Degree

MASTER OF SCIENCE
ATHENS, GEORGIA
2007

A STUDY OF NON-CODING ROX1 AND ROX2 RNAS IN DROSOPHILA SPECIES

by

PING HU

Major Professor:	Russell Malmberg
Committee:	Michael McEachern Paul Schliekelman

Electronic Version Approved:

Maureen Grasso
Dean of the Graduate School
The University of Georgia
August 2007

DEDICATION

To my mother Ruiyun Yue, my father Xuede Hu, and my sister Guang Hu, whose love and support gave me strength, courage and hope to overcome all difficulty in the life.

To my husband Huanli Liu, my children Yupeng Liu and Yuqi Liu, for always being my sources of love and happiness…

ACKNOWLEDGEMENTS

I would like to thank all people who helped me during my study and research. I especially want to thank my major professor, Dr. Russell Malmberg, for all the assistance and advice he has given me during my research. I am grateful to my committee members, Dr. Michael McEachern and Dr.Paul Schliekelman for their helpful advice and positive encouragement. My special thanks go to Dr. Kelly Dawe for opening up new fields of knowledge to me through my research.

TABLE OF CONTENTS

Page

LIST OF TABLES

Page

LIST OF FIGURES

Page

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

INTRODUCTION

According to the central dogma of molecular biology, messenger RNA (mRNA), which is transcribed from DNA, carries the genetic information and acts as a template in protein synthesis by ribosome. In the past 20 years, RNA has been discovered to have many new functions besides coding for proteins. The transcribed non-coding RNA (ncRNA) molecules can fold into stable secondary and complicated tertiary structures, directly functioning as structural, catalytic, or regulatory factors rather than produce protein products in the cells.

SOURCES OF NON-CODING RNAs

ncRNAs may come from three sources:

- Noncoding RNA genes encode ncRNAs. Examples are the non-coding rox1 and rox2 RNAs in male *Drosophila melanogaster* [1, 2], and the widely existing miRNAs. It has been reported that 10% of the genomic genes in mammals are miRNA genes [3]. Most ncRNAs are not polyadenylated.
- ncRNAs may derive from the introns of mRNAs coding proteins, for example, snoRNAs [4].
- Several important regulatory motifs are found in untranslated regions (UTRs) of mature eukaryotic mRNAs. Internal ribosome entry site (IRES) elements are found in the 5'-UTR region of picornavirus [5] and some other viruses; some specific mRNAs can be translated through IRES under stress conditions when cap-dependent translation is

blocked. Iron response elements (IRE) are found in both 5' UTR and 3'UTR to regulate the metabolism of iron in the cells [6, 7].

BIOLOGICAL FUNCTIONS OF ncRNAs

The functions of two classic ncRNAs (rRNA and tRNA) are well known. rRNAs are part of the ribosomal complex mediating amino acids transfer and polypeptide chain elongation. tRNAs play an important role in transporting amino acids to messenger RNA during the translation process. Both of them participate in the translation process. In the last 20 years, many studies have uncovered new biological functions of ncRNA, which include

- Maturation of RNAs: A large number of ncRNAs are involved in the maturation of RNAs. For example snRNAs, a component of spliceosome, play a key role in mRNA maturation by recognition of the intron splicing sites [8]; RNase P functions in 5'end maturation of tRNA [9]; guide RNAs (gRNAs) are involved in editing RNA Precursors [10].
- RNA modification: In eukaryotes, the site-specific modification is directed by C/D box or H/ACA box snoRNAs through base pairing with target RNAs. The function of the C/D box RNAs is to methylate 2-*O*-ribose of target nucleotides, while H/ACA box RNAs guide the conversion of specific uridine residues to pseudouridine [11].
- Regulation of gene expression and translation: The well known miRNAs, 21-23 nts sequence, imperfectly base pair with the 3'-UTR of transcript mRNA of target gene to inhibit its translation. Similarly, RNAi (22 nts) causes targeted gene silence by perfectly base pairing with mRNA to trigger its degradation [12, 13].

- Regulating the fidelity of DNA replication: Telomerase RNA provides the template for addition of the telomeric repeats to the ends of chromosomes [14].
- Regulation of dosage compensation: In female mammals, the 17 kb Xist (X- inactivate specific transcript) ncRNA spreads along the x-chromosome to inactivate most gene expression by a cis-acting mechanism [15]. By contrast, the 3.7 kb rox1 and 1.2 kb rox2 RNAs found in male *Drosophila melanogaster* can direct MSL (male-specific- lethal) complex spread along the whole X- chromosome to up-regulate the transcription rate of X-linked genes about 2-fold [1, 2, 16, 17].
- Some ncRNAs are involved in epigenetic regulation of imprinted gene silencing, where gene expression is restrained to the parental-specific alleles. The expression of the non-coding Air RNA can inhibit three paternal protein-coding genes (Igf2r/Slc22a2/Slc22a3) expression in mice [18]; the expression of H19 ncRNA gene found in human, mouse, rat and rabbit is exclusive to maternal chromosomes, and the 2.3 kb transcripts may function as a tumor suppressor [19].

As the number of genome sequencing projects increases, the number of non-coding genes found should also increase dramatically. The data from genomic projects reveals that the number of protein coding genes is lower than was expected. In humans, only about 1.4 % of total RNAs are translated to proteins; the functions of the rest of the RNAs are little known [20]. Interest in studying ncRNA has been growing in the bioinformatics field. Several well-annotated ncRNA databases have appeared, such as, the Rfam database with over 280,000 regions of 379 families [21]; the noncoding RNA database contains 109 " transitional " classes and nine groups [22]; the RNAdb includes over 800 known mammalian ncRNAs, but excludes tRNAs, rRNAs and

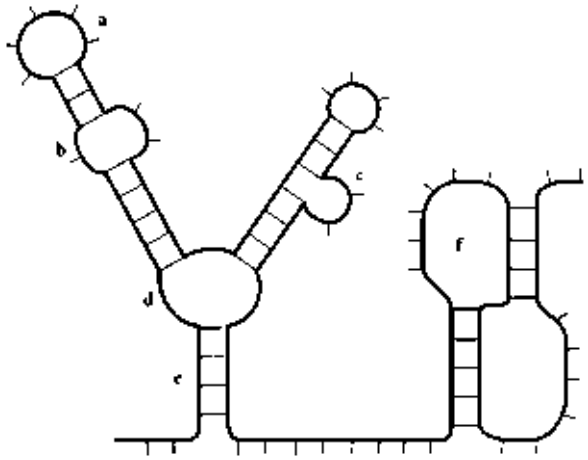
snRNAs [23] and the Arabidopsis small RNA project (ASRP). The functions of several classes of ncRNAs have been analyzed, and are summarized in Table 1.1 [24].

Table 1.1 Major classes of functional RNAs (Bompfűnewerer, et al., 2005)

Class	Size	Function	Phylogenetic Distribution
tRNA	70-80	translation	ubiquitous
rRNA			
16S/18S	1.5k	translation	ubiquitous
28S+5.8S/23S	3k	translation	ubiquitous
5S	130	translation	ubiquitous
RNase P	220-440	tRNA maturation	ubiquitous
MRP	250-350	endonuclease, 5.8S rRNA maturation	eukarya
snoRNA H/ACA	~130	pseudouridinylation in rRNAs	eukarya
C/D	60-80	ribose 2'-O-methylation in rRNAs	eukarya, archaea
telomerase	400-550		eukarya
snRNA U1,U2,U4,U5,U6	100-160	major spliceosome, mRNA maturation	eukarya
U11,U12	130-140	minor spliceosome, mRNA maturation	eukarya
SL	~100	trans-splicing	lower eukaryotes
U7	~65	histone mRNA maturation	eukaryotes
7SK	~300	transcriptional regulation	vertebrata
7SL/SRP	300-400	signal recognition particle	ubiquitous
vault	80-100	part of vault particle	vertebrate
Y	80-100	part of Ro particle	metazoans
tmRNA	300-400	tags protein for proteolysis	bacteria, chloroplasts, cyanoplasts
miRNA	~22	post-transcriptional regulation	multicellular organisms

RNA SECONDARY STRUCTURES

Single strand RNA molecules fold back on themselves to form different shapes. The interactions between nucleotides in the molecule determine what kinds of structures can be formed, such as stem-loops, internal loops or the more complicated pseudoknots. Figure 1.1 illustrates several RNA secondary structures [William Liu, CS374 lecture Notes, available at web http://ai.stanford.edu/~serafim/CS374_2004/].



a. hairpin loop
b. internal loop
c. bulge loop
d. multibranched loop
e. stem
f. pseudoknot

Figure 1.1 RNA secondary structure

BLAST searching and other kinds of gene finding algorithms and comparative genome analysis have been successfully applied to identify protein coding genes, homologs and potential functions, but these methods don't work as well on ncRNA genes. The primary sequence of ncRNAs varies over a relatively short evolution distance, making the use of BLAST difficult. Even though some ncRNAs share similar functions, they may have varied sequence lengths or nucleotide composition, and there are no ORFs. The known ncRNAs have a highly conserved secondary structure over evolution. A stable structure by itself is not a significant signal of function, since all single strand RNAs can easily form secondary structure with the canonical base pairing system A-U, G-C and additional base pairing G-U. Thus, a reasonable way to

identify ncRNAs may be to combine sequence analysis and conserved secondary structure analysis together in genomic searches.

Similarity searching is performed by dynamic programming algorithms based on both sequence similarity and structural similarity; the alignment score comes from both sequence conservation and structure conservation. Figure 1.2 is an example of how to identify structural homologs [William Liu, CS374 lecture Notes, http://ai.stanford.edu/~serafim/CS374_2004/].

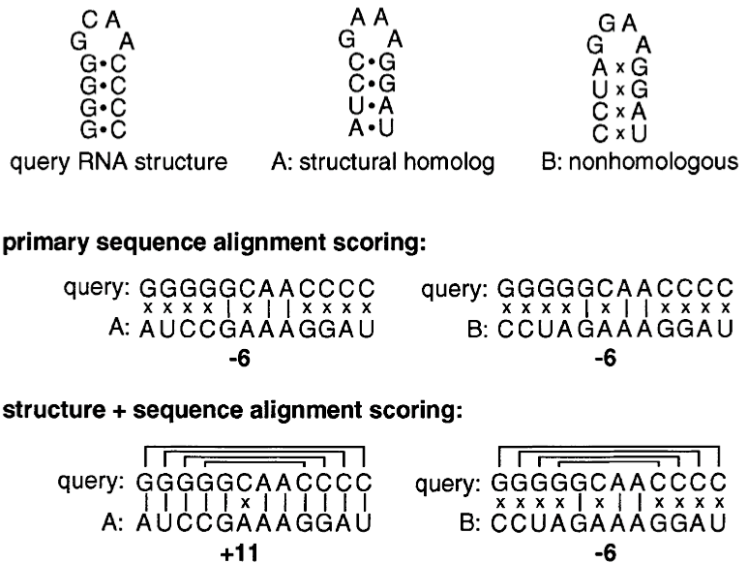


Figure 1.2 Sequence alignment scoring vs structural alignment scoring

COMPUTATIONAL METHODS TO STUDY ncRNAs

Computational approaches to study RNA structure from primary sequence data involve three problems: (1) structure prediction based on single RNA sequence; (2) two or multiple-sequence common consensus structure prediction and alignment; and (3) structural homology searching in databases or genomes.

Section 1 : Structure Prediction

Free Energy Minimization (FEM**)** dynamic programming is the most widely used method for predicting RNA secondary structure from a single sequence. It is motivated by the general concept established by Anfinsen [25] that the three dimensional conformation of a macromolecule forms based upon minimum free energy. A critical step in this line of research was the development of 'Mfold' software for RNA folding by Zuker and Stiegler [26]. This program predicts RNA secondary structure based on energy parameters determined according to the nearest-neighbor model. Most of these thermodynamic parameters are determined experimentally. This method cannot predict pseudoknots; also, Mfold produces multiple structural predictions with similar free energies. Sometimes the results are not completely reliable [27]. The Vienna RNA Package [28] is available at http://www.tbi.univie.ac.at/~ivo/RNA/, RNAfold predicts secondary structures with minimum energy and pair probabilities similar to Mfold; RNAcofold can predict a hybrid structure of two sequences; RNAduplex can predict possible hybridization sites between two sequences; RNAdistance can compare the similarity among secondary structures.

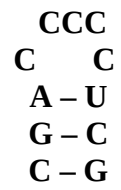
The main alternative algorithm is probabilistic modeling approaches using Stochastic Context Free Grammars (SCFGs) [29], which are capable of capturing the long range, nested, pairwise correlations, such as those induced by base-pairing in non-pseudoknot RNA secondary structures. An advantage of this method is that it is easy to combine with other sources of statistical information to structure prediction, but at the price of greater complexity in memory use and time. This algorithm requires a set of training data, and cannot model or profile the more complex structural motif of pseudoknots. SCFG is a modified Chomsky Grammar [http://en.wikipedia.org/wiki/Chomsky_hierarchy]. Figure 1.3 [William Liu, CS374 lecture

Notes, http://ai.stanford.edu/~serafim/CS374_2004/] describes the process of RNA folding based on the information of sequence cgacccccucg with a SCFG. Let S be the non-terminal node, which can be replaced by either non-terminal or terminal nodes. The structure of the sequence cgacccccucg is modeled as shown on the right of the Figure 1.3 using three rules.

Let the production rules be:

S → aSu | gSc | cSg | ccccc

S → cSg
→ cgScg
→ cgaSucg
→ cgacccccucg



This base-pairing sequence can also be described as: (c(g(a(ccccc)u)c)g)

Figure 1.3 An example of RNA modeling by Stochastic Context Free Grammars

Section 2: Multiple-sequence consensus structure prediction

A structure prediction can be achieved by comparative sequence analysis and structure conservation analysis. The dynamic programming algorithm formulated by Sankoff [30] can produce a sequence alignment and minimize free energy folding simultaneously. This method involves finding the optimal alignments with minimal cost and folding the common consensus structure with minimum free energy; unfortunately, the method requires too much memory and time to be practical. Modifications of this idea are widely applied in the several more recent algorithms discussed later.

The Covariance Model (CM) developed by Eddy and Durbin [31] employs a probabilistic approach to find the common consensus structure. It describes both the RNA secondary structure and the primary sequence consensus, and is the first optimal global algorithm for RNA secondary structure prediction based on pairwise covariations. The covariance model is

directly built from either aligned or unaligned RNA input sequences. The probabilities of insertions, deletions, and mismatches are calculated according to observed RNA sequences. An initial model is constructed based on the input sequences, then a new model is built by iteratively re-estimating the probabilities of emission and state transition based on an Expectation Maximization (EM) algorithm until the parameters converge. This cycle is repeated until both the model structure and its parameters are not significantly changing; the optimal global structure can be found by dynamic programming algorithm. Limitations of this method are that it can't detect non-pairwise interactions (base triples) or non-nested pairs (pseudoknots) structure.

2.1 Algorithms for finding the common secondary structure between two RNA sequences

- **Dynalign** [32] is an algorithm to predict a common secondary structure with minimum free energy for two RNA sequences, it doesn't requires any identity in the sequences. There are two parameter variants involved, gap penalty and a set of thermodynamic parameters. Dynalign doesn't predict pseudoknots and is limited to sequence length less than 300 nucleotides.
- **The Foldalign** [33] algorithm combines local alignment and maximum number of base-pairs together to predict the structure for two sequences. Foldalign optimizes the number of base pairs rather than minimizes free energy. Multibranch loop structures are not allowed to form in Foldalign.
- **CARNAC** [34] is another algorithm for pairwise folding of two unaligned RNA sequences. The program takes four steps to find the structure common to two RNA sequences: (1) searching for the best candidate stems; (2) finding the regions of high similarity between the two sequences, called *anchor points*; (3) performing a pairwise selection of stems, taking into account the information of anchor points and

covariations (4) constructing the common structure based on energy minimization from the set of pre-selected stems.

- **Pair HMMs on Tree Structures (PHMMTSs):** [35] Pair Hidden Markov Models can be used to find a structural pairwise alignment between an unfolded RNA sequence and a RNA sequence of known secondary structure. An RNA secondary structure can be represented by a tree, thus, a pair HMM on tree structures (PHMMTSs) is applied to find a secondary structural alignment between the unfolded RNA sequence and the tree of the known RNA secondary structure.

2.2 Algorithms for searching for the common consensus structure among multiple RNA sequences

- **MSARi** implements a distribution-mixture approach to detect conserved common stems. It is based on computing the statistical significance of short, contiguous potential secondary structure base-paired regions that are conserved between candidate orthologs and allows for small variations between alignments of orthologous base pairs [36] .
- **RNAalifold** is designed for constructing the consensus structure among aligned sequences; it considers both thermodynamic stability and sequence covariation, and it also introduces a base-pairing probability matrix for RNA folding [37].

2.3 Algorithms for predicting and searching for pseudoknots

RNA pseudoknots are functionally important in several known RNAs. For example, RNA pseudoknots are conserved in ribosomal RNAs, the catalytic core of group I introns, RNase P RNAs and telomeras RNAs [38].

- **PCSG (** parallel communicating system grammar) is an algorithm to model RNA

structures including pseudoknots [39]. This approach can automatically generate a pseudoknot prediction algorithm for each specified pseudoknot structure model. A 5 × 5 probability matrix that describes the probability distribution of all possible base pairs among the four nucleotides and gaps, as well as probabilities for the production rules, are needed.

- **PKNOT** is an algorithm (modified SCFG) for predicting optimal RNA secondary structure including pseudoknots [40]. It is built based on the standard RNA thermodynamic model with an augmented set of estimated pseudoknot weighting parameters and coaxial stacking energy. This method can't analyze sequences larger than 130 nucleotides, as the computational time needed is great.
- **The ILM (** iterative loop matching) algorithm combines thermodynamic stability and mutual-information scores to produce a secondary structure including pseudoknots [41] This algorithm can be applied on both multiple aligned sequences and single sequence; the length of each input sequence can be as long as 2000 nucleotides.

2.4 Other methods:

Some other algorithms were also developed for searching for specific classes of ncRNAs, such as tRNAscan-SE, snoscan and snoGPS for detecting tRNAs, methylation-guide C/D box snoRNAs and pseudouridylation-guide H/HAC box snoRNAs [42]. The microinspector program is used to identify the miRNA potential binding sites [43]. A simple hidden Markov model with two states (''RNA'' and ''background genome'') is applied and ncRNAs identified by screening for GC-rich regions in the AT-rich *Methanococcus jannaschii* and *Pyrococcus furiosus* genomes [44].In addition, some other methods that have been successfully applied to identify ncRNAs are summarized in the Table 1.2 [24].

Table 1.2 General Purpose algorithms for RNA Motif Detection

Program	Comparative or single organism	Description
Approaches which search for instances of a motif		
ERPIN	comparative	Input is a sequence alignment with consensus structure. For each helix and single strand a log-odds-score profile is defined which describes the motif.
PATSearch	single	Motif is defined by a language inspired by regular expressions.
fragrep	single	Detects patterns consisting of approximately matched gapless blocks with constrained inter-block distances.
Palingol	single	A constraint programming language particularly adapted for secondary structures. Allows both sequence and structure patterns, including pseudo-knots.
RNAMotif	comparative	Description of structural motif in terms of helices and sequence patterns. Putative hits are ranked according to user defined rules.
infernal	comparative	Toolkit for constructing covariance models and finding new members of a family. Input is a multiple alignment with structural annotation. With SCFGs a consensus model of RNA structure shared by these sequences is defined
Rsearch	single	Input is a single RNA sequence and its structural information. Rsearch is a local alignment algorithm which considers structural and sequence constraints. A base pair and single nucleotide substitution matrix for RNAs

		(RIBO-SUM) defines alignment scores.
FastR	single	Like a pairwise alignment algorithm that addresses structural and sequence conservation. Running time is highly decreased by preprocessing the target sequences. Only those targets sharing similar structural features with the query RNA are aligned.
Approaches which search for motifs from scratch		
SLASH	comparative	Inputs are unaligned sequences. foldalign defines highest scoring local alignments of these sequences according to sequence and structure constraints. COVE creates a SCFG model from those local alignments and does database searches.
RNAProfile	comparative	Input is a set of unaligned sequences. Motif is defined by the number of single hairpins it may contain. Greedy heuristic to find sequences in the input set which share a common motif with defined number of hairpins.
GPRM	comparative	Genetic programming approach to find structural RNA motifs that discriminate a set of input sequences from a set of randomized sequences
HyPa and HyPaLib	single	A search engine and pattern library for hybrid patterns", consisting of sequence and structure elements. The language also includes thermodynamic constraints. Currently, however, HyPaLib contains only some 60 patterns.

Section 3: Searching for ncRNA structures in genomes

- **QRNA** [45] is used for searching ncRNAs given two aligned sequences. The key idea is to test the pattern of substitutions observed in the pairwise alignment of two homologous sequences using three different probabilistic models: a pair HMM for a protein-coding RNA, a pair SCFG for non-coding RNAs, a pair HMM for non-transcribed DNA.
- **RNAz** [46] classifies multiple sequence alignments as ncRNA sequence or non-ncRNA based on two components: Z-score, which is the measure of RNA secondary structure thermodynamic stability, and the structure conservation index (SCI), a measure of the conservation of secondary structure constructed from the input multiple alignment. The alignment is classified as ncRNA or not by a support vector machine (SVM). It is suitable for large-scale genomic annotation whenever alignments can be obtained, but it requires the input as a structural aligned sequences protein-coding RNA, a pair SCFG for non-coding RNAs, and a pair HMM for non-transcribed DNA.
- **The Tree-Decomposition model** developed by Yinglei Song and Chunmei Liu is a novel RNA profile model, which can search for complex structures including those with pseudoknot structures [47]. A conformational graph of the consensus structure of a RNA family is specified based on a set of structurally aligned training data set. Whether there is a significant hit is determined by the Z-score, which is calculated from the log odds ratio of structural alignment scores to random sequences with the same base composition. This algorithm has higher sensitivity and specificity, and is much faster than the Covariance Model.

CHAPTER 2

COMPARATIVE BIOINFORMATIVE OF NON-CODING RNA ROX1 AND ROX2 IN DROSOPHILA

INTRODUCTION

In some mammals and flies, dosage compensation is essential for males to equalize the expression of X-linked genes between two sexes [48]. In Drosophila, the male-specific lethal complex (MSL) , which can significantly increase histone H4 acetylation at lysine 16 and cause chromatin remodeling [49] , spreads along the X-chromosome to up-regulate transcription rate of X-linked genes about 2-fold [50]. rox1 and rox2 are non-coding genes discovered in male *Drosophila melanogaster* [51]; both of them are located on the X-chromosome with major transcriptional products 3742 bps [52] and ~581 bps [53], respectively. It is believed that un-translated rox1 RNA and un-translated rox2 RNA provide a nucleation site for the MSL complex to remodel the male's x-chromosome by covalently modifying the histone H4 [49]. The rox RNAs join the complex at their synthesis sites in cis [54]. Deletions of either rox1 or rox2 RNA produce viable males, but the double deletion leads to a 95% reduction in viability [55, 56]. Both rox1 and rox2 RNAs have several alternative transcribed products by using multiple 5' and 3' splicing sites between two exons [57].

rox RNAs are highly unstable unless they are co-localized with MSL proteins [58] . Although rox1 and rox2 lack similarity in their primary sequences, they have redundant functions in dosage compensation. One possible explanation is that the interaction between the rox RNAs and MSL complex is not due to simple binding of MSL proteins to the rox RNA sequences; it may involve some more complicated unknown mechanism. It is possible that there is a core rox RNA structural element shared by rox1 and rox2, which is important in association with MSL

complex. Some studies suggest nascent transcripts of rox genes play an important role in initializing the assembly of MSL complex to the X-chromosome [59].

Stuckenholz et al [60] discovered that deletions of 10% of the rox1 RNA gene still can have rox1 RNA function at nearly normal activity level. They also found that ~900 nts near the 5'-end of rox1 RNA is very important to its function; if this region was deleted, rox1 RNA would completely lose its function. They found a stem-loop structure within the region of ~600 nts near the 3'-end of the rox1 RNA, this structure is essential to dosage compensation, and showed that mutations disrupting the stem-loop caused defects in its location and processing. Figure 2.1 describes the series of deletions they created in the rox1 RNA gene.

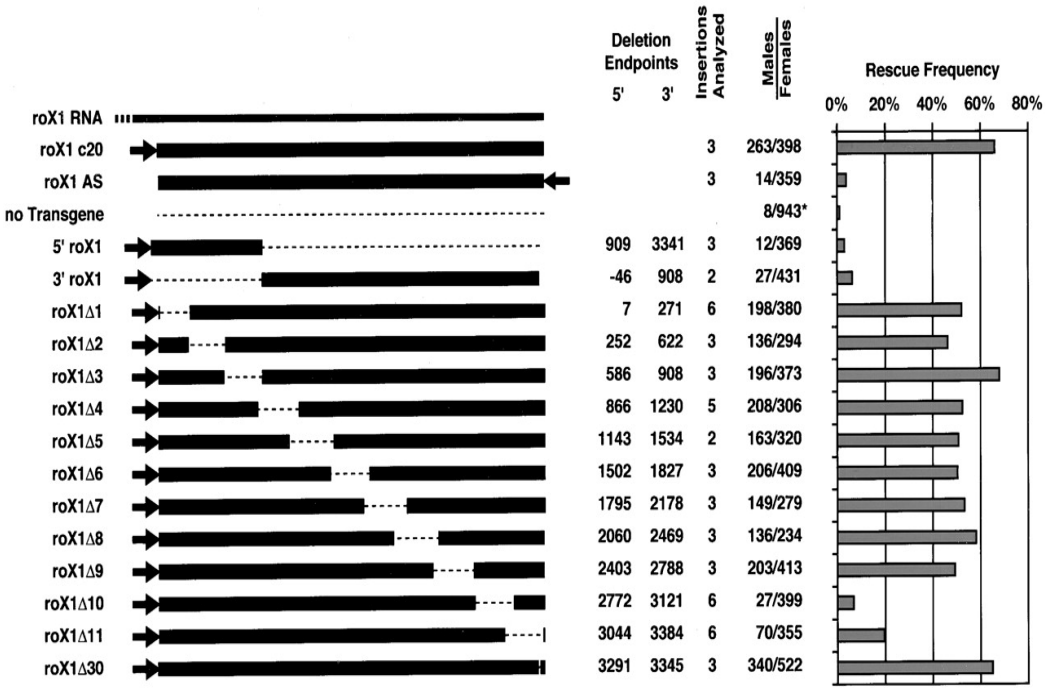


Figure 2.1 Rescue by rox1 deletion constructs. (Overview of the 5' and the 3' rox1 constructs and deletions rox1 1–rox1 30 (Stuckenholz, et al., 2003). Where rox1 RNA gene is located on the X- chromosome of the position from 3752978 ~ 3756719; while rox1 RNA c20 is

324 nts shorter than rox1 RNA from the 5'- end and located on the X- chromosome of the position from 3756442 ~ 3756719. The rescue frequency was defined as the ratio of males to their respective sisters when roX1$^-$ roX2$^-$ females were crossed to males carrying a *y*+ X-chromosome and a roX1 transgene balanced by either CyO or TM3, depending on whether the insertion of the transgene was on the second or third chromosome.(For details see original paper at http://www.genetics.org/cgi/reprint/164/3/1003 and supplements at http://www.genetics.org/supplemental/).

It is a challenge to investigate the causes of redundant function of rox1 and rox2 RNAs. Since their primary sequences are not homologous to each other, traditional comparative analysis of sequence alignment doesn't work well. Computational approaches have been successfully used in classifying RNAs with the same or similar functions, where they do not have similarities in their primary sequences, but they share some common conserved secondary structures. The RNAz [46] and RNAalifold programs [37] are designed for constructing the consensus structure among aligned sequences. They consider both thermodynamic stability and sequence covariation

The aim of this study is to explore the possible mechanism of functional redundancy within the rox1 RNA gene, to see if there is a similar structure shared by rox1 and rox2 RNAs, using a bioinformatic approach. As part of this, we identify new members of the rox RNA gene family in the other *Drosophila* species, *simulans, erecta, yakuba,* and *sechellia*.

MATERIALS AND METHODS

- **Blast search for homologs of rox1 and rox2 genes in *Drosophila* other species**

We used rox1(gi:1835653) with a sequence length of 3742 nts and rox2 (gi:1835654) with sequence a length of 1293 nts (downloaded from http:// www.ncbi.nlm.nih.gov) as query

sequences to perform blast searches against Drosophila and other sequences in flybase (at http:// flybase.bio.indiana.edu). The parameters used were the default parameters.

- **Multiple sequence alignments**

We selected candidate hits with p-value < 0.02 from *simulans, erecta, yakuba*, and *sechellia* species, then used these hit sequences as inputs to perform multiple sequence alignments by CLUSTAL W (1.83) program (at http://www.genebee.msu.ru/clustal/advanced.html). Based on the initial results of alignments, we expanded each hit region in both directions to get a longer conserved alignment , then checked the alignments with the CLUSTAL W (1.83) program, identifying highly conserved homologs of rox1 and rox2 gene. The default parameters were used in the CLUSTAL W (1.83) program.

- **RNA Structure Prediction**

We selected candidate conserved 3'end sequences of rox1 and rox2, performed a CLUSTAL W (1.83) alignment, then used the best alignment region (~200nts) as an input file to run the RNAalifold [37] program (http://rna.tbi.univie.ac.at/cgi-bin/alifold.cgi). The parameters used were the default parameters.

RESULTS

1. **The functional redundancy within rox1 RNA is not due to some sequence similarity within rox1 RNA**

According to Stuckenholz et al's (2003) report that deletions of 10% of the rox1 RNA gene still can keep rox1 RNA function at nearly normal activity level and ~900 nts near the 5'-end of rox1 RNA is very important to its function, we looked for the possible reasons of functional redundancy within rox1 RNA near 5'-end. We subdivided the 5' end of alternative

rox1 (gi:12657620) gene product with length about 900 nts into three sub-regions that correspond to Stuckenholz.et al's experimental analyses.

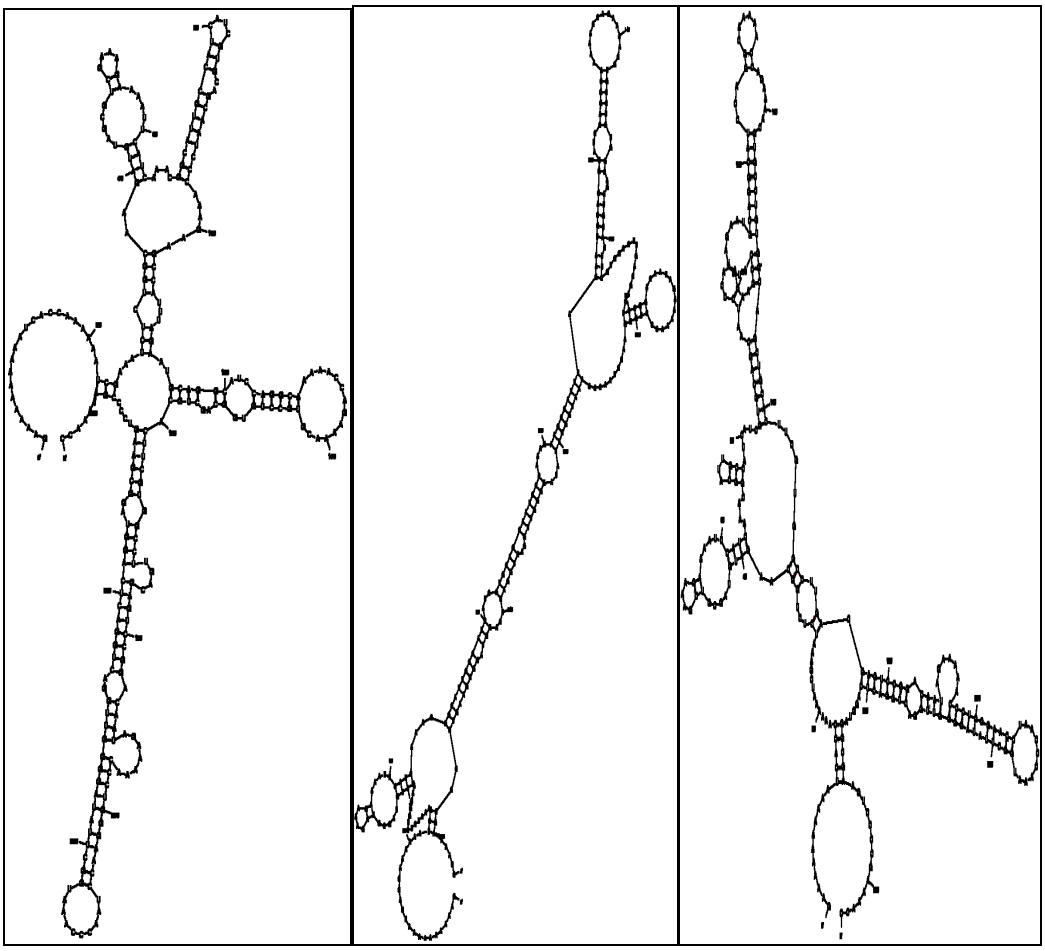
Sub-sequence-1: 7-271 (264 bps from 5'end of RNA)

Sub-sequence-2: 252-622 (377 bps from 5'end of RNA)

Sub-sequence-3: 586-908 (422 bps from 5'end of RNA)

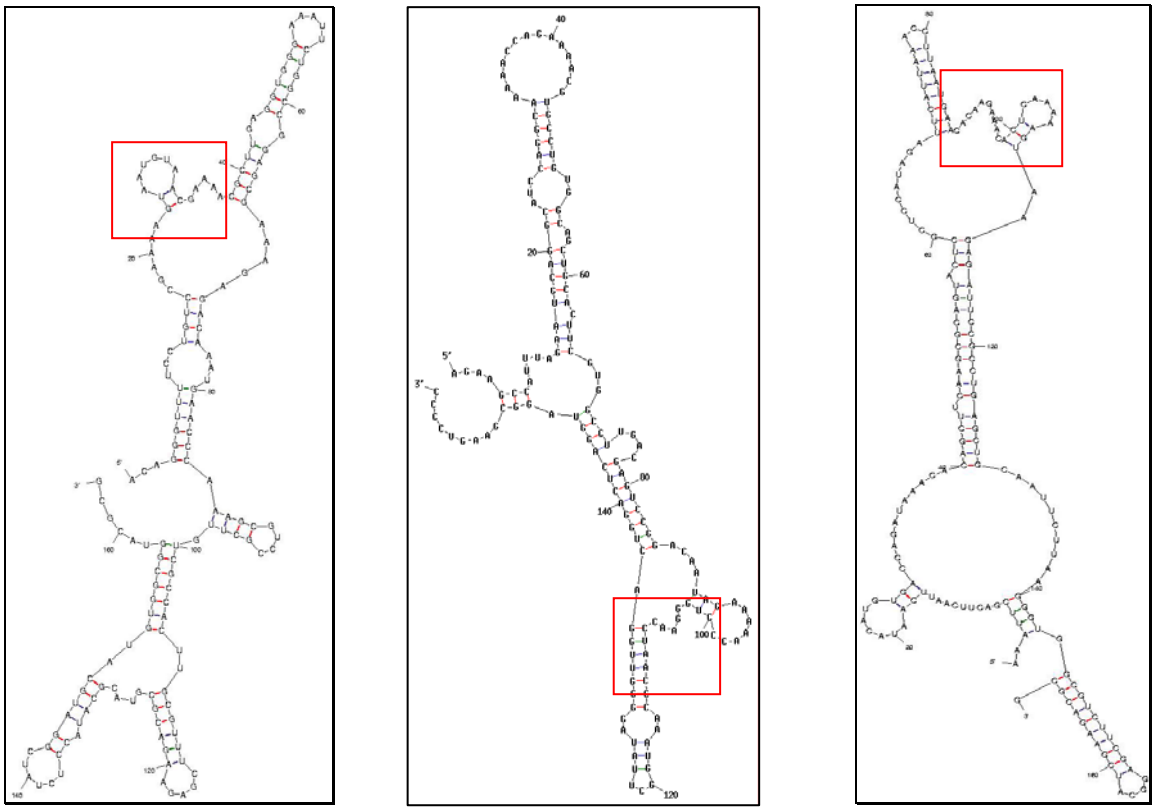
First, to test whether there is a similar sequence or pattern existing among these three sub-regions, two methods were applied: Pairwise Blast (bl2seq) and Multiple Sequence Alignment CLUSTAL W (1.83). All the parameters used in the programs are default parameters. Both programs give the same result that there is no significant sequential homology within the 5'-end of rox1.

Second, we looked for some possible common secondary structures shared by these three sub-regions. We ran the Mfold program (http://bioweb.pasteur.fr/seqanal/interfaces/mfold-simple.html) for each of the three sub-sequences to investigate the possible functional structures shared by them. All the parameters used in the programs were default parameters. Most of the time, the Mfold program will produce multiple structures for a given sequence. For example, Figure 2.2 displays 3 of the 8 candidate structures produced by Mfold for the sequence rox1-sub1 (7-271), the energy ranged from - 69.30 Kcal/mol to -72.60 Kcal/mol. We can see the energy of the two structures in panel B and panel C in Figure 2.2 are close to each other, but their structures were quite different. I picked the most similar structures among these three rox1-sub sequences after removing overlapping parts, and show the results in Figure 2.3. The red box regions indicate candidate common structure.



A dG = -72.60 Kal /mol B dG = -69.50 Kal /mol C dG = -69.30 Kal /mol

Figure 2.2 Predicted structures of rox1-sub1 (7-271) by Mfold



rox1-sub1: 7-271

dG = -48.16 Kal/mol

rox1-sub2: 252-622

dG = -43.1 Kal/mol

rox1-sub3: 586-908

dG = -32.38 Kal/mol

Figure 2.3 Predicted structures of three sub-regions of rox1 RNA from Mfold.

The boxes indicate similar structures

For the next step, we performed a similar experiment by running the RNAalifold program (http://rna.tbi.univie.ac.at/cgi-bin/alifold.cgi) and RNAz algorithm to check if the predicted common structure shared by these three regions of rox1 RNA is consistent with the results from Mfold program. We used the default parameters. The RNAalifold program produces one common structure for a set of aligned sequences. Figure 2.4 is the best alignment of these three sub-regions from CLUSTAL W (1.83) obtained by removing overlapping parts among these three regions. We used these aligned sequences as input sequence to run the RNAalifold program and RNAz algorithm, and obtained the predicted structure shared by these three regions

(rox1-sub1: 7-271 ; rox1-sub2: 252-622 and rox1-sub3: 586-908) as displayed in the Figure 2.5. The red box region indicates a potential common structure shared by these three regions. It seems that both Mfold and RNAalifold programs produced a similar stem-loop structure.

```
                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    5         15         25         35         45         55         65         75
rox1sub1         ACAGGGTTTT CCTGTCCGA- -AAAGTAATG TAACGAAAAC -GCTTGAGGT GGG------- ---------A AATTC-T--G
rox1sub2         AGAAGCCATT TAGA----AT GCAGGCATCC AGGCAAAAAC --CAGAAAAC GTGCCTGTGG CAGCTGCAC- --TTCGT--G
rox1sub3         AAATTCGACT TCAATTCAAT ACATGTGACC AGATAAACAC AGCTTCAAGC GCA---GTAC TCGCTCCATA GATTCATTAA
Clustal Consens  * *      *         *    * *           ** **   *   *    *                       *** *

                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                    85         95        105        115        125        135        145        155
rox1sub1         GCCGGAGCGA A----AGAGA CAA--ATGAA CCCAAAGCGT CCGCTTGTCG CCACTTGC-G TTTCGAGAAG ACGCGTACGC
rox1sub2         GCCTTGACGA GTCCGGACAA TAG--AAAAA CCCTGGGAAC CTAAC----- CCAAATG--G CTTATAGGGG TTGGAC-TGG
rox1sub3         ACCTTAATGA ACACAAGAAA CACTCAAAAA GTAAGAGATT CCGCC----- TGAGCTGCAA TTCTTAAGGG TGGCGT-C--
Clustal Consens   **     **          *  *   *  **       *    *           *  **     *   *   *   *

                 ....|....| ....|....| ....|....| ..
                   165        175        185
rox1sub1         ATACCTCTAT CGGATGCATG TGGCGGTACG CG
rox1sub2         AC-------T CAGGTAGGCG AAG----TCC CC
rox1sub3         -T-------T CGAGGCATCG AAG----ACG CG
Clustal Consens           * *        *   *     *  *
```

Figure 2.4 Alignment of three sub-regions of rox1(rox1-sub1: 7-271; rox1-sub2: 252-622; rox1-sub3: 586-908) from CLUSTAL W (1.83).

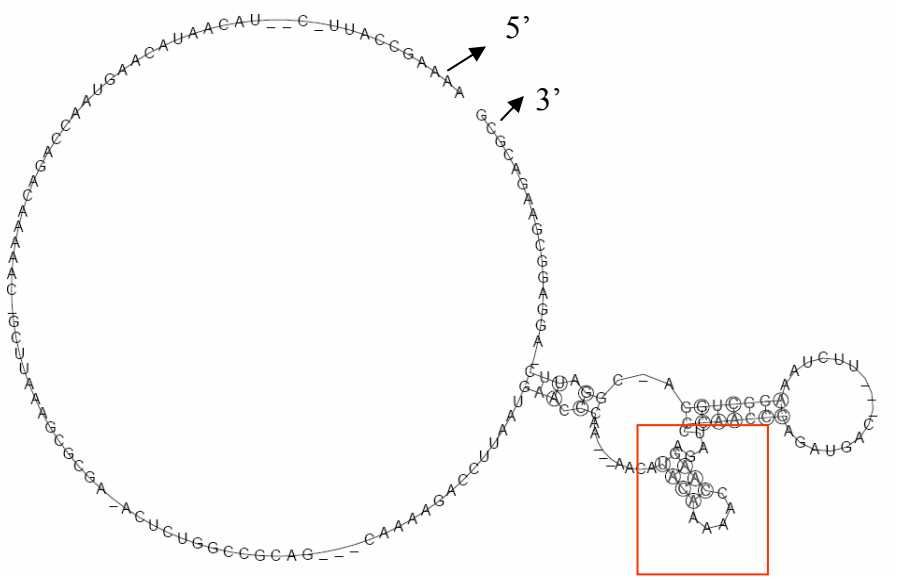


Figure 2.5 Common structure shared within 5’-end of rox1 RNA. (Partial output from RNAz: Sequences: 3 (number of input sequences); Columns: 174 (length of input sequences, including gaps); Consensus MFE: -2.38 Energy; Mean pairwise identity: 38.1 (average of identity between two sequences); Structure conservation index: 0.05; Mean z-score: -1.53 ; Mean single sequence MFE(average of minimum free energy for each sequence): -49.45; SVM RNA-class probability: 0.282826 (predicted probability of functional structure by RNAz) where the nucleotides in the circles mean non-conserved nucleotides.The box indicates potential common conserved structure.

2. Identify candidate members of rox RNAs family in other Drosophila species

It has been widely reported that non-coding rox1 and rox2 RNA are involved in dosage compensation in male *Drosophila melanogaster,* but it hasn’t been reported if rox1 and rox2 RNAs also exist in other *Drosophila* species. Here, we applied comparative bioinformatic methods to identify putative rox1 and rox2 RNAs in other *Drosophila* species as a prelude to consensus structure prediction.

We did a blast search for new members of the rox1 and rox2 RNA family in other *Drosophila* species with query sequences of *Drosophila melanogaster* rox1(gi:1835653) and rox2 (gi:1835654) in flybase (at http://flybase.net/blast/); we found some highly conserved homologues of rox1 and rox2 RNAs, respectively, and estimated the length of candidate rox RNAs according to the extent of conservation (see Figure 2.6 and Figure 2.7). The boxes of sub1, sub2, sub3, sub4, sub5 and sub6 indicate the conserved regions of rox1 RNA; the average of the mean pairwise identity of these regions is 0.89 among *Drosophila* species. Similarly, the sub1 and sub2 regions of rox2 RNAs have mean pairwise identity of 0.91 and 0.93 respectively. The solid lines show the regions that have a lot of sequence variation and the dashed lines show the rox1 RNA sequence with full length of 3742 nts and rox2 RNA sequence with length of 581 nts in *Drosophila melangoster*.

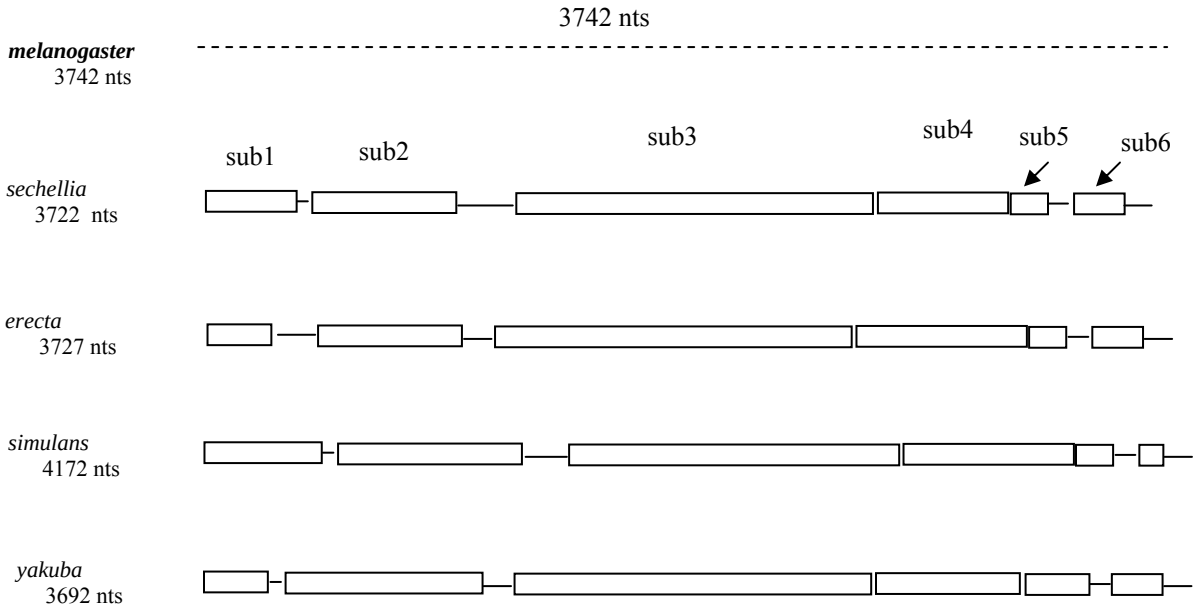


Figure 2.6 Candidate rox1 RNAs in *Drosophila simulans, yakuba, sechelli,* and *erecta*

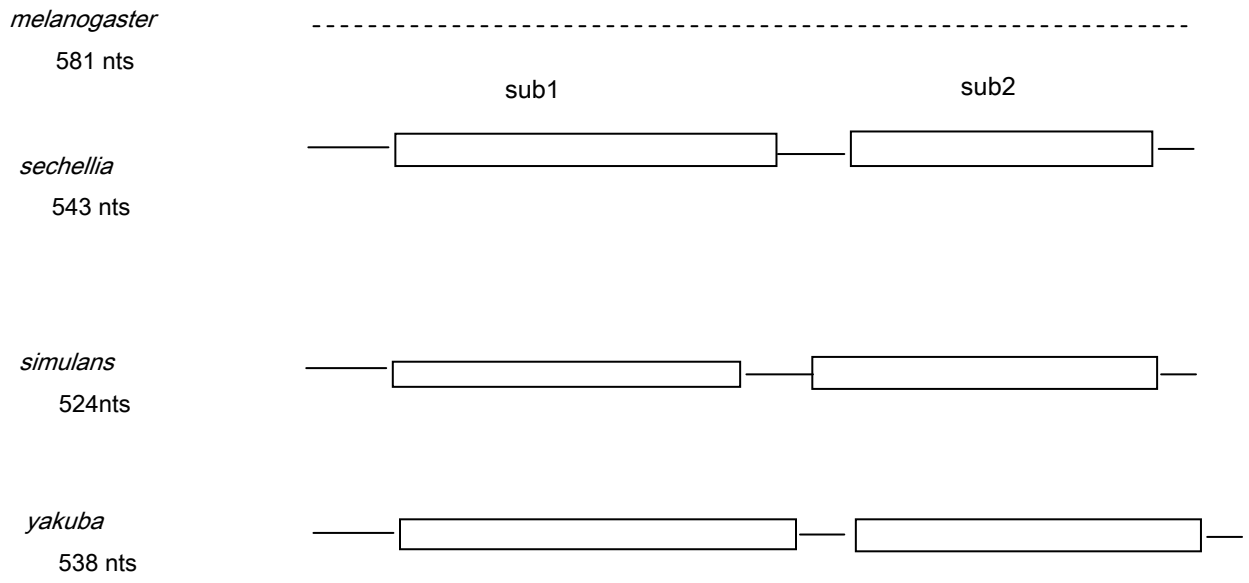


Figure 2.7 Candidate rox2 RNAs in *Drosophila simulans, yakuba, sechellia*

3. There are some common conserved predicted structures of rox RNAs shared among different Drosophila species

We picked the sub2 region of rox1 RNA as described in Figure 2.6. The alignment was performed by the program clustalw(1.83) (Figure 2.8).We used this set of aligned sequences to run the RNAalifold and RNAz programs, and obtained a predicted common consensus structure for this region.(Figure 2.9). Since the RNAz algorithm works better with input sequence length less than 200 nts, we shortened the sequences by removing 130 nts from 3'-end and removing 20 nts from 5'-end from the alignment in the Figure 2.8, then re-ran Clustalw (1.83) to obtain the alignment described in Figure 2.10. There is no significant difference between these two sets of alignments except the length of aligned sequences. We used the alignment showed in Figure 2.10 to run RNAalifold program and obtained the predicted structure shown in Figure 2.11.

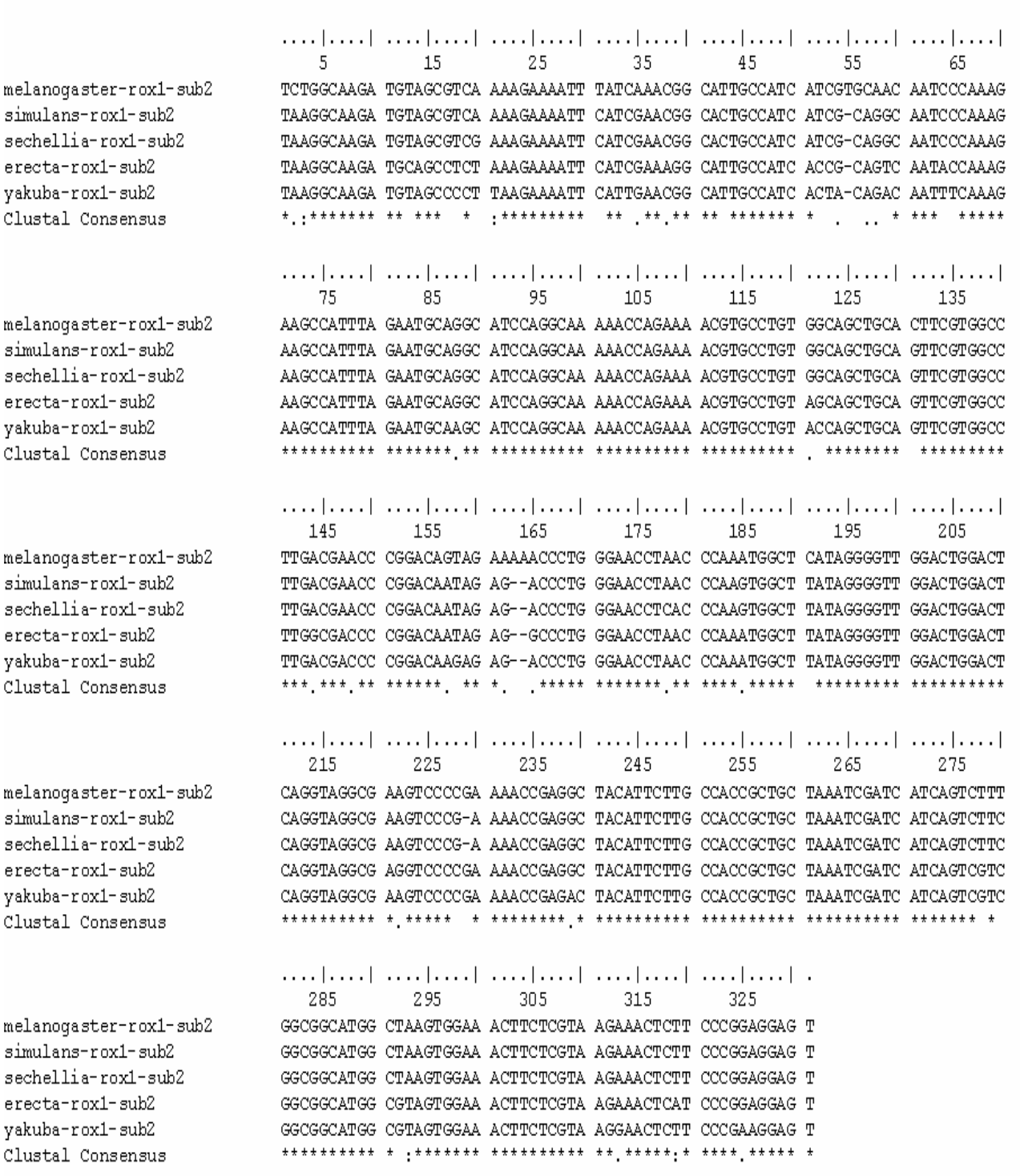


Figure 2.8 Sequence alignments of rox1-sub2 region among *Drosophila Melanogaster, simulans, sechellia, erecta, yakuba*

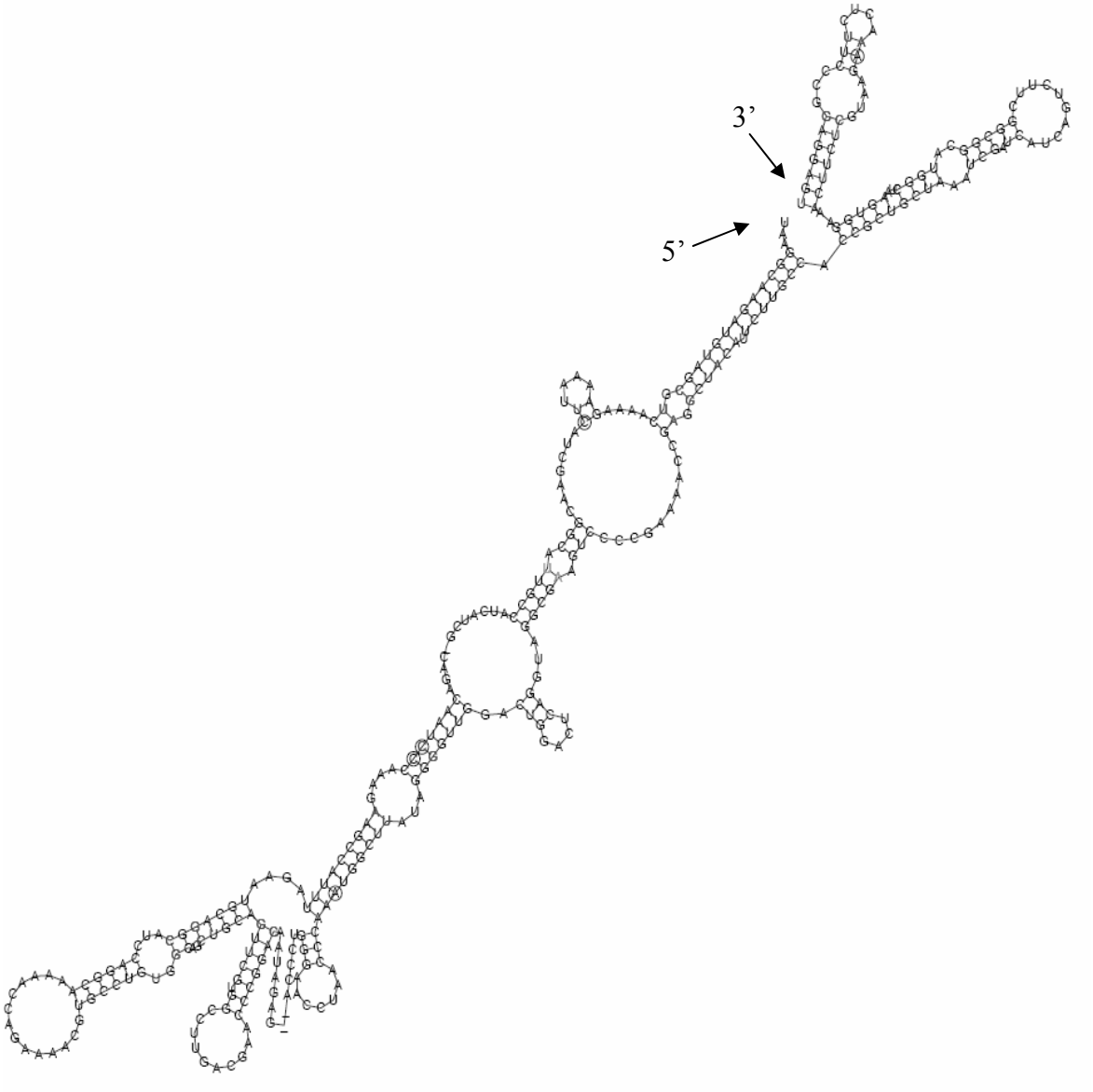


Figure 2.9 Predicted common consensus structure for rox1-sub2 region among *Drosophila melanogaster, simulans, sechellia, erecta, yakuba* from the RNAz.

(Partial RNAz output: Sequences: 5; Columns: 331; Mean pairwise identity: 93.16; Mean single sequence MFE: -106.46; Consensus MFE: -90.42 Energy; Mean z-score: -1.33; SVM RNA-class probability: 0.6599)

```
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                   5         15         25         35         45         55         65
melanogaster-rox1-sub1-shorten  AAAGAAAATT TATCAAACGG CATTGCCATC ATCGTGCAAC AATCCCAAAG AAGCCATTTA GAATGCAGGC
simulans-rox1-sub1-shorten      AAAGAAAATT CATCGAACGG CACTGCCATC ATCG-CAGGC AATCCCAAAG AAGCCATTTA GAATGCAGGC
sechellia-rox1-sub1-shorten     AAAGAAAATT CATCGAACGG CACTGCCATC ATCG-CAGGC AATCCCAAAG AAGCCATTTA GAATGCAGGC
erecta-rox1-sub1-shorten        AAAGAAAATT CATCGAAAGG CATTGCCATC ACCG-CAGTC AATACCAAAG AAGCCATTTA GAATGCAGGC
yakuba-rox1-sub1-shorten        TAAGAAAATT CATTGAACGG CATTGCCATC ACTA-CAGAC AATTTCAAAG AAGCCATTTA GAATGCAAGC
Clustal Consensus               :*********  ** .**.** ** ******* *  .  .. * ***  ***** ********** *******.**

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                   75         85         95        105        115        125        135
melanogaster-rox1-sub1-shorten  ATCCAGGCAA AAACCAGAAA ACGTGCCTGT GGCAGCTGCA CTTCGTGGCC TTGACGAACC CGGACAGTAG
simulans-rox1-sub1-shorten      ATCCAGGCAA AAACCAGAAA ACGTGCCTGT GGCAGCTGCA GTTCGTGGCC TTGACGAACC CGGACAATAG
sechellia-rox1-sub1-shorten     ATCCAGGCAA AAACCAGAAA ACGTGCCTGT GGCAGCTGCA GTTCGTGGCC TTGACGAACC CGGACAATAG
erecta-rox1-sub1-shorten        ATCCAGGCAA AAACCAGAAA ACGTGCCTGT AGCAGCTGCA GTTCGTGGCC TTGGCGACCC CGGACAATAG
yakuba-rox1-sub1-shorten        ATCCAGGCAA AAACCAGAAA ACGTGCCTGT ACCAGCTGCA GTTCGTGGCC TTGACGACCC CGGACAAGAG
Clustal Consensus               ********** ********** ********** . ********  ********* ***.***.** ******. **

                                ....|....| ....|....| ....|....| ....|....| .
                                  145        155        165        175
melanogaster-rox1-sub1-shorten  AAAAACCCTG GGAACCTAAC CCAAATGGCT CATAGGGGTT G
simulans-rox1-sub1-shorten      AG--ACCCTG GGAACCTAAC CCAAGTGGCT TATAGGGGTT G
sechellia-rox1-sub1-shorten     AG--ACCCTG GGAACCTCAC CCAAGTGGCT TATAGGGGTT G
erecta-rox1-sub1-shorten        AG--GCCCTG GGAACCTAAC CCAAATGGCT TATAGGGGTT G
yakuba-rox1-sub1-shorten        AG--ACCCTG GGAACCTAAC CCAAATGGCT TATAGGGGTT G
Clustal Consensus               *.  .***** *******.** ****.*****  ********* *
```

Figure 2.10 Shortened sequence alignments of rox1-sub2 region among *Drosophila melanogaster, simulans, sechellia, erecta, yakuba*

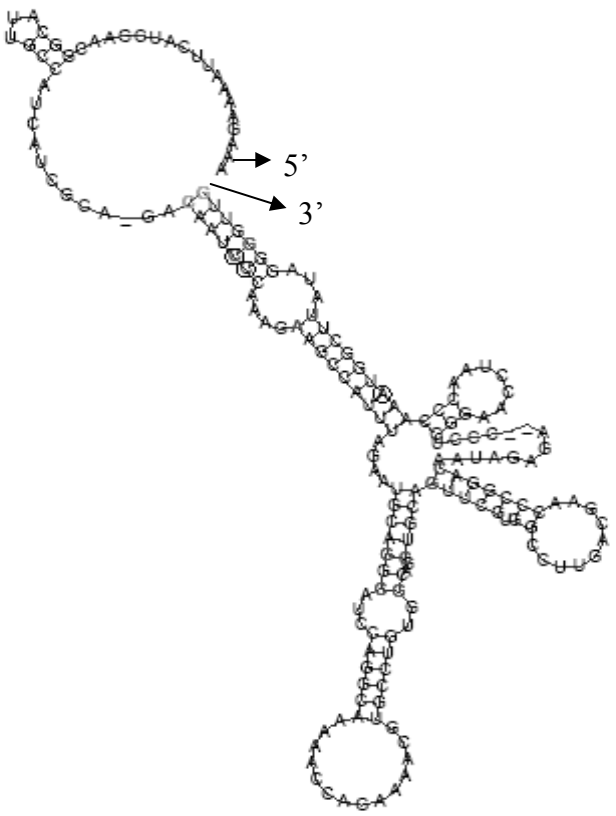


Figure 2.11 Predicted common consensus structure for shortened rox1-sub2 region among *Drosophila melanogaster, simulans, sechellia, erecta, yakuba* from the RNAz.
(Partial RNAz output: Sequences: 5;Columns: 181; Mean pairwise identity: 92.35;
Mean single sequence MFE: -49.39; Consensus MFE: -44.06; Mean z-score: -0.98;
SVM RNA-class probability: 0.267851)

4. We found there is potential small common consensus structure near the 3’end shared by both rox1 and rox2 RNAs

Previous studies indicated that nascent synthesized rox RNAs play an important role in initializing the MSL complex assembly and binding to the X-chromosome, so we proposed that there could be a common structure shared by rox1 and rox2 RNAs. We aligned rox1 and rox2 RNAs near the 5’-end (~200 nts) among several *Drosophila* species by running Clustalw (1.83) program, but we failed to find significant homologues in their primary sequences. In addition,

the 3'-ends of rox RNAs are also very important to their function, so we looked for a potential common structure for rox1 and rox2 RNAs around this region. Similarly, we aligned the 3'-end of rox1 RNA and rox2 RNA (~100 nts) among *Drosophila* species using the Clustalw (1.83) program. We obtained two sets of alignments (see Figure 2.12 and Figure 2.13), the only difference between these two sets of sequences is that the second set has three additional sequences of rox1- *erecta*, rox1- *sechellia* and rox1- *simulans*; then we used these two sets of aligned sequences to run RNAalifold and RNAz programs, obtaining a common structure predicted to be shared by both rox1 and rox2 RNAs (Figure 2.14).

Although these two sets of sequences yielded the same structure, their RNAz outputs are quiet different, especially the prediction probabilities were reduced from 0.920797 with 6 sequences to 0.6523 with 9 sequences.

```
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        5         15         25         35         45         55         65
rox1-melanogaster   TAAAAACTTG CTGATCAACG TTCTACGCAG TTCTTAAAAA GATGTTGAAA TGAACACAGC CAAAGCAAGT
rox1-yakuba         TAAAAACTTG CTGATCAACG TTCTACGCAG TTCTTAAAAC ATGTTAAAAA TGAATACAAC CAATGCAAGT
rox2-melanogaster   CAACATTGTA CAAGTCGCAA TGCAAACTGA AGTCTTAAAA GACGTGTAAA ATGTTGCAAA TTAAGCAAAT
rox2-sechellia      CAACATTGTA CAAGTCGCAA TGCAAACTGA AGTCTTAAAA GACGTGTAAA ATGTTGCAAA TTGAGCAAAT
rox2-erecta         CAACATTGTA CAAGTCGCAA TGAAAACTAA AGTCTTAAAA GACGTGTAAA ATGTTGCAAA TTAAGCAAAT
rox2-yakuba         CAACATTGTA CGAGTCGCAA TGAGAACTAA AGTCTTAAAA GACGTGTAAA ATGTTGCAAA TTAAGCAAAT
Clustal Consensus    ** *   *  *   **     *   *          * ***         ***       **       **** *

                    ....|....| ....|....| ....|....| ....|....| .
                        75         85         95        105
rox1-melanogaster   AAAAA----T GTGTGGA-AA CGTTATACGA ATCTTCACCA A
rox1-yakuba         AAAAA----T GTGTGAA-AA CGTTATACGA ATCTTCACCA A
rox2-melanogaster   ATATATGCAT ATATGGGTAA CGTTTTACGC GCCTTAACCA G
rox2-sechellia      ATATATGCAT ATATGGGTAA TGTTTTACGC GCCTTAACCA G
rox2-erecta         ATATATGCAT ACATGGGTAA CGTTTTACGC GCCTTAACCA G
rox2-yakuba         ATATATGCAT ATATGGGTAA CGTTTTACGC GCCTTAACCA G
Clustal Consensus   *        *    **   **  *** ****    *** **** .
```

Figure 2.12 Alignment of the 3'-end of rox1 and rox2 RNAs (rox1-*melanogaster*, rox1-*yakuba;* rox2-*melanogaster*, rox2-*sechellia*, rox2-*erecta,* rox2-*yakuba)*

```
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                        5         15         25         35         45         55         65
melanogaster-rox1   TAAAAACTTG CTGATCAACG TTCTACGCAG TTCTTAAAAA GATGTTGAAA TGAACACAGC CAAAGCAAGT
simulans-rox1       TAAAAACTTG CTGATCAACG TTCTACGCAG TTCTTAAAAA GATGTTGAAA TGAACACAGC CAAAGCAAGT
sechellia-rox1      TAAAAACTTG CTGATCAACG TTCTACGCAG TTCTTAGAAA GATGTTGAAA TGAACACAGC CAAAGCAAGT
erecta-rox1         TAAAAACTTG CTAATCAACG TTCTACGCAG TTCTTAAAAA GATGTTGAAA TGAACACAAC CAATGCAAGT
yakuba-rox1         TAAAAACTTG CTGATCAACG TTCTACGCAG TTCTTAAAAC ATGTTAAAAA TGAATACAAC CAATGCAAGT
melanogaster-rox2   CAACATTGTA CAAGTCGCAA TGCAAACTGA AGTCTTAAAA GACGTGTAAA ATGTTGCAAA TTAAGCAAAT
sechellia-rox2      CAACATTGTA CAAGTCGCAA TGCAAACTGA AGTCTTAAAA GACGTGTAAA ATGTTGCAAA TTGAGCAAAT
erecta-rox2         CAACATTGTA CAAGTCGCAA TGAAAACTAA AGTCTTAAAA GACGTGTAAA ATGTTGCAAA TTAAGCAAAT
yakuba-rox2         CAACATTGTA CGAGTCGCAA TGAGAACTAA AGTCTTAAAA GACGTGTAAA ATGTTGCAAA TTAAGCAAAT
Clustal Consensus    ** *     *             *         *  **     *  ***       *       **** *

                    ....|....| ....|....| ....|....| ....|....| .
                        75         85         95        105
melanogaster-rox1   AAAAA----T GTGTGGA-AA CGTTATACGA ATCTTCACCA A
simulans-rox1       AAAAA----T GTGTGGA-AA CGTTATACGA ATCTTCACCA A
sechellia-rox1      AAAAA----T GTGTGGA-AA CGTTATACGA ATCTTCACCA A
erecta-rox1         AAAAA----T GTGTGGA-AA CGTTATACGA ATCTTCACCA A
yakuba-rox1         AAAAA----T GTGTGAA-AA CGTTATACGA ATCTTCACCA A
melanogaster-rox2   ATATATGCAT ATATGGGTAA CGTTTTACGC GCCTTAACCA G
sechellia-rox2      ATATATGCAT ATATGGGTAA TGTTTTACGC GCCTTAACCA G
erecta-rox2         ATATATGCAT ACATGGGTAA CGTTTTACGC GCCTTAACCA G
yakuba-rox2         ATATATGCAT ATATGGGTAA CGTTTTACGC GCCTTAACCA G
Clustal Consensus   * * *    *    **   **  *** ****    *** ****
```

Figure 2.13 Alignment of 3'-end of rox1 and rox2 RNAs (rox1-*melanogaster*, rox1-*simulans,* rox1-*sechellia,* rox1-*erecta*, rox1-*yakuba;* rox2- *melanogaster*, rox2-*sechellia*, rox2-*erecta,* rox2-*yakuba*)

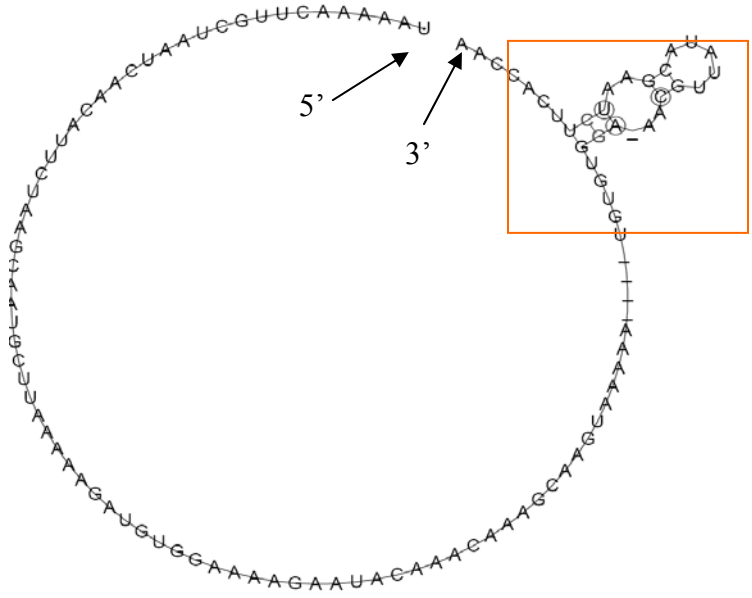


Figure 2.14 Predicted common structure shared by 3’rox1 and rox2 RNAs.
(Partial RNAz output Sequences: 6; Columns: 112; Consensus MFE: -8.80 Energy; Structure conservation index: 0.37; Mean pairwise identity: 73.14; Mean single sequence MFE: -23.96; Mean z-score: -2.34; SVM RNA-class probability: 0.920797.Where the nucleotides in the circles mean non-conserved nucleotides. The box indicates potential common conserved structure)

CHAPTER 3

CONCLUSIONS AND FUTURE STUDIES

In some mammals and flies, dosage compensation is essential for males to equalize the expression of X-linked genes between two sexes [48]. In Drosophila, the male-specific lethal complex (MSL) spreads along the X-chromosome to up-regulate transcription rate of X-linked genes about 2-fold. It is believed that un-translated rox1 RNA and un-translated rox2 RNA provide a nucleation site for the MSL complex to remodel the male's X-chromosome by covalently modifying the terminal tails of histone [49] . The rox RNAs join the complex at their synthesis sites in cis. Simultaneous mutation of both rox genes can cause a striking male-specific reduction in viability.

In Stuckenholz et al's study (2003), they deleted 10% of the rox1 RNA gene and found that rox1 RNA still can function at nearly normal activity level, but the mechanism is unknown. We explored possible reasons for functional redundancy within rox1 RNA in the 900 nts near the 5'-end. The results from Pairwise Blast (bl2seq), Lalign [61] and multiple sequence alignment by CLUSTAL W (1.83) all indicated that there is no significant sequential homology within the 5'-end of rox1 RNA. We can conclude that the functional redundancy within rox1 RNA is not due to some sequence similarity within rox1 RNA. The RNAz program predicted there is a stem loop structure shared by these three sub-regions (see Figure 2.5). The redundant function of the rox1 5'-end ~900 nts may be due to a stem loop structure with weak signal; the predicted probability of the functional structure (SVM RNA-class probability) by RNAz is 0.282826. It might be worth testing the function of this stem loop structure by changing base pairs to disrupt the stem to check if rox1 RNA still can function normally.

rox1 and rox2 RNAs show functional redundancy in dosage compensation although they lack similarity in their primary sequences. This redundant function may be related to a potential core rox RNA structural element shared by rox1 and rox2. Previous studies uncovered the observation that nascent rox RNA transcripts play an important role in initializing the assembly of MSL complex to the X-chromosome [59]. We aligned rox1 and rox2 RNAs near the 5'-end (~200 nts) among several *Drosophila* species, but we failed to find significant homologies in their primary sequence. In addition, the 3'-ends of rox RNAs are also very important to their function, We aligned the 3'-end of rox1 RNA and rox2 RNA (~100 nts) among *Drosophila* species and obtained a predicted common structure (see Figure 2.14), which is shared by both rox1 and rox2 RNAs with high a probability of 0.920797. This common consensus structure (a double internal loop) was produced by the RNAz program. We observed that the number of input sequences had a big effect on the prediction probabilities, such as the SVM RNA-class probabilities were reduced from 0.920797 with 6 sequences to 0.6523 with 9 sequence even though the two sets of alignments are almost the same (see Figure 2.12 and Figure 2.13). Thus, a

limitation of this program is that RNAz works well for alignments with sequence less

than 6.

Although non-coding rox1 and rox2 RNA have been widely studied in *Drosophila melanogaster,* it hasn't been reported if rox1 and rox2 RNAs also exist in other *Drosophila* species. Here, we identified putative rox1 and rox2 RNAs in other *Drosophila* species (see Figure 2.6 and Figure 2.7) . The conservation of nucleotides in their primary sequences is very high among several *Drosophila* species. The average mean pairwise identity of sub1 (256 nts described in Figure 2.7) and sub2 (187 nts described in Figure 2.7) of rox2 RNAs among

Drosophila melanogaster, simulans, yakuba, sechellia is up to 0.92, and 443 nts out of a total 581 nts of the major product of rox2 RNA were very conserved. Similarly, we found that 3613 nts out of a total of 3742 nts of rox1 RNA were conserved and the mean pairwise identity was as high as 0.89 among *Drosophila melanogaster, simulans, yakuba, sechelli and erecta.*

Future experiments should include RT-PCR to confirm our prediction of these candidate RNA members of rox RNA gene family in Drosophila species. The method would be applied as described in Smith et al's paper (2000) [62]. A forward primer for rox2 RNA is 5'-GGTAGCTCGGATGGCCATCGAAAGGGTA-3', and a reverse primer is 5'-GACTGGTTAAGGCGCGTAAAACGTTACC-3'. For the rox1 RNA, we can use following three pairs of primers: forward primers are 5'-GGCTAAGTGGAAACTTCTCGTAAGAAACTC-3'; 5'-CCCAGAAGAAACTGCCACTGC-3'; 5'- CCCAGAAGAAACTGCCACTGC-3', and reverse primers are 5'-TGAATCCCGGGTGGATACGATTGTAG-3'; 5'- AATGTCCCTTTTCGAGCG-3'; and 5'-TCCGCGAGGCTCCAAGTTTCG-3' respectively.

We ran the RNAalifold and RNAz programs for each conserved sub-region of rox RNAs in Figure 2.6 and Figure 2.7, and we obtained the predicted the common conserved structures of rox RNAs shared among different Drosophila species. We only listed the results of rox1-sub2 region here (see Figure 2.9 and Figure 2.11) because although most regions are very conserved, there is still some variation within this region. We also noticed that the SVM RNA-class probability was reduced from 0.6599 to 0.267851 as the same set of alignments shortened (see Figure 2.8 and Figure 2.10); the RNAalifold algorithm is sensitive to the length of the alignments and usually works well with the sequences length less than 200 nts.

It difficult to evaluate the prediction result of RNA secondary structure from various kinds of software because the predicted structure is highly dependent on the length of the RNA sequences and the base composition. The correct prediction rate of the widely applied Zuker's Mfold is still not high since there are not enough experimental thermostable parameters available to build models. It is also difficult to choose among the multiple output predictions. To confirm the results from the RNAalifold program, I also ran a similar program, the RNAz algorithm to predict the common structures shared by several *Drosophila* species. The RNAz algorithms can detect evolutionarily conserved and thermodynamically stable RNA secondary structures from multiple sequence alignments. The results of the RNAz can be affected by several factors, so how to determine a reliable conclusion based on the RNAz output is complex. From the results described above, we can see the SVM-RNA class probability is very sensitive to the number of sequences in an alignment and the length of an alignment. Here will discuss several factors that influence the interpretation of if a set of aligned RNA sequences containing a common consensus structure (see RNAz Manual).

(1) The Structure Conservation Index (SCI) measures the structure conservation, which is the ratio of consensus RNA structure Minimum Free Energy (E_A) to individual single RNA structure Minimum Free Energy ($\bar{E}$) , SCI = $E_A/\bar{E}$. SCI depends on the number of sequences in the alignment and mean pairwise identity. The RNAz works well for an alignment limited to 6 sequences. When the SCI is above or close to the mean pairwise identity, this indicates a strong signal that the structure is a conserved fold. For example, an RNAz output with the mean pairwise identity of 0.95 and SCI of 0.80 doesn't mean the structure is more conserved than a set of alignment with same number of sequences with mean pairwise identity of 0.60 and SCI of 0.75. (2) The z-score of Minimum of Free Energy is not strictly normally distributed, but a z-

score below -3 usually indicates a very stable structure. In the meanwhile, for a given z-score, the higher the mean pairwise identity of the alignment is, the higher false positive will be. An alignment with mean pairwise identity of 0.95 is more likely to obtain a z-score of - 4 by chance than an alignment with mean pairwise identity of 0.70. (3) The SVM RNA-class classification probability is also called the "decision-value", this value is not a P-value since there is no underlying statistical model applied. The RNAz program uses an ad hoc machine learning machine to calculate this classification probability. It is hard to say a prediction with a SVM RNA-class classification probability of 0.95 is more reliable than a prediction with a SVM RNA-class classification probability of 0.90 without considering other factors. The limitations of applying RNAz are that RNAz could produce reliable prediction of the RNA structure for a set of alignment only when: the number of sequences in the alignment around 5 or 6, but no more than 6; the length of sequences is less than 200 nucleotides; the mean pairwise identity of the alignments is around 0.80, but no less than 0.60.

Our prediction of the common structure shared by rox1 and rox2 near 3'-end , a double internal loop structure, is a significantly statistical signal since the z-score is as low as -2.34, mean pairwise identity is 0.73 and SVM RNA-class probability is 0.92 except the SCI is a little low as 0.37. All these features suggest it might be worth doing an experiment to test this structure. We could use the in-line probing assay method [63, 64] to confirm our prediction, where we will label the 5'-end of rox RNAs with ^{32}P and use RNase T_1 to digest the RNAs of *Drosophila* species sequence, then run a gel to resolve the fragments.

My experience with these programs suggests the RNAalifold and RNAz can produce structural predictions that are worth performing lab experiments to test.

REFERENCES

1. Stuckenholz C, Meller VH, Kuroda MI. Functional redundancy within rox1, a noncoding RNA involved in dosage compensation in Drosophila melanogaster. Genetics, 2003, 164 (3):1003-14.

2. Park Y, Mengus G, et al. Sequence-specific targeting of Drosophila rox genes by the MSL dosage compensation complex. Molecular Cell, 2003,11(4):977-86.

3. Rodriguez A, Griffiths-Jones S, Ashurst JL, Bradley A. Identification of mammalian microRNA host genes and transcription units. Genome Research., 2004,14(10A):1902-10.

4. Liang D, Zhou H, Zhang P, Chen YQ, Chen X, Chen CL, Qu LH. A novel gene organization: intronic snoRNA gene clusters from Oryza sativa. Nucleic Acids Research, 2002, 30(14):3262-72.

5. Martinez-Salas E, Fernandez-Miragall O. Picornavirus IRES: Structure Function Relationship. Current Pharmaceutical Design, 2004, 10 (30) : 3757-3767(11)

6. Rogers JT et al. An Iron-responsive Element Type II in the 5'-Untranslated Region of the Alzheimer's Amyloid Precursor Protein Transcript. The Journal of Biological Chemistry, 2002, 277(47):45518-28.

7. Cmejla R, Petrak J, Cmejlova J. A novel iron responsive element in the 3'UTR of human MRCKalpha. Biochemical and Biophysical Research Communications, 2006, 3, 341(1):158-66. pre-published online, 2006, Jan 6.

8. Johnson T.L.; Abelson J. Identification of Mammalian microRNA Host Genes and Transcription Units. Genes and Development, 2001,15(15):1957-1970.

9. Sharin E, Schein A, Mann H, Ben-Asouli Y, Jarrous N. RNase P: role of distinct protein cofactors in tRNA substrate recognition and RNA-based catalysis. Nucleic Acids Research, 2005, 9 , 33(16):5120-32.

10. Leung SS, Koslowsky DJ. RNA editing in Trypanosoma brucei: characterization of gRNA U-tail interactions with partially edited mRNA substrates**.** Nucleic Acids Research, 2001,29(3):703-9.

11. Kiss T. Small nucleolar RNAs: an abundant group of noncoding RNAs with diverse cellular functions. Cell, 2002, 109: 145–148.

12. Nelson P, Kiriakidou M, Sharma A, Maniataki E, Mourelatos Z. The microRNA world: small is mighty. Trends in Biochemical Sciences, 2003, 28(10):534-40.

13. Andrew Fire**.** RNA-triggered gene silencing. Trends in Genetics, 1999, 15 (9) :358–363.

14. Lingner J, Cech TR. Telomerase and chromosome end maintenance. Current Opinion in Genetics & Development. 1998 , 8(2):226-32.

15. Chow JC, Yen Z, Ziesche SM, Brown CJ. Silencing of the mammalian X chromosome. Annual Review of Genomics and Human Genetics**.** 2005, 6 :69-92.

16. PARK, Y., R. L. KELLEY, H. OH, M. I. KURODA, and V. H. MELLER. Extent of chromatin spreading determined by rox RNA recruitment of MSL proteins. Science, 2002, 298:1620-1623.

17. Richard L. Kelley and Mitzi I. Kuroda**.** The Drosophila rox1 RNA gene can overcome silent chromatin by recruiting the male-specific lethal dosage compensation complex. Genetics, 2003, 164 : 565-574.

18. Sleutels F, Zwart R, Barlow DP. The non-coding Air RNA is required for silencing autosomal imprinted genes. Nature, 2002, 14; 415(6873):810-813.

19. Volker A. Erdmann et al. Collection of mRNA-like non-coding RNAs. Nucleic Acids Research, 1999, 27: 192-195.

20. David Baltimore. Our genome unveiled. Nature, 2001, 409 : 814-816.

21. Griffiths-Jones S, Moxon S, Marshall M, Khanna A, Eddy SR, Bateman A. Rfam: annotating non-coding RNAs in complete genomes. Nucleic Acids Research, 2005 , 33(Database issue):D121-4

22. Liu C, Bai B, Skogerbo G, Cai L, Deng W, Zhang Y, Bu D, Zhao Y, Chen R. NONCODE: an integrated knowledge database of non-coding RNAs. Nucleic Acids Research, 2005, 33(Database issue):D112-5.

23. Pang KC, et al. RNAdb--a comprehensive mammalian noncoding RNA database. Nucleic Acids Research, 2005, 33 (Database issue):D125-30.

24. Athanasius F. Bompfűnewerer, et al. Evolutionary patterns of non-coding RNAs . Theory in Biosciences, 2005 (123) 301–369.

25. Anfinsen C.B.. Principles that govern the folding of protein chains. Science, 1973, 181:223-230.

26. Zuker. M, et al. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. Nucleic Acids Research, 1981,9:133-148.

27. R.D. Dowell. et al. Evaluation of several lightweight stochastic context-free grammars for RNA secondary structure prediction. BMC Bioinformatics, 2004, 5:71.

28. Ivo L. Hofacker. Vienna RNA secondary structure server. Nucleic Acids Research, 2003, 31(13): 3429-343.

29. Yasubumi Sakakibara, Michael Brown, Richard Hughey1, I.Saira Mian2, Kimmen Sjolander, Rebecca C.Underwood and David Haussler. Stochastic context-free grammars for tRNA modeling, 1994, 22:5112-5120.

30. Sankoff D. Simultaneous solution of the RNA folding, alignment and protosequence problems. SIAM Journal on Applied Mathematics,1985, 45:810-825.

31. Eddy SR, Durbin R. RNA Sequence Analysis Using Covariance Models**.** Nucleic Acids Research, 1994, 22**:**2079-2088.

32. Mathews D, Turner D. Dynalign: An algorithm for finding the secondary structure common to two RNA sequences. Journal of Molecular Biology, 2002, 317(2):191-203.

33. Gorodkin J, Heyer L, Stormo G. Finding the most significant common sequence and structure motifs in a set of RNA sequences. Nucleic Acids Research,1997, 25(18):3724-3732.

34. Perriquet O, Touzet H, Dauchet M: Finding the common structure shared by two homologous RNAs. Bioinformatics, 2003, 19:108-116.

35. Yasubumi Sakakibara. Pair hidden Markov models on tree structures. Bioinformatics, 2003, 19:232-240.

36. A.Coventry et al. MSARi: Multiple sequence alignments for statistical detection of RNA secondary structure. PNAS, 2004,101:12102-12107**.**

37. Hofacker I, Fekete M, Stadler P. Secondary structure prediction for aligned RNA sequences. Journal of Molecular Biology, 2002, 319:1059–1066.

38. Edwin ten Dam, Kees Pleij,t and David Draper. Structural and Functional Aspects of RNA Pseudoknots. BioChemistry, 1992,Volume 31, Number 47 December I.

39. Liming Cai, Russell L. Malmberg, and Yunzhou Wu .Stochastic modeling of RNA pseudoknotted structures: a grammatical approach**.** Bioinformatics, 2003, 19: 66-73.

40. Elena Rivas and Sean R. Eddy. A dynamic programming algorithm for RNA structure prediction including pseudoknots .Journal of Molecular Biology,1998, 285:2053-2068.

41. Jianhua Ruan, Gary D. Stormo, and Weixiong Zhang**.** ILM: a web server for predicting RNA secondary structures with pseudoknots. Nucleic Acids Research, 2004, 32:146-149.

42. Peter Schattner, Angela N. Brooks1 and Todd M. Lowe. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Research, 2005, 33: 686–689.

43. Ventsislav Rusinov, Vesselin Baev, Ivan Nikiforov Minkov and Martin Tabler. MicroInspector: a web tool for detection of miRNA binding sites in an RNA sequence Nucleic Acids Research, 2005, 33: 696-700.

44. Robert J. Klein, Ziva Misulovin, and Sean R. Eddy. Noncoding RNA genes identified in AT-rich hyperthermophiles. PNAS, 2002, 99:7542-7547.

45. Elena Rivas and Sean R Eddy. Noncoding RNA gene detection using comparative sequence analysis. BMC Bioinformatics, 2001, 2:8.

46. Washietl. S. et al. Fast and reliable prediction of noncoding RNAs. PNAS, 2005, 102: 2454-2459.

47. Yinglei Song, Chunmei Liu, Russell L. Malmberg, Fangfang Pan, Liming Cai. Tree Decomposition Based Fast Search of RNA Structures Including Pseudoknots in Genomes. IEEE Computer Society Bioinformatics Conference, 2005, 223-234.

48. Edith Heard and Christine M. Disteche. Dosage compensation in mammals: fine-tuning the expression of the X chromosome. Genes and development, 2006, 20:1848-1867.

49. Smith ER, Pannuti A, Gu W, Steurnagel A, Cook RG, Allis CD, Lucchesi JC.The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation , Mol. Cell. Biol., 2000, 20 (1), 312-318.

50. Meller VH, Kuroda MI. Sex and the single X chromosome. Adv Genet, 2002; 46:1-24.

51. Meller, V. H., Wu, K.-H., Roman, G., Kuroda, M. I. and Davis, R. L. roX1 RNA paints the X chromosome of male Drosophila and is regulated by the dosage compensation system. Cell, 1997, 88, 445-457.

52. Amrein H, Axel R. Genes expressed in neurons of adult male Drosophila. Cell, 1997, Feb 21; 88(4):459-69.

53. Weigang Gu, Xierong Wei, Antonio Pannuti and John C. Lucchesi. Targeting the chromatin-remodeling MSL complex of Drosophila to its sites of action on the X chromosome requires both acetyl transferase and ATPase activities. EMBO, 2000, 19,
5202–5211

54. Kelley RL, Meller VH, Gordadze PR, Roman G, Davis RL, Kuroda MI. Epigenetic spreading of the Drosophila dosage compensation complex from roX RNA genes into flanking chromatin. Cell, 1999; 98:513-22.

55. Victoria H. Meller1 and Barbara P. Rattner. The rox genes encode redundant male-specific lethal transcripts required for targeting of the MSL complex. EMBO Jounal. 2002, 21: 1084–1091.

56. Hubert Amrein and Richard Axel. Genes Expressed in Neurons of Adult Male Drosophila. Cell. 1997, 88: 459-469.

57. Park, Y., Oh, H., Meller, V.H. and Kuroda, M.I. Variable Splicing of Non-Coding roX2 RNAs Influences Targeting of MSL Dosage Compensation Complexes in Drosophila, RNA Biol, 2005, Oct 27;2 (4): 17114930

58. Franke, A. and Baker, B.S. The rox1 and rox2 RNAs are essential components of the compensasome, which mediates dosage compensation in Drosophila, Molecular cell, 1999, 4, 117-122.

59. Park Y, Kelley RL, Oh H, Kuroda MI, Meller VH. . Extent of chromatin spreading determined by roX RNA recruitment of MSL proteins, Science. 2002 Nov 22; 298(5598):1620-3.

60. Stuckenholz, C., Meller, V.H. and Kuroda, M.I. Functional redundancy within roX1, a noncoding RNA involved in dosage compensation in Drosophila melanogaster, Genetics, 2003,164, 1003-1014.

61. Huang, X. and Miller, W. A time-efficient, linear-space local similarity algorithm. Adv. Appl. Math., 1991, 12, 337-357.

62. Smith ER, Pannuti A, Gu W, Steurnagel A, Cook RG, Allis CD, Lucchesi JC. The drosophila MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. Mol Cell Biol., 2000, Jan;20(1):312-8.

63. Elena Puerta-Fernandez[*], Jeffrey E. Barrick[†,‡], Adam Roth[‡], and Ronald R. Breaker . Identification of a large noncoding RNA in extremophilic eubacteria, PNAS, 2006, vol(103), 19490-19495.

64. G. A. Soukup and R. R. Breaker. Relationship between internucleotide linkage geometry and the stability of RNA, RNA, 1999(5), 1308–1325.

APPENDICES

A. The alignment of sub1- rox1 RNAs among Drosophila Species Clustalw (1.83)

```
                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             5         15         25         35         45         55         65
rox1-sub1-melanogaster   -GTTACGTTC GGAGTGGAAA ATGGAAATTA AGTGAAATAT CCAGTGATCG ATCGGTAATA GTAAATTGTT
rox1-sub1-sechellia      -TTTGCGTTC GGAGTTGAAA ATGGAAATTC AGTGAAATAT GTAGTGATCG ATCGGTAATA ATTAATTGCT
rox1-sub1-erecta         -ATTACGTTC GGAGTTGAAA ATGGAAATGT AGTGAAATG- T---CAAGTG ATCGGTAATA GAAAACGGTT
rox1-sub1-yakuba         ATTTACGTTC GGAATTGAAA GTGAAAAAGA AGTGAATCAC T---TGACAG ATCGTTAATA GTAAGCTGTT
rox1-sub1-simulans       -TTTGCGTTC GGAGTTGAAA ATGGAAATTC AGTGAAATAT GTAGTGATCG ATCGGTAATA ATTAATTGCT
Clustal Consensus          ** ***** *** * ****  ** ***    ******           *  * **** *****    *   * *

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             75         85         95        105        115        125        135
rox1-sub1-melanogaster   TGATACGTTT AGGCCAGTTT ATAAGGCAAA TTCACAGCAG TTGTAAGTAA ATT-TAATCA GAGACCAGGG
rox1-sub1-sechellia      TGATACGTTT AAGCAAGCTT GTAAGGCAAA TACACAAAAG TTGTAAGTAA ATT-TAATCG GATACCAGGG
rox1-sub1-erecta         TGATACGTTT AAGCAAGCTT GTAAGCCTAT TTCATAGCAG TTGTAAGTAA TTTGTAATCA AAGACCAGGG
rox1-sub1-yakuba         TAATACGTTT AAGCAAGACT GTAAGCCCAT TTCACAGCAG TTGTATGTAA TTT-TAATCT TAGACCAGGG
rox1-sub1-simulans       TGATACGTTT AAGCAAGTTT GTAAGGCAAA TACACAACAG TTGTAAGTAA ATT-TAATCG GATACCAGGG
Clustal Consensus        * ******** * ** **  *  **** * *  * ** *  ** ***** ****  ** *****   * *******

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            145        155        165        175        185        195        205
rox1-sub1-melanogaster   CACCACACCC GAAAAGCGTG CAGATATTA- -----GAAGA CATGGGCGTA GTTTCATATA CGAGCTGTCC
rox1-sub1-sechellia      CACCACACCC GAAAAGCGTG GAGATATTA- ------GAGA CATGGGCGTA GTTTCACGTA CGAGCTGACC
rox1-sub1-erecta         CACCGCACTT CAAAAGCGTG CTGATATTA- -----GAAGA CAGGGGCGTA GTTTCATATA CGAGCAGTCC
rox1-sub1-yakuba         CTCTACACTC GAAAACCGTG CTGATATTGC TGATTGAAAA CAGGGGCGTA GTTTTATGAA CAAGCAGTCC
rox1-sub1-simulans       CACCACACCC GAAAAGCGTG GAGATATTA- ------AAGA CATGGGCGTA TTTTCACGTA CCAGCTGTCC
Clustal Consensus        * *  ***    **** ****   ******         * * ** *******  *** *   * * *** * **

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            215        225        235        245        255        265        275
rox1-sub1-melanogaster   CCTTCGGCTT TTTCGACAAG TGGCAGCCCT AATGGCCCTC GTTTTTTCGC CGACAAGCAT TTAATGCGTA
rox1-sub1-sechellia      CCTTCGGCTT TTTCGACAAG TGGCAGCCCT AATGGCAGTC GTTTTTTCGC CGACATGCAT TTAATGCGTA
rox1-sub1-erecta         TCTTCGGCTT TTTCGACAAG TGGCAGCCCT AATGGCCGTC GTTTTTTCGC CGACAAGCAT TTAATGCGTA
rox1-sub1-yakuba         CCTTCGGCTT TTTCGACAAG TGGCAGCCCT TATGGCCGTC GTTTTTTCGC CGACAAGCAT TTAATGCGTA
rox1-sub1-simulans       CCTTCGGCTT TTTCGACAAG TGGCAGCCCT AATGGCCGTC GTTTTTTCGC CGAACAGCAT TTAATGGCTA
Clustal Consensus         ********* ********** **********  *****  ** ********** ***   **** ******  **

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ...
                            285        295        305        315        325        335
rox1-sub1-melanogaster   GTCACCGAAG AAAAGTGTTA GTTACCAGGG CCTGCCCTTT TAAAATTAAA TTTAAATT-- ---
rox1-sub1-sechellia      GTCACCGAAG AGAAGTGTTA GTAAC-CGGG CCTGCCCTTT -AAAATTAAA TTTAAATTGA G--
rox1-sub1-erecta         GTCACCGAAG AGAAGTGTTA GTTAC-AGGG CCTGGCCTTT -AAAATTAAA TTTAAATTAA AGA
rox1-sub1-yakuba         GTCACCGAAG AGAAGTGTTA GCTAC-AGGG TCTGCCCTTT -AAAATTGAA ATTAATTT-- ---
rox1-sub1-simulans       TTCCCGGAAA AAAATTGTTA -TAAC-CGGG CCTGCCCTTT -AAAATTNNN NNNNNNNNNN N--
Clustal Consensus         ** * ***  * ** *****    **  ***  *** *****  ******
```

B. The alignment of sub2- rox1 RNAs among Drosophila Species Clustalw (1.83)

```
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                   5        15        25        35        45        55        65
melanogaster-rox1-sub2          TCTGGCAAGA TGTAGCGTCA AAAGAAAATT TATCAAACGG CATTGCCATC ATCGTGCAAC AATCCCAAAG
simulans-rox1-sub2              TAAGGCAAGA TGTAGCGTCA AAAGAAAATT CATCGAACGG CACTGCCATC ATCG-CAGGC AATCCCAAAG
sechellia-rox1-sub2             TAAGGCAAGA TGTAGCGTCG AAAGAAAATT CATCGAACGG CACTGCCATC ATCG-CAGGC AATCCCAAAG
erecta-rox1-sub2                TAAGGCAAGA TGCAGCCTCT AAAGAAAATT CATCGAAAGG CATTGCCATC ACCG-CAGTC AATACCAAAG
yakuba-rox1-sub2                TAAGGCAAGA TGTAGCCCCT TAAGAAAATT CATTGAACGG CATTGCCATC ACTA-CAGAC AATTTCAAAG
Clustal Consensus               *.:******* ** ***  *  :*********  ** .**.** ** ******* *  .  .. * ***  *****

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                   75        85        95        105       115       125       135
melanogaster-rox1-sub2          AAGCCATTTA GAATGCAGGC ATCCAGGCAA AAACCAGAAA ACGTGCCTGT GGCAGCTGCA CTTCGTGGCC
simulans-rox1-sub2              AAGCCATTTA GAATGCAGGC ATCCAGGCAA AAACCAGAAA ACGTGCCTGT GGCAGCTGCA GTTCGTGGCC
sechellia-rox1-sub2             AAGCCATTTA GAATGCAGGC ATCCAGGCAA AAACCAGAAA ACGTGCCTGT GGCAGCTGCA GTTCGTGGCC
erecta-rox1-sub2                AAGCCATTTA GAATGCAGGC ATCCAGGCAA AAACCAGAAA ACGTGCCTGT AGCAGCTGCA GTTCGTGGCC
yakuba-rox1-sub2                AAGCCATTTA GAATGCAAGC ATCCAGGCAA AAACCAGAAA ACGTGCCTGT ACCAGCTGCA GTTCGTGGCC
Clustal Consensus               ********** *******.** ********** ********** ********** . ********  *********

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  145       155       165       175       185       195       205
melanogaster-rox1-sub2          TTGACGAACC CGGACAGTAG AAAAACCCTG GGAACCTAAC CCAAATGGCT CATAGGGGTT GGACTGGACT
simulans-rox1-sub2              TTGACGAACC CGGACAATAG AG--ACCCTG GGAACCTAAC CCAAGTGGCT TATAGGGGTT GGACTGGACT
sechellia-rox1-sub2             TTGACGAACC CGGACAATAG AG--ACCCTG GGAACCTCAC CCAAGTGGCT TATAGGGGTT GGACTGGACT
erecta-rox1-sub2                TTGGCGACCC CGGACAATAG AG--GCCCTG GGAACCTAAC CCAAATGGCT TATAGGGGTT GGACTGGACT
yakuba-rox1-sub2                TTGACGACCC CGGACAAGAG AG--ACCCTG GGAACCTAAC CCAAATGGCT TATAGGGGTT GGACTGGACT
Clustal Consensus               ***.***.** ******. ** *.  .***** *******.** ****.*****  ********* **********

                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                  215       225       235       245       255       265       275
melanogaster-rox1-sub2          CAGGTAGGCG AAGTCCCCGA AAACCGAGGC TACATTCTTG CCACCGCTGC TAAATCGATC ATCAGTCTTT
simulans-rox1-sub2              CAGGTAGGCG AAGTCCCG-A AAACCGAGGC TACATTCTTG CCACCGCTGC TAAATCGATC ATCAGTCTTC
sechellia-rox1-sub2             CAGGTAGGCG AAGTCCCG-A AAACCGAGGC TACATTCTTG CCACCGCTGC TAAATCGATC ATCAGTCTTC
erecta-rox1-sub2                CAGGTAGGCG AGGTCCCCGA AAACCGAGGC TACATTCTTG CCACCGCTGC TAAATCGATC ATCAGTCGTC
yakuba-rox1-sub2                CAGGTAGGCG AAGTCCCCGA AAACCGAGAC TACATTCTTG CCACCGCTGC TAAATCGATC ATCAGTCGTC
Clustal Consensus               ********** *.*****  * ********.* ********** ********** ********** ******* *

                                ....|....| ....|....| ....|....| ....|....| ....|....| .
                                  285       295       305       315       325
melanogaster-rox1-sub2          GGCGGCATGG CTAAGTGGAA ACTTCTCGTA AGAAACTCTT CCCGGAGGAG T
simulans-rox1-sub2              GGCGGCATGG CTAAGTGGAA ACTTCTCGTA AGAAACTCTT CCCGGAGGAG T
sechellia-rox1-sub2             GGCGGCATGG CTAAGTGGAA ACTTCTCGTA AGAAACTCTT CCCGGAGGAG T
erecta-rox1-sub2                GGCGGCATGG CGTAGTGGAA ACTTCTCGTA AGAAACTCAT CCCGGAGGAG T
yakuba-rox1-sub2                GGCGGCATGG CGTAGTGGAA ACTTCTCGTA AGGAACTCTT CCCGAAGGAG T
Clustal Consensus               ********** * :******* ********** **.*****:* ****.***** *
```

```
                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             5        15        25        35        45        55        65
melanogaster-rox1-sub2   TCTGGCAAGA TGTAGCGTCA AAAGAAAATT TATCAAACGG CATTGCCATC ATCGTGCAAC AATCCCAAAG
simulans-rox1-sub2       TAAGGCAAGA TGTAGCGTCA AAAGAAAATT CATCGAACGG CACTGCCATC ATCG-CAGGC AATCCCAAAG
sechellia-rox1-sub2      TAAGGCAAGA TGTAGCGTCG AAAGAAAATT CATCGAACGG CACTGCCATC ATCG-CAGGC AATCCCAAAG
erecta-rox1-sub2         TAAGGCAAGA TGCAGCCTCT AAAGAAAATT CATCGAAAGG CATTGCCATC ACCG-CAGTC AATACCAAAG
yakuba-rox1-sub2         TAAGGCAAGA TGTAGCCCCT TAAGAAAATT CATTGAACGG CATTGCCATC ACTA-CAGAC AATTTCAAAG
Clustal Consensus        *.:******* ** ***  *  :*********  ** .**.** ** ******* *  .  .. * ***  *****

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             75        85        95       105       115       125       135
melanogaster-rox1-sub2   AAGCCATTTA GAATGCAGGC ATCCAGGCAA AAACCAGAAA ACGTGCCTGT GGCAGCTGCA CTTCGTGGCC
simulans-rox1-sub2       AAGCCATTTA GAATGCAGGC ATCCAGGCAA AAACCAGAAA ACGTGCCTGT GGCAGCTGCA GTTCGTGGCC
sechellia-rox1-sub2      AAGCCATTTA GAATGCAGGC ATCCAGGCAA AAACCAGAAA ACGTGCCTGT GGCAGCTGCA GTTCGTGGCC
erecta-rox1-sub2         AAGCCATTTA GAATGCAGGC ATCCAGGCAA AAACCAGAAA ACGTGCCTGT AGCAGCTGCA GTTCGTGGCC
yakuba-rox1-sub2         AAGCCATTTA GAATGCAAGC ATCCAGGCAA AAACCAGAAA ACGTGCCTGT ACCAGCTGCA GTTCGTGGCC
Clustal Consensus        ********** *******.** ********** ********** ********** . ********  *********

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                           145       155       165       175       185       195       205
melanogaster-rox1-sub2   TTGACGAACC CGGACAGTAG AAAAACCCTG GGAACCTAAC CCAAATGGCT CATAGGGGTT GGACTGGACT
simulans-rox1-sub2       TTGACGAACC CGGACAATAG AG--ACCCTG GGAACCTAAC CCAAGTGGCT TATAGGGGTT GGACTGGACT
sechellia-rox1-sub2      TTGACGAACC CGGACAATAG AG--ACCCTG GGAACCTCAC CCAAGTGGCT TATAGGGGTT GGACTGGACT
erecta-rox1-sub2         TTGGCGACCC CGGACAATAG AG--GCCCTG GGAACCTAAC CCAAATGGCT TATAGGGGTT GGACTGGACT
yakuba-rox1-sub2         TTGACGACCC CGGACAAGAG AG--ACCCTG GGAACCTAAC CCAAATGGCT TATAGGGGTT GGACTGGACT
Clustal Consensus        ***.***.** ******. ** *.  .***** *******.** ****.*****  ********* **********

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                           215       225       235       245       255       265       275
melanogaster-rox1-sub2   CAGGTAGGCG AAGTCCCCGA AAACCGAGGC TACATTCTTG CCACCGCTGC TAAATCGATC ATCAGTCTTT
simulans-rox1-sub2       CAGGTAGGCG AAGTCCCG-A AAACCGAGGC TACATTCTTG CCACCGCTGC TAAATCGATC ATCAGTCTTC
sechellia-rox1-sub2      CAGGTAGGCG AAGTCCCG-A AAACCGAGGC TACATTCTTG CCACCGCTGC TAAATCGATC ATCAGTCTTC
erecta-rox1-sub2         CAGGTAGGCG AGGTCCCCGA AAACCGAGGC TACATTCTTG CCACCGCTGC TAAATCGATC ATCAGTCGTC
yakuba-rox1-sub2         CAGGTAGGCG AAGTCCCCGA AAACCGAGAC TACATTCTTG CCACCGCTGC TAAATCGATC ATCAGTCGTC
Clustal Consensus        ********** *.*****  * ********.* ********** ********** ********** ******* *

                         ....|....| ....|....| ....|....| ....|....| ....|....| .
                           285       295       305       315       325
melanogaster-rox1-sub2   GGCGGCATGG CTAAGTGGAA ACTTCTCGTA AGAAACTCTT CCCGGAGGAG T
simulans-rox1-sub2       GGCGGCATGG CTAAGTGGAA ACTTCTCGTA AGAAACTCTT CCCGGAGGAG T
sechellia-rox1-sub2      GGCGGCATGG CTAAGTGGAA ACTTCTCGTA AGAAACTCTT CCCGGAGGAG T
erecta-rox1-sub2         GGCGGCATGG CGTAGTGGAA ACTTCTCGTA AGAAACTCAT CCCGGAGGAG T
yakuba-rox1-sub2         GGCGGCATGG CGTAGTGGAA ACTTCTCGTA AGGAACTCTT CCCGAAGGAG T
Clustal Consensus        ********** * :******* ********** **.*****:* ****.***** *
```

C. The alignment of sub3- rox1 RNAs among Drosophila Species Clustalw (1.83)

```
                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             5         15         25         35         45         55         65
rox1-sub3-melanogaster   GCTACTGATA GGGTGACCTA ACGCAACAGC AACGAGATGT GCTCCATCGG GCAGTGATAA CTTCTTGGCA
rox1-sub3-simulans       GCTACAGATA GGGAGACCTA ACGAAACAGC AGCGAGATGT GCTCCATCGG GCAGTGATAA CTTCTTGGCA
rox1-sub3-sechellia      GCTACAGATA GGGAGACCTA ACGCAACAGC AGCGAGATGT GCTCCATCGG GCAGTGATAA CTTCTTGGCA
rox1-sub3-erecta         GCTACAACTA GGGAGACCTA ACGCAACCGC ACCGAAACGT GTTCCAACGG GCAGTGACAA CTTCTGGGCA
rox1-sub3-yakuba         GCTAGAGCTA GGGAGACATA ACGCAACAGC ACCGAGATAT GTTCCATCGG GCAGTGACAA CTTCTGGGCA
Clustal Consensus        ****    ** *** *** ** *** *** ** * *** *  * * **** *** ******* ** ***** ****

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             75         85         95        105        115        125        135
rox1-sub3-melanogaster   ACCCCA---- -ACCTTAAAA ATTCGACT-- TCAATTCA-T ACATGTGACC AGATAAACAC AGCTTCAAGC
rox1-sub3-simulans       TCCCCA---- -ACCTTAAAA ATCCGACT-- TCAATTCAAT ACATGTGACC AGATAAGCCC AGCTTCAAGC
rox1-sub3-sechellia      TCCCCA---- -ACCTTAAAA ATCCGACT-- TCAATTCAAT A--------C AGATAAGCCC AGCTTCAAGC
rox1-sub3-erecta         TCCCCAGCCC CACCTTAAAA ATCTGACTCC TCAATTCAAT ACATGTGACC AGAAACACAC AGCTTCAAGC
rox1-sub3-yakuba         TCCCCAGCCC CACCTTAAAA ATCCGTCTCC TCAATTCAAT ACATGCGACC AGAAACACAC AGCTTCAAGC
Clustal Consensus         *****      ********* **  * **   ******** * *        * *** *  * * **********

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            145        155        165        175        185        195        205
rox1-sub3-melanogaster   GCATTACTCG CTCCATAGAT TCATTAAACT TTAATGAACA CAAGAAACAC TGAAAAAGTA AGAGATTCTT
rox1-sub3-simulans       CCAGTACTCG CTCCATAGAT TCATTAAACC TTAATGAACC CAAGAAACAC TAAAAACGTA AGAGATTCTT
rox1-sub3-sechellia      CCAGTACTCG CTCCATAGAT TCATTAAACC TTAATGAACC CAAGAAACAC TCAAAAAGTA AGAGATTCTT
rox1-sub3-erecta         GCAATACTCG CTTCATAGAA TCATTAAAAC TTGATGAAGC AAAGAAACCA TCAAAAAGTA AGAGATTCTT
rox1-sub3-yakuba         CCAATACTCG CTCCATAGAA CCATTAAACC TTGATGAAGC AAAGTAACCA TCAAAAAGCA AGAGATCCTC
Clustal Consensus         ** ****** ** ******   *******   ** *****    *** ***   * **** * * ****** **

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            215        225        235        245        255        265        275
rox1-sub3-melanogaster   CGCCTGAGCT GCAATTCTTA AGGGT--GGC GTCTTCGAGG CATCGAAGAC GCGGCTATCC GAAGGAGCGA
rox1-sub3-simulans       CGCCTGAGCT GCAATTCTTA GGGGT--GGC GTCTTCGAGG CATCGAAGAC GCGGCTATCT GAAAGAGCGA
rox1-sub3-sechellia      CGCCTGAGCT GCAATTCTTG CAATTCTGGC GTCTTCGAGG CATCGAAGAC GCGGCTATCT GAAAGAGCGA
rox1-sub3-erecta         CACCTGAGCT GCAATTCTTC CGGGT--GGC GTCTTCGAGG CATCGAAGAC GCGGCTGTCT GAAATAGCGA
rox1-sub3-yakuba         CACCTGAGCT GCAATTCTTC CGGGT--GGC GTCTTCGAGG CATCGAAGAA GCGGCTGTCT GAAAGAGCGA
Clustal Consensus        * ******** *********      *  *** ********** *********  ****** **  ***  *****

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            285        295        305        315        325        335        345
rox1-sub3-melanogaster   AAGTGGTGCG AGTGGTGCAA CTTGTGCCAC TGCCAGCGGA GTGGATGAGC CCCATATAAA CTGAACCTTG
rox1-sub3-simulans       AAGTGGTGCG AGTGGTGCAA CTTGTGCCAC TGCCAGCGGA GTGGATGAGC CCCATATAAA CTGAACCTTG
rox1-sub3-sechellia      AAGTGGTGCG AGTGGTGCAA CTTGTGCCAC TGCCAGCGGA GTGGATGAGC CCCATATAAA CTGAACCTTG
rox1-sub3-erecta         AAGTGGTGCG AGTGGTGCAA CTTGTGCCAC TGCCAGCGGA GTGGATGAGT CCCATATAAA ATGAACCTTG
rox1-sub3-yakuba         AAGTGGTGCG AGTGGTGCAA CTTGAGCCAC TGCCAGCGGA GTGGATGATT CCCATATAAA ATAAACCTTG
Clustal Consensus        ********** ********** **** ***** ********** ********   **********  * *******

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            355        365        375        385        395        405        415
rox1-sub3-melanogaster   CAGACCTACA AAGACCCTAT GGCCCATAAA GCCGGACAGG G---TTTTCC TGTCCGAAAA GTAATGTAAC
rox1-sub3-simulans       CAGACCTACA AAGACCCTGT GGCCCATAAA TCCGGAGGGA GAAATACTCC CGCCCGAAAA GTAATGAACC
rox1-sub3-sechellia      CAGACCTACA AAGACCCTGT GGCAGATAAA TCCGGAGGGA GAAATACTCC CGCCCGAAAA GTAATGAAAC
rox1-sub3-erecta         CAGACTTA-G GAGACCCTAT GGCCCATAAA TCCAGGGGGA GCAATACTCC TGCCCGGAAA GTAAGGAAAC
rox1-sub3-yakuba         CAGACTTACG AAGACCCTAT GGCCCATAAG TCCAGGGGGA GCAATACTCC CGCCCGGAAA GTAAGGAAAC
Clustal Consensus        ***** **    ******* * ***  ****   ** *   *  *   *  ***  * *** *** **** * * *

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            425        435        445        455        465        475        485
rox1-sub3-melanogaster   GAAAACGCTT GAGGTGGGGG GTTCTGGCCG GAACGAAAGA GA-----CAA ATGAACCC-A AAGCGTCCGC
rox1-sub3-simulans       GAAAACGCTT GAGATGGAAA AGTCTGGCCG GAACGAAAGA GA-----CAA ATGAACCC-A GAGCGTCCGC
rox1-sub3-sechellia      GAAAACGCTT GAGATGGAAA AGTCTGGCCG GAACGAAAGA GA-----CAA ATGAACCC-A GAGCGTCCGC
rox1-sub3-erecta         GAAAACGCTT GAGATGGAAA AGTCTGACTG GAACGAAAGA GA-----CAA AGGAGCCCCA GAGCATTCGC
rox1-sub3-yakuba         GAAAACGCTT GAGATGGAAA AGTCTGACTG GAACGAAAGA GGAGCCCCAA AGGAGCCCTA GAGCATTCGC
Clustal Consensus        ********** *** ***      **** * * ********** *      *** * ** *** *  *** * ***
```

```
                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             495        505        515        525        535        545        555
rox1-sub3-melanogaster    TTGTCGCCAC TTGCGTTTCG AGAAGACGCG TACGCATACC TCTATCGGAT GCATGTGGCG GTACGCGGAT
rox1-sub3-simulans        TTGTCGCCAC TTGCGTTTCG AGAAGACGCG TACGCATACC TCTATCGAAC GCATGTGGCG CTACGCGGAT
rox1-sub3-sechellia       TTGTCGCCAC TTGCGTTTCG AGAAGACGCG TACGCATACC TCTGTCGAAC GCATGTGGCG CTACGCGGAT
rox1-sub3-erecta          TTGGCGCCAC CTGCGTTTCG AGAAGACGCG TACGCATACC TCTATCGGTC GCATGTGGCG GTACGCGGAT
rox1-sub3-yakuba          TTGGCGCCAC CTGCGTTCCG AGAAGACGCG TACGCATACC TCTATCGGTC GCATGTGGCG GTACGCGGAT
Clustal Consensus         *** ****** ****** ** ********** ********** *** ***    **********  *********

                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             565        575        585        595        605        615        625
rox1-sub3-melanogaster    GCGAGCGAGA CAATGATACT GTAGACAAGG AGAGACGGCT CCCTCTCTCC ATCACTCTCT GTCGGGCTGG
rox1-sub3-simulans        GCGAGCGAGA CAATGATACT GTAGACAAGG AGAGACGGCT CCCTCTCTCC ATCACTCTCT GTCGCGCTGG
rox1-sub3-sechellia       GCGAGCGAGA CAATGATACT GTAGACAAGG AGAGACGGCT CCCTCTCTCC ATCACTCTCT GTCGCGCTGG
rox1-sub3-erecta          GCGAGCGAGA CAACAATACT GTAGACAAGG AGAGACGGCT CCCTCTCTCC ATCACTCTCT GTCGCGCTGG
rox1-sub3-yakuba          GCGAGCGAGA CAACGATACT GTAGACAAGG AGAGACGGCT CCCTCTCTCC ATCACTCTCT GTCGCGCTGG
Clustal Consensus         ********** ***  ***** ********** ********** ********** ********** **** *****

                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             635        645        655        665        675        685        695
rox1-sub3-melanogaster    CCAGGCGTTT GCATCGCAAG CGGGTGAGTG AGACGGCCAT AGGGCGGACT GCAAGTCCCG AAAGAGAGCG
rox1-sub3-simulans        CCAGGCGTTT GCATCGCAAG CGGTGGAGTG AGACGGCCAT AGGGCGGACT GCAAGTCCCG AAAGAGAGCG
rox1-sub3-sechellia       CCAGGCGTTT GCATCGCAAG CGGTTGAGTG AGACGGCCAT AGGGCGGACT GCAAGTCCCG AAAGAGAGCG
rox1-sub3-erecta          CCAGGCGTTT GCATCGCAAG CGGGTGAGTG AGATGGCCAT AGGGCGGACT GCAAGGCCCG AAAGAGAGCA
rox1-sub3-yakuba          CCAGGCGTTT GCATCGCAAG CGGGTGAGTG AGACGGCCAT AGGGCGGACT GCAAGGCCCG AAAGAGAGCG
Clustal Consensus         ********** ********** ***  ***** *** ****** ********** ***** **** *********

                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             705        715        725        735        745        755        765
rox1-sub3-melanogaster    GTAAAGCCAT ATAACCCCTA TAAGCAATAA ACGACGCATC GAAAGTTGCG TATAACGGCA TGGGCATCCC
rox1-sub3-simulans        GTAAAGCCAT ATAACCCCGA TAAACAATAA TCGACACATC GAAAGATGCG TATAACG-CA TGGGCATCCC
rox1-sub3-sechellia       GTAAAGCCAT ATAACCCCGA TAAACAATAA TCGACACATC GAAAGATGCG TATAACG-CA TGGGCATCCC
rox1-sub3-erecta          GTAAAGCTGT ATAACACCGA TAAGCAATAA TCGACGCATC GAAAGATGCG TATAACA-CA CGGGCATCCC
rox1-sub3-yakuba          GTAAAGCTAT ATAACCCCGA TAAGCAATAA TCGACGCATC GAACGAGGCG TATAACA-CA CGGGCATCCC
Clustal Consensus         *******  * ***** ** * *** ******  **** **** *** *  *** ******  **  *********

                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             775        785        795        805        815        825        835
rox1-sub3-melanogaster    TCGGAAAGGA AAAAGTTCAG ACCGCACTCA CGAATACCGC TCCGCTCTGG AAAGACCGGC CAAATCTGTA
rox1-sub3-simulans        TCGGAAAGCA AAGTGTTCAG ACCGCACTCA CGAATACCGC TCCGCTCTGA AAAGACCGGC CAAATCTGTA
rox1-sub3-sechellia       TCGGAAAGCA AAGTGTTCAG ACCGCACTCA CGAATACCGC TCCGCTCTGA AAAGACCGTC CAAATCTGTA
rox1-sub3-erecta          TCGGAAAGCA AAGAGTTCAG ACCGCACTCA CGAATACCGC TCCGCTCTGA AAGGACCGGC CAAACCAGAA
rox1-sub3-yakuba          TCGTAAAGCA AAGAGTTCAG ACCGCACTCA CGAATACCGC TCCGCTCTAA AAGGACCGGC CTAATCTGCA
Clustal Consensus         *** **** * **  ****** ********** ********** ********   ** ***** * * ** * * *

                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             845        855        865        875        885        895        905
rox1-sub3-melanogaster    AGAAATGCTC GGCCAGGAGC GTAGAATCGC C-ATAATGCA GTAAACGACT GCAAAAGCAG CTATAACTCA
rox1-sub3-simulans        GGAAATGCTT GGCCAGGAGC CTAGAATCGC CCATAATGCA GTAAACGACT GCAAAAGCAG CTATAACTCA
rox1-sub3-sechellia       AGAAATGCTT GGCCAGGAGC CTAGAATCGC C-ATAATGCA GTAAACGACT GCAAAAGCAG CTATAACTCT
rox1-sub3-erecta          AGAAATGCTC TGCCAGGAGC GTAGAATCGC A-ATAAAGCA GTAAACAACT GCCAAAGCAG CAATAACTCA
rox1-sub3-yakuba          AGAAATGCTG GGTTAGGAGT GTAGAATCAC C-ATAAAGCA GTATACAACT GCCCAAGCAG CTATAACTCA
Clustal Consensus          ********   *  *****   ******* *   **** *** *** ** *** **  ****** * *******

                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             915        925        935        945        955        965        975
rox1-sub3-melanogaster    AACAGAAGTT GACAAAAGCA AGAGCTATGC GCTCTATTAA AGTCACAATA AGTGTCATCG TAACGTTCAC
rox1-sub3-simulans        AACCGAATCT GACAAAAGCA AGAGCTATGC GCTCTATTAA AGTCA--AAA AGTGTCATCG TAACGTTCAC
rox1-sub3-sechellia       AACCGAATCT GACAAAAGCA AGAGCTATGC GCTCTATTAA AGTCACAAAA AGTGTCATCG TAACGTTCAC
rox1-sub3-erecta          AACCAAAGCT GACAAA-GCC AGAGCTATGC GCTCTATTAA AGCCAAAGAA AGTGTCATCG TAACGTTCAC
rox1-sub3-yakuba          TATCAAAGCT GACAAA-GCA AGAGCTATGC GCTCTATTTA AGTCACAGAA AGTGTCATCG TAACGTTCAC
Clustal Consensus          *   **  * ****** **  ********** ******** * ** **    * ********** **********

                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             985        995       1005       1015       1025       1035       1045
rox1-sub3-melanogaster    CACTTTTCAC ATTTCATCAA AGTCCATTTC TGGTCCGTGC AAAGCAGTAA TGTAGAGATT CTGTTTCATT
rox1-sub3-simulans        CACTCTTCAC ATTTCATCAA AGTCCATTTC TGGTTCGTGC TAAGCAGTAT TGAAGAGATT ATGTTTCATT
rox1-sub3-sechellia       CACTCTTCAC ATTTCATCAA AGTCCATTTC TGGTTCGTGC TAAGCTGTAT TGAAGAGATT ATGTTTCATT
rox1-sub3-erecta          CACTCTTCAC ATTCCATCAA AGTCCATTTC TGGTCCCTAC AAAGCAGTAT TGTAGAGGTT CTTTTCTATT
rox1-sub3-yakuba          CACTCTTCAT ATTCCATCAA AGTCCATTTC TGGTCCGTAC AAAGCAGTAT TGTAAAGGTT ATCTTCCAGT
Clustal Consensus         **** ****  *** ****** ********** **** * * *  **** ***  ** * ** **  * **  * *
```

```
                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             1055       1065       1075       1085       1095       1105       1115
rox1-sub3-melanogaster    TGA------- AATTGTTGGT TATCACTGCC CTAGGAACTC ATCAATTTTT GT--GATTAT AGTGACTGAA
rox1-sub3-simulans        TGA------- AATTGTTGGT TATCACTGCC CTAGGAACTC ATCAATTTAT GT--GATTAT AGTGACGGAA
rox1-sub3-sechellia       TGA------- AATTGTTGGT TATCACTGCC CTAGGAACTC ATCAATTTAT GT--GATTAT AGTGACTAAA
rox1-sub3-erecta          TGTTTTATAT AGTTGTTGGT TATCACTGCC C-AGGAACTC ATCAATTTTT TTTGGATTAT ACGGACTGAA
rox1-sub3-yakuba          TGTGGTACAT ATTTGTTGGT TATCACTGCC CTAGGAACTC ACCAATTTTT TG-GGATTAT ACGGACTGAT
Clustal Consensus         **         * ******** ********** * ******** * ****** *     ****** *  ***  *

                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             1125       1135       1145       1155       1165       1175       1185
rox1-sub3-melanogaster    GTATTTGTTT TGAGGGAATT AGAAGCCAGA AGATTTAGAA GCCATGCCGC CGCTTAAGTA AGATCTAAAA
rox1-sub3-simulans        GTATTTGTTA TGAGGGACTT --------GT AGAATTAGGA GCCATGCCGC CGCTTAACTA AGATCTAAAA
rox1-sub3-sechellia       GTATTTGTTA TGAGGGACTT --------GT AGAATTAGGA GCCATGCCGC CGCTTAACTA AGATCTAAAA
rox1-sub3-erecta          GTGTTTGTTA TGAAGAACTT --------GT AGAATTAGGA GCCATGCCGC CGCTTAACTA AGATCCAAAA
rox1-sub3-yakuba          GTGTTTGTTA TGAGGAACAT --------GT AGAATTAGGA GCCATGCCGC CGCTTAACTA AGATCCAAAA
Clustal Consensus         ** ******  *** * *  *         *  *** **** * ********** ******* ** ***** ****

                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             1195       1205       1215       1225       1235       1245       1255
rox1-sub3-melanogaster    GCCAAGAAGC TACACCCAGA AGAAACTGCC ACTGCATAGA TTCTTGAAAT GAGCAAAGAA ATCTAAACTG
rox1-sub3-simulans        GCCAAGAAGC TACACCCAGA AGAAACTGCC GCTGCATAGA TTCTTAAAAT GAGCAAAAAA G-CTGAACTG
rox1-sub3-sechellia       GCCAAGAAGC TACACCCAGA AGAAACTGCC GCTGCATAGA TTCTTAAAAT GAGCAAAAAA G-CTGACCTG
rox1-sub3-erecta          GCCAAGAAGC TACGCCCAGA AGAAACTGCC ACTGCATAGA TTCTTGAAGT GAGCAAGAAA G-CTGAACTG
rox1-sub3-yakuba          GCCGAGAAGC TACGCCCAGA AGAAACTGCC GCTGCATAGA TTCTT----- GAGAAATAAA G-CTGAACTA
Clustal Consensus         *** ****** *** ****** **********  ********* *****      *** **  **   ** * **

                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             1265       1275       1285       1295       1305       1315       1325
rox1-sub3-melanogaster    AAATGGTTAT CCACCAACTG AAGGCCATCG AATCTATTAA GTTTTATCCA CGTTAGCACT GTTCAAAATC
rox1-sub3-simulans        AAATGGCTAT CCACCGACTG AAGGCCATCG AAACTATCAG ATTTTATCCG CGTTAGCACT GTTCAAAATC
rox1-sub3-sechellia       AAATGGCTAT CCACCGACTG AAGGCCATCG AAACTTTTAG ATTTTATCCG CGTTAGCACT GTTCAAAATC
rox1-sub3-erecta          AAATGGTTAT CCACCGACTA CAGGGCATCT AACCTATTAA ACTTAATG-A AGACAGCACC TTTCAAAACC
rox1-sub3-yakuba          AAATGATTAT CCACCAACTG AAGGGCATCT AAACTATTAT GTTTAATCCA AGTCAGCATT GTTCAAAATC
Clustal Consensus         *****  *** ***** ***   *** ****  ** ** * *    ** **     *  ****    ******* *

                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             1335       1345       1355       1365       1375       1385       1395
rox1-sub3-melanogaster    TCAGGCTAGG CCACTAAACG AACT------ ATTCATCACA AAAGCCTCAC ACAAAATATT CATTCATGAA
rox1-sub3-simulans        TCAAGCTAGG CCACTAAACG AACT------ ATTCATCACA AAAACC--AC ACAAAATATT AATTCATGAA
rox1-sub3-sechellia       TCAAGCTAGG CCACTAAACG AACT------ ATTCATCACA AAAACC--GC ACAAAATATT CATTCATGAA
rox1-sub3-erecta          TCAAGCTAGG CCACGAAACG AAC------A ATTCATCACA AAAGCC--AC AACAAATATT GACTCAAGAA
rox1-sub3-yakuba          TCAAGCTAGG CCACAAAACG AATTCATCGA ATTCATTATA AAAGCC--AA AA-AAATATC GACTCAAGAA
Clustal Consensus         *** ****** **** ***** **         ****** * * *** **     *  ******   * *** ***

                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             1405       1415       1425       1435       1445       1455       1465
rox1-sub3-melanogaster    TCAGTCTAGC CTATTTTTTC TTCTGCTTTA CCATTTGAAC AAGTACTC-- -AGAAATCGC TCGAAAAGGG
rox1-sub3-simulans        TCAGTCTAGC TTACTTTT-C TATTGCTTTG CCATTTTAAC AAATACTC-- -AGAAATCGC TCGAAAAGGG
rox1-sub3-sechellia       TCAGTCTAGC TTATTTTT-C TATTGCTTTG CCATTTTAAC AAGTACTC-- -AGAAATCGC TCGAAAAGGG
rox1-sub3-erecta          TCAGTTTAGC TCATCTG--- CACTCCTT-- ----TCTAAC AAGTACTCAG CAGAAATCAC TCGAAAAGGG
rox1-sub3-yakuba          TCAGTCTAGC TCATCTGTTC CATTGCTT-G CGATTTTGAC AAGTACTC-- -AGAAATCGC TCGAAAAAGG
Clustal Consensus         ***** ****   *  *       * ***       *   ** ** *****    ******* * ******* **

                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             1475       1485       1495       1505       1515       1525       1535
rox1-sub3-melanogaster    ACATTTTGAT TTTGCTTTGA TTTTGTGCCT TAGTTTTTCA ACTAACCAAA ATATTGCTTA CAGAAATTAA
rox1-sub3-simulans        ACATTTTTAT TTTGCTTTGA TTTTGTGCCT TAGTTTTG-A ACTAACCAAA ATATTGCTTG CAGAAATTAA
rox1-sub3-sechellia       ACATTTTTAT TTTGCTTTGA TTTTGTGCCT TAGTTTTG-A ACTAACCAAA ATATTGCTTG CAGAAATTAA
rox1-sub3-erecta          ACATTTTGAT TTTGCTTTAT TTATGTGCTT TAGTTTCG-A ACTAACCAAA ATATTGCTTG CAGAAATTAA
rox1-sub3-yakuba          ACATTTTGAT TTTGCTTT-- --ATGTGCCT TAGTTTTG-A ACTAACCAAA ATTTTGCTTT CAGAAATTAA
Clustal Consensus         ******* ** ********      ***** * ******   * ********** ** ******  **********

                          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             1545       1555       1565       1575       1585       1595       1605
rox1-sub3-melanogaster    TCAGTTTATT AACTGAAAAA GCCCTTAAAT TAACTGCCAG GAGGTGAGTT TTTGTCTTCG ATTCTTGGAA
rox1-sub3-simulans        TCAGTTTATT AACTGAAAAA GCCCTTAAAT TAACTGCCAG GAGGTGAGTT TTTGTCTTCG ATTCTTGGAA
rox1-sub3-sechellia       TCAGTTTATT AACTGAAAAA GCCCTTAAAT TAACTGCCAG GAGGTGAGTT TTTGTCTTCG ATTCTTGGAA
rox1-sub3-erecta          TCAGTTTATT AGCTGAAAAA GCCCTTAAAT TAACTGCCAG GAGGTGAGTT TTTGTCTTCG ATTCTTGGAA
rox1-sub3-yakuba          TCAGTT-ATT AGCTGAAAAA GCCCTTAAAT TAACTGGCAG GAGGTGAGTT TTTGTCTTTG ATTCTTGGAA
Clustal Consensus         ****** *** * ******** ********** ****** *** ********** ******** * **********

                          ....|....| ....|....| ....|....| ....|....| ....|....
                             1615       1625       1635       1645       1655
rox1-sub3-melanogaster    AACGTTGAGT AGGGGCCCAA CTAAAGGAAA TAGGCCAGAC AAAGCCTGT
rox1-sub3-simulans        AACGTTGAGT AGGGGCCCAA CTAAAGGAAA TAGGCCAGAC AAAGCCTGT
rox1-sub3-sechellia       AACGTTGAGT AGGGGCCCAA CTAAAGGAAA TAGGCCAGAC AAAGCCTGT
rox1-sub3-erecta          AACGTTGAGT AGGGGCCCAA CTAGAGGAAA AAGGACAGGC AAAGCCTGT
rox1-sub3-yakuba          AACGTTGAGT AGTGGCCCAA CTAGAGGAAA CAGGCCAGGC AAAGCCTGT
Clustal Consensus         ********** ** ******* *** ******  *** *** * *********
```

D. The alignment of sub4- rox1 RNAs among Drosophila Species Clustalw (1.83)

```
                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            5         15         25         35         45         55         65
rox1-sub4-melanogaste    AAAAAAAACC TATAAAAATA AAAAGTTCTA AAAGAGGCCA ATGTGAAACT GTTATACAAG CTG-GCCTAA
rox1-sub4-simulans       TGTAAAAACC TATACAAATA TAAAGTTCTA AAAGAAGCCA ATGTGAAACT GTTATACAAG CTG-GTCGAA
rox1-sub4-sechellia      TGTAAAAACC TATACAAATA TAAAGTTCTA AAAGAAGCCA ATGTGAAACT GTTATACAAG CTG-GTCGAA
rox1-sub4-erecta         ATAAAAAACC TATAAAAATA ACAAGTTCCA ACAGAAGCCA ATATGAAACT GTTATACAAG CCAAGTCAAA
rox1-sub4-yakuba         ATAAAAAACC TATAAAAATA ACAAGTTCCA ATAAAAGCCA ATGTGAAATT GTTATACAAG CTG-GTCAAA
Clustal Consensus           ******* **** *****   ****** * * * * **** ** ***** * ********** *   * * **

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            75         85         95        105        115        125        135
rox1-sub4-melanogaste    AGTTATGCGT AGAAGAAAAC GTATAAAAGT GAATAAATGT TTGCAAACGG ATGTTTAAAA CATGCGTTTT
rox1-sub4-simulans       AGTTATGCGA TGAAGAAAAC GTATAAAAGT GAATAAATGT TCTCAAACGG ATGATTAAA- CATGCGTTTG
rox1-sub4-sechellia      AGTTATGCGA TGAAGAAAAC GTATAAAAGT GAATAAATGT TATCAAACGG ATGATTAAA- CATGCGTTTG
rox1-sub4-erecta         AGTTATGCGA AGAAGAAAAG GAAGGAATGT CAAGAAATGT TTGCAAACGG ATGATAAAAA CATGCGTTTC
rox1-sub4-yakuba         AGTTATGCGA AGAT--AAAC AAAGAAAAGT GAAGAAATGT TTGCAAATGG ATATCAAAAA CATGCGTTTT
Clustal Consensus        *********   **   ***    *  ** **  ** ****** *  **** ** **    ***  *********

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            145        155        165        175        185        195        205
rox1-sub4-melanogaste    CTAAACGAAA GATGAA-AAA TAATATCAAA AAGCGAATTA AGTGCTGTTC AATCGTTTGA CCAGCAGTTG
rox1-sub4-simulans       -TAAACGAAT GATGAA-AAA TAATATCAAA AAGCGAATTA AGTGCTGTTC AATCGTTTGT CCAGCACTTG
rox1-sub4-sechellia      -TAAACGAAT GATGAA-AAA TAATATCAAA AAGCGAATTA AGTGCTGTTC AATCGTTTGT CCAGCACTTG
rox1-sub4-erecta         GTGAACGAAA GAAATCCAAA AAATATTAAA A-GCGAATTA AGTGTTGTTC AATCGTTTGG CCAACATTTG
rox1-sub4-yakuba         GTGAACAAAA GAAGTACAAA AAATATTAAA AAGCGAATTA AGTGTTGTTC AATCGTTTGC CCAACAGTTG
Clustal Consensus         * *** **  **     ***  ***** *** * ******** **** ***** *********  *** ** ***

                         ....|....| ....|....| ....|....| ....|....| ....|...
                            215        225        235        245        255
rox1-sub4-melanogaste    ATTTGCGAAT GGCTGAAAAT ATTCGCTTAA ACTTATGGCA AAGCTTGT
rox1-sub4-simulans       ATTTGCGAAT GGCTGAAACT ATTCGCTTAA ACTTATGGCA AAGCTTGT
rox1-sub4-sechellia      ATTTGCGAAT GGCTGAAACT ATTCGCTTAA ACTTATGGCA AAGCGTGT
rox1-sub4-erecta         ATTTGAGAAT GGCTGAAAAT ATTCGCTTAA ACTTATGTCA AAGCTTGT
rox1-sub4-yakuba         ATTTGCGAAT GGCTGAAAAT ATTCGCTTAA ACTTTAGTCA AAGCTTGT
Clustal Consensus        ***** **** ******** * ********** ****  * ** **** ***
```

E. The alignment of sub5- rox1 RNAs among Drosophila Species Clustalw (1.83)

```
                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            5         15         25         35         45         55         65
rox1-sub5-melanogast     ---TTAAAAT AAA---AAAT AAAAAATTTT --AAAAGAAA TAAAAATATC CTTACACCGG GCACCTACAA
rox1-sub5-simulans       ---TTAACAT AAA---AAAT AAAAAATTA- --AAAAAAAA TATAAATATC CTTACACCGG GCACCTACAA
rox1-sub5-sechellia      ---TTAACAT AAA---AAAT AAAAAATTA- --AAAAAAAA TAAAAATATC CTTACACCGG GCACCTACAA
rox1-sub5-erecta         ATATTCGCTT AAACTTATGT CAAAGCTTGT ACAAATAAAA AAG----ATC CTTCTACTGG GCACCGACGA
rox1-sub5-yakuba         --------TT AAACTTTAGT CAAAGCTTGT TAAAATAAAA GAAGAACATC CTTATACTGG GCGCCTACAA
Clustal Consensus                * *** *  *** **    *** *** *    *** *** ** ** ** ** ** *

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                            75         85         95        105        115        125        135
rox1-sub5-melanogast     TGCGTATCCA CCCGGGATTC AAGCAAAAGC TATTATTCTA TTGGCCTTGA TTATTAACAT ATTTCTGTGC
rox1-sub5-simulans       TGCGTTTCCA CCCGGGATTC AAGCAAAAGC TATTATTCTA TTGGCCTTGA TTATTAACAT ATTTCTATGC
rox1-sub5-sechellia      TGCGTTTCCA CCCGGGATTC AAGCAAAAGC TATTATTCTA TTGGCCTTGA TTATTAACAT ATTTCTATGC
rox1-sub5-erecta         TGCGTTTTCA CCCGGGATTT AAGCAAAAGC TATTAGTCTA TTGGCCTTGA TTATTAACCT ATTTCTATGC
rox1-sub5-yakuba         TGCGTTTACA CCCGGGATTT AAGCAAAAGC TATTAGTCTA TTGGTCTTAA TTATTAACCT ATTTCTATGC
Clustal Consensus        ***** * ** *********  ********** ***** **** **** *** * ******** * ****** ***

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                           145        155        165        175        185        195        205
rox1-sub5-melanogast     TGTGTGTTCT TTCAGATGAG TTTTGAACAG AGACGAAACT TGGAGCCTCG CGGAGCCAAA GTTTTAAATT
rox1-sub5-simulans       TGTGTGTTCT TTCAGATGAG TTTTGAACAG AGACGAAACT TGGAGCCTCG CGGAGCCAAA GTTTTCAATT
rox1-sub5-sechellia      TGTGTATTCT TTCAGATGAG TTTTGAACAG AGACGAAACT TGGAGCCTCG CGGAGCCAAT GTTTTCAATT
rox1-sub5-erecta         TGTGTGTTCT TTCAGATGAG TTTTGAACAG AGACGAAACT TGGAGCCTCG CGGAGCCAAA GTTTTCAATT
rox1-sub5-yakuba         TGTGTGTTCT TTCAGATGAG TTTTGAACAG AGACGAAACT TGGAGCCTCG CGGAGCCAAA GTTTTGAATT
Clustal Consensus        ***** **** ********** ********** ********** ********** *********  ***** ****

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                           215        225        235        245        255        265        275
rox1-sub5-melanogast     GATTTTTGTT TTTAATTCCC CATATTCGAG TCGATTTCCC GCAGTCTTTC ATT-----AA TTTTCTTAAA
rox1-sub5-simulans       GATTTTTGTT TTTAATTGCC TGTATTCGAG TCGATTTCCA TTAATTTTCC ATTTAATTAA TATTCTTAAA
rox1-sub5-sechellia      GATTTTTGTT TTTAATTGCC TGTATTCGAG TCGATTTCCA TTAATTTTCC ATTTAATTAA TATTCTTAAA
rox1-sub5-erecta         GTTTTTTCTT TTTAATTGCC TGTATTCGAG TCGATTTCCA GCACCCTTTC ATA-----AG TTTTCTTAAA
rox1-sub5-yakuba         GATTTTT--T TTTAACTGCC TGTACTCGAG TCGATTTCTA GCAACCTTTC ATA-----AG TTTACTTTAA
Clustal Consensus        * *****  * ***** * **   ** ***** ********     *   ** * **      *  * * *** **

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                           285        295        305        315        325        335        345
rox1-sub5-melanogast     CATTGACTGG CTAATTCTTG AACCTCAGTT TTCATTCAAA AGTCATCTTT TCTTCGTTGG TCCCTGCTTG
rox1-sub5-simulans       CATTAGCCGA CTAATTCTTG AACCTAAATT TCCATTCAAA AGTCATCTTT TCTTCGTTGG TCCCTGCTTG
rox1-sub5-sechellia      CATTAGCCGA CTAATTCTTG AACCTAAATT TCCATTCAAA AGTCATCTTT TCTTCGTTGG TCCCTGCTTG
rox1-sub5-erecta         CATTCGCTGG CTTATTCTTG AATCTGAGTT TCCATTCAAA AATCATCTTT TCTTCGTTGG TCGGTGCTTG
rox1-sub5-yakuba         CATTTGCTGG CTAATTCTTG AATCTCAATT TCCATTCAAA AGTCATATTT TCTTCGTTGG TCCCTGCTTG
Clustal Consensus        ****  * *  ** ******* ** ** * ** * ******** * **** *** ********** **  ******

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                           355        365        375        385        395        405        415
rox1-sub5-melanogast     AGCTTGAAAC ATTGTGCCAA GAATTTTTTT CTGTTCAATT GAACCTTTTC TGTGCAAATG TCCTGCAGAA
rox1-sub5-simulans       AGCTTGAAAC ATTGTGCCAA GAATTTTTT- CTGTTCAATT GAACCTTTTC TGTGCAAATG TCCTCCAGAA
rox1-sub5-sechellia      AGCTTGAAAC ATTGTGCCAA GAATTTTTT- CTGTTCAATT GAACCTTTTC TCTGCAAATG TCCTGCAGAA
rox1-sub5-erecta         AGAT-GAAAC ATTGTGCCAA GTATTTTTT- CTGTTCAATT GAACTTTATC CGTGCAAATG TCCTGCAGGA
rox1-sub5-yakuba         AGCT-GAATC ATTGTGCCTA GTAATTTTC- CTGTTCAATT GAACCTTTTC TGTGCTAATG TCCTACAGGA
Clustal Consensus        ** * *** * ******** * * * ****   ********** **** ** **   *** **** **** *** *

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                           425        435        445        455        465        475        485
rox1-sub5-melanogast     GAGGAAACCC ATTACAGACT TGTGCTTTCT CCTGAATGTG ATATGCAAGC AGGTGTAACA AAGCCTATGA
rox1-sub5-simulans       GAGGAAACCC ATTACAGACT TGTGCTTTCT TCTGAATGTG ATATGCAAGC AGGTGTAACA AAGCCTATGA
rox1-sub5-sechellia      GAGGAAACCC ATTACAGACT TGTGCTTTCT CCTGAATGTG ATATGCAAGC AGGTGTAACA AAGCCTATGA
rox1-sub5-erecta         AGGGAAACCC ATTACAGGCT TGGGCTTTCT TCTGAATGTG ATATGCAAGC AGGTGTAATA AAGCCTATGA
rox1-sub5-yakuba         AAGGAAACCC ATTACAGGAT TGTGCTTTCT CCTGAATGTG ATATGCAAGC AGGTGTAATA AAGCCTATGA
Clustal Consensus          ******** *******  * ** *******  ********** ********** ******** * **********
```

```
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       495        505        515        525        535        545        555
rox1-sub5-melanogast AATCCGGTCC AACCCCACAT CAGGCCATAG CCAAGAAGCT CCGCCTAATA CAAGAGAAGA TTATACTAAC
rox1-sub5-simulans   AATCCGGTCC TACCCCACAT CAGGCCATAG CCAAGAAGCT CCGCCTAATA CAAGAGAAGA TTATACTAAC
rox1-sub5-sechellia  AATCCGGTCC TACCCCACAT CAGGCCATAG CCAAGAAGCT CCGCCTAATA CAAGAGAAGA TTATACTAAC
rox1-sub5-erecta     AATCCAGTCC TACCCCACAT CAGGCCACAG CCAAGAAGCT CCGCCTAATA CAAGAGAAGA TTATACTAAT
rox1-sub5-yakuba     AATTCGATCC TACCCCACAT CAGGCCACAG CCATGAAGCT CCGCCTAATA CAAGAGAAGA TTATACTAAT
Clustal Consensus    *** *  ***  ********* ******* ** *** ****** ********** ********** *********

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       565        575        585        595        605        615        625
rox1-sub5-melanogast CAGTTTGCTA TCTTTCAGAT CCGACCAGAA GTAGATCGTG TTCTGTGAAC T--AACCCCT TCAGTGTTCA
rox1-sub5-simulans   CGTATTGCTA TCTTTCAGAT CCGACCAGAT GTAGATCATA TTCTGTGAAC T--AACCCCT TCTGTGTTCA
rox1-sub5-sechellia  CAGATTGCTA TCTTTCAGAT CCGACCAGAT GTAGATCATA TTCTGTGAAC T--AACCC-T TCTGTGTTCA
rox1-sub5-erecta     CAGATTGCTA TCTTTCAGAT CCGGCCCGAA GTAGATCCTG TTCTA-GAAC T--AACCCCT TCTGTTTACA
rox1-sub5-yakuba     CAGATTGCTA TCTTTCAGAT CTGACCAGAA GTAGATCCTG TTCTATGAAC TCTAACCCCT TCTGTTTTCA
Clustal Consensus    *   ****** ********** * * ** **  ******* *  ****  **** *  ***** * ** ** * **

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       635        645        655        665        675        685        695
rox1-sub5-melanogast GCACCTCGTC AATTGTTCAT AATTGTTCCT TTTATTTTAT GTTGTGTT-- ---ATCAAAT TAACTTCCGT
rox1-sub5-simulans   GCACCTCGTC AATTGTTCTA AAATGTTTCT TTTATGTTAT GTTGTGTTGT GTTATCTAAG TATCTTCGGT
rox1-sub5-sechellia  GCACCTCGTC AATTGTTCTA AATTGTTTCT TTTATGTTAT GTTGTGTTGT GTTATCTAAG CATCTTCGGT
rox1-sub5-erecta     GATTCTCGTC AGTTGTTCTC AATCGTTTCT TTTATGT--- -GTGTATT-- ---ACCCAAG TAACTTCAAT
rox1-sub5-yakuba     GCTTCTCGTC AGTTGTTCTC AATCGTTTAT TTTGTGT--- -GTGTTTT-- ---ACCCAAG A---------
Clustal Consensus    *   ****** * ******   **  ***  * *** * *      *** **      * * **

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       705        715        725        735        745        755        765
rox1-sub5-melanogast TGTATTTTA- CCCAGTCCCC TTCCTTGACT TTCTAATAAT TTTCCATGTT TTGACAGATC CTTTTTTGTC
rox1-sub5-simulans   TGTATTTTA- CCCAGTCCCC TT----GACT TTCTATTAAT TTTCCATGTT TTAACATATC --TTCGTGTC
rox1-sub5-sechellia  TGTATTTTA- CCCAGACCCC TT----GACT TTCTATTAAT TTTCCATGTT TTAACATATC --TTCGTGTC
rox1-sub5-erecta     TTTTTTTTAA CCGAGTCCCC TTCCTATACT TGCTGTTTAT TTTCTATGTT TTAACATAAC --TACGTATC
rox1-sub5-yakuba     --TTTTTTAA CCTG------ ---------T TGCTATTCAT TTTCT-TCTT TTAACATAAC --TCCGTATC
Clustal Consensus      * *****  **                 * * **  * ** ****  * ** ** *** * *   *   * **

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       775        785        795        805        815        825        835
rox1-sub5-melanogast CCACCCGAAT AACCAACCAT ACTATTCTTA TATAAG---- GTTCGTGTTT CGGAAAACGC ATTAAAAGGC
rox1-sub5-simulans   CCATTCGAAT CACCAACCAT ACTATTATTA TCTAAG---- GTTCGTGTTT CGGAAAACGC TCTAAAAGGC
rox1-sub5-sechellia  CAATTCGAAT CACCAACCAT ACTATTATTA TCTAAG---- GTTCGTGTTT CGGAAAACGC TCTAAAAGGC
rox1-sub5-erecta     CCATTCAGAT CACCAAACAT ACTTTTCTTA CCTAAGCTAG ATTTGTATTT CGGAAAACGC ACCAAAAGGC
rox1-sub5-yakuba     CCATTTAAAT CACCAAATAT ACTTTTCCTA CCTAAGCTAG GTTCGTGTTT CGGAAAACGC ACTAAAAGGC
Clustal Consensus    * *     **  *****  ** *** **  **   ****      ** ** *** **********    *******

                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                       845        855        865        875        885        895        905
rox1-sub5-melanogast GTAATTTTAA ATCGTTTTCC GAAATGGGAA TCACATTTAA ACAATATTTT GAACTGCGTA AAACGAATAA
rox1-sub5-simulans   GCAATTTTAA ATCGTTTTCC GAAATGGGAA TCAAGTATAA GCCATATTTT GAACTGCGTA AAACGAATAA
rox1-sub5-sechellia  GCAATTTTAA ATCGTTTTCC GAAATGGGAA TCAAATATAA GCCATATTTT GAACTGCGTA AAACGAATAA
rox1-sub5-erecta     GTAATTTAGA ATCGTTTTCC GAAATGGGAA TCAAAGATAA GCAATATTTT GAACTGCGTA AAATGAATAA
rox1-sub5-yakuba     GTAGTTTTGA ATCGTTTTCC GAAATGGGAA TCAAATATAA GCAATATTTT GAACTGCGTA AAACGAATAA
Clustal Consensus    * * ***  * ********** ********** ***    ***  * ******* ********** *** ******

                    ....|....| ....|....| ....|....| ....|....| ....|....| ..
                       915        925        935        945        955
rox1-sub5-melanogast ATGGAACGGT ATTGAAAGCC TATGCATTCA TTACGGTTCA AGAAGTATAA CT
rox1-sub5-simulans   ATGGAACGGT ATTGAAAGCC TACGCTTTCA TTACGGTTCT AGAAGTATAA CT
rox1-sub5-sechellia  ATGGAACGGT ATTGAAAGCC TACGCTTTCA TTACGGTTCT AGAAGTATAA CT
rox1-sub5-erecta     ATTGAACGGT ATTGAAAGCC TACGCTTTCA TTACGGTTCA TGAAGTATAA CT
rox1-sub5-yakuba     ATGGAACGGT ATTGAAAGCC TAAGCTTTCA TTACGGTTCA AAAAGTATAA CT
Clustal Consensus    ** ******* ********** ** ** **** *********    ******** **
```

F. The alignment of sub6- rox1 RNAs among Drosophila Species Clustalw (1.83)

```
                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             5         15         25         35         45         55         65
rox1-sub6-melanogaster   GAAAAACACA TTTACTAACA AATAAAAACT TGCTGATCAA CGTTCTACGC AGTTCTTAAA AAGATGTTGA
rox1-sub6-simulans       GAAAAACACA TTTACTAACA AATAAAAACT TGCTGATCAA CGTTCTACGC AGTTCTTAAA AAGATGTTGA
rox1-sub6-sechellia      GAAAAACACA TTTACTAACA AATAAAAACT TGCTGATCAA CGTTCTACGC AGTTCTTAGA AAGATGTTGA
rox1-sub6-erecta         GAAAAACACA TTTACTAACA AATAAAAACT TGCTAATCAA CGTTCTACGC AGTTCTTAAA AAGATGTTGA
rox1-sub6-yakuba         ACTAAACACA TTTAATAACA AATAAAAACT TGCTGATCAA CGTTCTACGC AGTTCTTA-A AACATGTTAA
Clustal Consensus           ******* **** ***** ********** **** ***** ********** ******** * ** ***** *

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|
                             75         85         95        105        115        125        135
rox1-sub6-melanogaster   AA-TGAACAC AGCCAAAGCA AGTAAAAATG TGTGGAAACG TTATACGAAT CTTCACCAAG TGCCC
rox1-sub6-simulans       AA-TGAACAC AGCCAAAGCA AGTAAAAATG TGTGGAAACG TTATACGAAT CTTCACCAAG TGCCC
rox1-sub6-sechellia      AA-TGAACAC AGCCAAAGCA AGTAAAAATG TGTGGAAACG TTATACGAAT CTTCACCAAG TGCCC
rox1-sub6-erecta         AA-TGAACAC AACCAATGCA AGTAAAAATG TGTGGAAACG TTATACGAAT CTTCACCAAC TGCCC
rox1-sub6-yakuba         AAATGAATAC AACCAATGCA AGTAAAAATG TGTGAAAACG TTATACGAAT CTTCACCAAG TGCCC
Clustal Consensus        ** **** ** * **** *** ********** **** ***** ********** *********  *****
```

G. The alignment of rox1 RNA among Drosophila species between sub1 and sub2

```
                                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                     5         15         25         35         45         55         65
between-sub1-sub2-rox1-melanog   ---------- ---------- ---------- ---------- ---------- ---------- ----------
between-sub1-sub2-rox1-simulan   GAGAATAGAC ATGGGGTAGT TCACGTACGA GCGTCCCTTC GCTTTTCGAC AAGTGCAGCC CTAATGCCGT
between-sub1-sub2-rox1-sechell   ---------- ---------- ---------- ---------- ---------- ---------- ----------
between-sub1-sub2-rox1-erecta    ---------- ---------- ---------- ---------- ---------- ---------- ----------
between-sub1-sub2-rox1-yakuba    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Clustal Consensus

                                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                     75         85         95        105        115        125        135
between-sub1-sub2-rox1-melanog   ---------- ---------- ---------- ---------- ---------- ---------- ----------
between-sub1-sub2-rox1-simulan   CGTTTTTCGC CGACAGCATT TATGGGTAGT CACCGAAGAG AAGTGTTAGT AACCGGGCCT GCCCTTTAAA
between-sub1-sub2-rox1-sechell   ---------- ---------- ---------- ---------- ---------- ---------- ----------
between-sub1-sub2-rox1-erecta    ---------- ---------- ---------- ---------- ---------- ---------- ----------
between-sub1-sub2-rox1-yakuba    ---------- ---------- ---------- ---------- ---------- ---------- ----------
Clustal Consensus

                                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|
                                    145        155        165        175        185        195
between-sub1-sub2-rox1-melanog   ---------- -TGAAAAAAA AATGACCAAA AAATCGAAAT CT-------- -----
between-sub1-sub2-rox1-simulan   TTAAATTTAA ATTGAGGAAA AATTACCTAA -CATCCAA-T CTTCGGATAT TTCAT
between-sub1-sub2-rox1-sechell   ---------- -----GGAAA AATCACCTAA ACATCCAAAT CTTCGGATAT TTCAT
between-sub1-sub2-rox1-erecta    ---------- -------AAA AATTACTTAA ACATCAAAAT GATCGGTTAT TTCAT
between-sub1-sub2-rox1-yakuba    ---------- --TGAAGGAA AATTACTTAA ACATCCAAAT TGTCGGCTAT TTCAT
Clustal Consensus                                    ** *** **  **   *** ** *
```

H. The alignment of rox1 RNA among Drosophila species between sub2 and sub3

```
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                   5        15        25        35        45        55        65
between-sub2-sub3-rox1-melanog  -------TGT -GGAATTTTT GTTAAATTTT TTTTAAAGTG C-ATTTTTTT T---AAGGTG AAATTAGCTA
between-sub2-sub3-rox1-sechell  TGCCAAGTGT GGAA--TTTT TGTTATATTT TTTTAAAGTG C--GTTTTTT TTTAAAGGTG AAATTAG---
between-sub2-sub3-rox1-erecta   TTACTAGTGT -ACA--TTTT TTTAAAATTT TTTTAAAGTG ACATTTTTTT T---AAGGTG AAATTAGCTA
between-sub2-sub3-rox1-simulan  TGCCAAGTGT GGAA--TTTT TGTTATATTT TTTTAAAGTG C--GTTTTTT T--AAAGGTG AAATTAG---
between-sub2-sub3-rox1-yakuba   TTTTTGGTGT TATAATTTTT TTTTTTTTTT ATTAAAAGTG A---TTTTTT T---AAGGTG AAATTAGCTA
Clustal Consensus                     ***    *  ****   *    ***  ** ******     ****** *   ****** *******

                                ....|....| ..
                                   75
between-sub2-sub3-rox1-melanog  GAGTGTGTAC TG
between-sub2-sub3-rox1-sechell  ---------- --
between-sub2-sub3-rox1-erecta   GAAAATGTAC TG
between-sub2-sub3-rox1-simulan  ---------- --
between-sub2-sub3-rox1-yakuba   GAGAGTGTAC TG
Clustal Consensus
```

I. The alignment of sub1- rox2 RNAs among Drosophila Species Clustalw (1.83)

```
                                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                     5         15         25         35         45         55         65
rox2-sub1-melanogaster           CGTTTAGGTA GCTCGGATGG CCATCGAAAG GGTAAATTGG TGTTACATAT AGCTTTAGAG ATCGTTTCGA
rox2-sub1-sechellia              CGTTTAGGTA GCTCGGATGG CCATCGGAAG GGTAAATTGG TGTTACATAT AGCTTTAGAG ATCGTTTCGA
rox2-sub1-erecta                 CGTTTAAGGT GCTCAGATGG CCATCTAAAG GGTAAATTGT TGTTACATAT AGCTTTAGAG ATAGTTTCGA
rox2-sub1-yakuba                 CGTTTAGGTT GCACAGATGG CCATCTAAAG GGTAAATTGG TGTTA----T AGCTTTAGAG ATCGTTTCGA
Clustal Consensus                ****** *   ** * ***** *****  *** *********  *****    * ********** ** *******

                                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                     75         85         95        105        115        125        135
rox2-sub1-melanogaster           ATCACATTGA TAATCGTTCG AAACGTTCTC CGAAGCAA-- AATCAAGCAA GAGTAACGAT TTCCGCATAG
rox2-sub1-sechellia              ATCACATTGA TAATGGTGCG AAACGTTCTC CGAAGCAA-- AATCA--TAA GAGTAACGAT TTCCGCACAG
rox2-sub1-erecta                 ATCACATTGA TAATCGTGCG AAACGTTCTC CGAAGCAA-- ACTCA--AAC AAGTAACGAT TTCCACACAG
rox2-sub1-yakuba                 ATCACATTGA CAATCATGCG AAACGTTCTC CGAAGCAAAA ACTCA--TAC AAGTAACGAT TTCCGAACAG
Clustal Consensus                **********  ***  * ** ********** ********   * ***   *   ********* ****  * **

                                 ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                    145        155        165        175        185        195        205
rox2-sub1-melanogaster           TCGAAAATGT TTAAGTTGAA TTGTCTTACG GACAGTGAGA TGAGTACGAC TATTTGGAAA TCACAAA-CG
rox2-sub1-sechellia              TCGAAAATGT TTAAGTTGAA TTGTCTTACG GACAGTGAGA TGAGTACGAC TATTTGGAAA TCAAAAAGCG
rox2-sub1-erecta                 TCGAAA-TGT TTAAGTTGAA TTGTCTTACG GACAGTAAGA TGAGTACGAC TATTTGGAAA TCACAAA-CG
rox2-sub1-yakuba                 TCGAAA-TGT TGAAGTTGAA TTGTCTTACG GACAGTAAGA TGAGTACGAC TATTTGGAAA TCACAAA-CG
Clustal Consensus                ****** *** * ******** ********** ****** *** ********** ********** *** *** **

                                 ....|....| ....|....| ....|....| ....|....| ....|....| .
                                    215        225        235        245        255
rox2-sub1-melanogaster           AATTGTTTTC ATGGTTGACG CGCTTGTCAA GCTACAAAAC AAAATGAATG A
rox2-sub1-sechellia              AATTGTTTTC ATGGTTGACA CGTTTGTCAA GCTTCAAAAC AAAATGAATG A
rox2-sub1-erecta                 AATTGTTTTC ATGGTTGACA CGCTTGCCAA GCTGCAAAAC AAAATGAAAG A
rox2-sub1-yakuba                 AATTGTTTTC ATGATTGACA CGCTTGTCAA GCTGCAAAAC AAAATGAAAG A
Clustal Consensus                ********** *** *****  ** *** *** *** ****** ******** * *
```

J. The alignment of sub2- rox2 RNAs among Drosophila Species Clustalw (1.83)

```
                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             5         15         25         35         45         55         65
rox2-sub2-melano         TCTTGGAACG CAACATTGTA CAAGTCGCAA TGCAAACTGA AGTCTTAAAA GACGTGTAAA ATGTTGCAAA
rox2-sub2-sechellia      TCTTGGAACG CAACATTGTA CAAGTCGCAA TGCAAACTGA AGTCTTAAAA GACGTGTAAA ATGTTGCAAA
rox2-sub2-erecta         -TTTAGAACG CAACATTGTA CAAGTCGCAA TGAAAACTAA AGTCTTAAAA GACGTGTAAA ATGTTGCAAA
rox2-sub2-yakuba         TCTATGAAGG CAACATTGTA CGAGTCGCAA TGAGAACTAA AGTCTTAAAA GACGTGTAAA ATGTTGCAAA
Clustal Consensus          *  *** * ********** * ******** **  **** * ********** ********** **********

                         ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                             75         85         95        105        115        125        135
rox2-sub2-melano         TTAAGCAAAT ATATATGCAT ATATGGGTAA CGTTTTACGC GCCTTAACCA GTCAAAATAC AAAATAAATT
rox2-sub2-sechellia      TTGAGCAAAT ATATATGCAT ATATGGGTAA TGTTTTACGC GCCTTAACCA GTCAAAATAC TAAATAAATT
rox2-sub2-erecta         TTAAGCAAAT ATATATGCAT ACATGGGTAA CGTTTTACGC GCCTTAACCA GTCAAAATAC TAAA--AATG
rox2-sub2-yakuba         TTAAGCAAAT ATATATGCAT ATATGGGTAA CGTTTTACGC GCCTTAACCA GTCAAAACAC TAAATAAATT
Clustal Consensus        ** ******* ********** * ********  ********* ********** ******* **  ***  ***

                         ....|....| ....|....| ....|....| ....|....| ....|..
                            145        155        165        175        185
rox2-sub2-melano         GGTAAATTTC ATATAACTAG TGAAATGTTA TACGAAACTT AACAATT
rox2-sub2-sechellia      GGTAAATTTC ATATAACTAG TGAAATGTTA TACGAAACTT AACAATT
rox2-sub2-erecta         TATAAATTTC ATATAACTAG TGAAATGTTG TACGAAACTT AGCAATT
rox2-sub2-yakuba         GGCAAATTTC ATATAACTAG TGAAATGTTA TACGAAACTT GCCAATT
Clustal Consensus           ******* ********** *********  **********   *****
```

K. The alignment of rox2 RNA among Drosophila species before sub1

```
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
                                   5         15         25         35         45         55         65
roX2-before-sub1-melanogaster   -----TGTTG CGGCATTCG- CG-GC-CTGG TCACACTAAG CTAGGGCTAC TTTTTATATC ATAAGTCGAG
roX2-before-sub1-sechellia      --TTGAATTA GATTATCCGA CACGTTGCGC TCACGACAAT CCTAATTCCA TCCTCTTCTC TTTCGT----
roX2-before-sub1-erecta         CCTAATTCCA CAG-ATGGG- CGCGT-GTGG GAGAAGCGTC CTTGGG--AC TGGCGACGGA TT--GTGGAG
roX2-before-sub1-yakuba         --AAGAAATA CAATACAAAA AGCAAATCTA TGAAGGCAAC ATTGTACGAG TCGCAATGAG A---ACTAA-
Clustal Consensus                           *                                       *

                                .

roX2-before-sub1-melanogaster   C
roX2-before-sub1-sechellia      -
roX2-before-sub1-erecta         C
roX2-before-sub1-yakuba         -
Clustal Consensus
```

L. The alignment of rox2 RNA among Drosophila species between sub1 and sub2

```
                                ....|....| ....|....| ....|....| ....|....| ....|....| ....|...
                                   5         15         25         35         45         55
between-sub1-sub2-rox2-melano   TATACAATAT ACAATATACA ATATGCAATA CAATACAATA CAAGACAA-A AAAATGTG
between-sub1-sub2-rox2-sechel   ---------- ----AATACA ATACACAATA CTATATAATA CAAGACAAGA AAAATGTA
between-sub1-sub2-rox2-erecta   ---------- ---------- ----GAAATA CAATACAAGA -AAAATAT-- ---ATTT-
between-sub1-sub2-rox2-yakuba   ---------- ---------- ----GAAATA CAATACAA-- -AAAGCAA-- ---ATC--
Clustal Consensus                                       **** * *** **    **   *       **
```