Biology Master's Theses   Department of Biology

11-2009

Defining the Limits and Mapping the Control Sequences of the Gene, Enhancer of Rudimentary, in Drosophila melanogaster

Nicholas Rizzo
The College at Brockport

Follow this and additional works at: http://digitalcommons.brockport.edu/bio_theses

Part of the Genetics and Genomics Commons

Defining the Limits and Mapping the Control Sequences of the Gene, *enhancer of rudimentary,* in *Drosophila melanogaster*

A Thesis

by

Nicholas Rizzo
SUNY Brockport
350 New Campus Drive
Brockport, New York 14420

THESIS DEFENSE

APPROVED	NOT APPROVED	MASTER'S DEGREE ADVISORY COMMITTEE

APPROVED   **NOT APPROVED**   **MASTER'S DEGREE ADVISORY COMMITTEE**

X _____

[signature]   11/12/09
Major Advisor   Date

X _____

Laurie B. Cook   11/12/09
Committee Member   Date

X _____

[signature]   11/12/09
Committee Member   Date

[signature]

Chairman, Graduate Committee

[signature]

Chairman, Dept. of Biological Sciences

Table of Contents

ABSTRACT

The *enhancer of rudimentary* gene, *e(r)* is a highly conserved gene located on the X chromosome in *Drosophila melanogaster*. The gene encodes a transcriptional co-factor of 104 amino acids in length, and although the exact function is still unclear it has been shown to be involved in pyrimidine biosynthesis, oogenesis, and the cell cycle. Mutations in *e(r)* are characterized by low viability and decreased fertility. They have also been shown to enhance phenotypes caused by mutations in two genes involved in neurogenesis, *notch* and *deltex*, suggesting a possible role in neurogenesis. In the present study, transgenes of *e(r)* were used to map the sequences necessary for normal expression. All of the sequences necessary for normal *e(r)* expression mapped to a 2,618 bp fragment that included the transcribed region plus 787 bp upstream of the start of transcription. The two introns of *e(r)* were shown not to contain transcription control sequences. These studies also mapped a transcription enhancer to a 5' region from -338 to -787 and a transcription silencer to a 5' region from -45 to -338. Along with the transgene study, deletions caused by P-element excision, exclusively located in the control region of *e(r)* were isolated and analyzed to determine effect on gene activity. The further refined region was then used with the EvoPrinter program, which uses the 12 other Drosophila species as a comparison to determine any conservation of these sequences and TESS which uses input sequence to look for possible transcription factor binding sites.

INTRODUCTION

Gene Expression and Regulation

With rare exceptions, every cell of an organism contains the same genetic information as the others. However, cells differentiate into various types, which ultimately differ in structure and function. This process is governed by the expression of certain genes, and the repression of others. Possible reasons for gene regulation are tissue-specificity, environmental cues, or stages of differentiation in development. Gene regulation is essential for viruses, prokaryotes and eukaryotes as it increases the versatility and adaptability of an organism by allowing the cell to express protein when needed[1]. Gene regulation is achieved through various methods such as chemical and structural modification of chromatin, altering the stability of mRNA molecules, and altering the ability of a ribosome to translate mRNA to protein[1]. The most widely used way, is altering the rate of transcription of a particular gene, in other words controlling the expression of the DNA itself. This can be done in several ways, mainly through the actions of transcription factors such as activators and repressors. These transcription factors will bind to specific sequences of DNA that are sometimes located upstream from the coding region. This region is termed the 5' regulatory region, or simply the control region. They may also bind to regions downstream of the coding region or within the gene itself such as with an intron[2]. Enhancers are regions of DNA that bind activators and allow normal or increased expression of a gene. Silencers are regions sequences of DNA that bind repressors and decrease or silence the expression of a gene. Together with the appropriate transcription factors, these regulatory elements allow for varying levels of gene expression throughout different tissues and different stages of development.

Drosophila Genetics

Drosophila melanogaster, the common fruit fly, is also one of the most valuable of organisms in biological research, particularly in genetics and developmental biology. Drosophila has been used as a model organism for research for almost a century, and today, several thousand scientists are working on many different aspects of the fruit fly. Some of the reasons people work on it are the extensive knowledge base of the fruit fly and its ease of handling. A more simple reason is the organism's practicality: it is a small animal, with a short life cycle of just two weeks, and it is cheap and easy to keep large numbers. Mutant flies, with defects in any of several thousand genes are available, and the entire genome has recently been sequenced[4].

Drosophila melanogaster has three pairs of autosomal chromosomes, an X and a Y chromosome. There are two arms to each of the three autosomes (2, 3 and 4, although the fourths right arm is tiny) referred to as L and R, for left and right. Each arm of the chromosomes is also numbered in cytogenic units, which allows the investigator to know expected recombination frequencies between two genes on the same chromosome arms. Genes are identified by numerical location or by recombination units. Determination of sex is based on the ratio of X chromosomes to autosomes; females having two X chromosomes (ratio of 1) and males having a single X (ratio of 0.5). Recombination only occurs in the females, which is a quality that geneticists exploit in certain crosses. The life cycle of the fruit fly is very simple; eggs are laid which develop into larvae, and the larvae develop into pupae, and then into adult flies. The normal generation time, with flies being stored at 25°C is 10 days.

There are also some simple rules for the nomenclature used in Drosophila genetics. Chromosomes are written in the order of X/Y; 2; 3; 4 and genotypes are only listed when a mutation is present and are italicized. Recessive mutations are lower case (i.e. *w* for the white

gene) and dominant mutations are capitalized (i.e. *Cy* for curly). Particular alleles are superscripted; if more than one mutation on a chromosome are present they are listed in order from left to right corresponding to left and right arms; and if mutations are homozygous they are written once, and if they are heterozygous than the wild-type allele is denoted as a plus-sign (e.g. *y, w; y w/+ +* respectively).

The Drosophila model system also has very powerful tools that make it such a useful and effective system. Examples of these tools are balancers, phenotypic markers and non-recombination in males. Balancers are chromosomes containing multiple inversions within the same chromosome, dominant markers, several recessive markers, and lethality or sterility as homozygotes. There are balancers for each Drosophila chromosome except for chromosome 4 because of its small size. The inversions in the balancers do not allow homologous recombination to occur, which allows researchers to follow a mutation throughout successive generations without losing it to recombination. They also allow the maintenance of deficiencies that would otherwise be lethal. The phenotypic markers are another factor that makes the system such a valuable one to geneticists. There exist numerous markers that affect eye phenotypes, body phenotypes, bristle phenotype, larval phenotype, wing phenotype and more. These markers allow the production of homozygous recessive stocks through multiple generation crosses. Lastly, non-recombination in males allows a mutation to be maintained in a non-balanced state[9].

P Element Transformation

An extremely powerful technique using the Drosophila model system with a battery of applications is P-element transformation. With the use of P-elements, one can reintroduce a wild-type copy of a gene into a null or mutant background, or even introduced altered genes or tagged

genes such as GFP labeled. P-elements are a family of transposable elements exclusive to Drosophila that can randomly insert themselves into the genome in the presence of an enzyme called a transposase. P-elements are constructed without the transposase gene, but are co-injected with another DNA construct containing the transposase gene. Phenotypic markers are also used here to select transformants; the aforementioned markers such as eye color and body color being the most common. These two constructs are injected into the posterior pole of the embryo where the germ line will develop, which will ensure the generation of a transmissible transformant. The flies that develop are backcrossed to the injected stocks several times to ensure selection of germline transformants. These transformants are then crossed to a double balancer to produce a homozygous stock.

P-elements can also create deletions in the gene-of-interest by inducing their excision. Work has been done in Drosophila where the goal was to insert P-elements throughout the genome specifically in the upstream or coding region of as many genes as possible. This allows researchers to obtain stocks in which there gene-of-interest contains a P-element insertion, and with the introduction of transposase containing P-element, can cause the excision of the P-element. When the excisions occur, many times adjacent sequences of DNA from the insertion site are excised along with the P-element. This results in deletions of various sizes, and with many stocks of these flies, one can generate and analyze many different mutant alleles for their gene. There are things to consider when using this method such as remnants of the P-element in your gene after excision and also inversions, which may be difficult to detect[6].

The *enhancer of rudimentary* gene – *e(r)*

The *enhancer of rudimentary* gene, *e(r)*, is a gene located on the X chromosome in *Drosophila melanogaster*, that encodes a transcriptional co-factor of 104 amino acids in length. It was discovered as a recessive enhancer of a wing phenotype of a mutation in the *rudimentary* gene, *r*, in Drosophila[3,4]. Although exact function has not been revealed, it has been shown to act as a transcriptional co-factor[5]. Studies on the structure of ER have shown a hydrophobic and hydrophilic portion to the protein, which is indicative of protein-protein binding at the hydrophobic region[6]. ER has been localized to the nucleus in *Xenopus*, binding to the transcription factor DCoH/PCD (dimerization cofactor of HNF1/pterin-4alpha-carboinolamine dehydratase)[5]. The ER protein has also been localized to the nucleus in Drosophila and human cells, furthering the proposal of its role as a transcriptional co-factor[7].

The *r* gene encodes the first three activities in pyrimidine biosynthesis and shows a recessive wing truncation in *r* mutants. This mutant wing phenotype is sensitive to changes in *r* expression and is useful in determining involvement of cis and trans-acting regulatory mutations[3]. This genetic interaction shows possible involvement in the cell cycle for the *e(r)* gene. There are several other genes which *e(r)* also shows a genetic interaction with. *Notch* is a highly conserved gene throughout all metazoa, which encodes a single-pass trans membrane receptor protein involved in a signaling pathway that is critical for nerve cell specification as well as other developmental processes. Upon ligand binding, the intracellular domain of the *notch* receptor is cleaved and translocates to the nucleus where it will bind to other transcription factors and induce gene expression. A mutant allele of *notch*, N^{nd-p}, in combination with a deletion in *e(r)*, results in lethality in hemizygous males and also in homozygous females. *Deltex*, a gene that is involved downstream in the *notch* signaling pathway also genetically

interacts with *e(r)*. This interaction is shown through a lethal combination of a *deltex* mutation *dx*, and mutations in *e(r)*. Mutations in *e(r)* are not lethal by themselves; however they do produce a decrease in the viability and fertility. It has been shown that *e(r)* is highly expressed in the ovaries; specifically the nurse cells and suggests possible ovarian function[4]. The decrease in fertility in *e(r)* mutants reflects this possible role. This decrease in viability and fertility is also indicative of a functional importance of *e(r)*. Although mutations are not lethal, in evolutionary terms, mutations in *e(r)* would ultimately lead to a low success rate and a decrease in the ability of the organism to compete. These organisms would most likely be selected against and eventually be eliminated from the population. This gives support and reasoning to the evidence of *e(r)*'s high conservation. There are also two different sized transcripts of *e(r)*, the larger of the two being female specific. These two transcripts differ not only in size, but also by their polyadenylation sites[4]. This female-specific mRNA is found in the nurse cells of the ovaries[4]. This information also gives support to the suggested role in oogenesis of *e(r)*.

The aims of this study are to define the limits of the *e(r)* gene and to map the sequences necessary for normal activity. This will be accomplished by two different approaches. The first approach involves the use of a set of different transgenes, where sequences containing the *e(r)* gene will be used to rescue or restore deficiencies caused by mutations in *e(r)*. The second approach will utilize the excision of a P-element to create a deletion in the control region of *e(r)* and assess whether expression remains normal. Also, using bioinformatic tools, attempt to determine what possible transcription factors may be involved in the regulation of *e(r)* and their respective binding sites.

MATERIALS and METHODS

Crosses and Phenotypic Assessment

Crosses were set up in Drosophila vials with 5-7 males and 6-10 females. Vials contained food source composed of yeast, agar, glucose, water and mold inhibitor. Adult flies were transferred to new vials every 3 days. Flies were observed under a dissecting microscope to view phenotype. The first assessment was performed in an attempt to rescue a lethality of a mutant notch allele - N^{nd-p}, and a deletion in $e(r)$ - $Df(1)e(r)37-6$. The females used in these crosses have a mutation in the yellow gene, y, resulting in a yellow body color; a mutation in the white gene, w, resulting in a white eye color; a deletion of 1406 bp in the $e(r)$ gene, $Df(1)e(r)37-6$; a mutation in notch, N^{nd-p}; and carry a balancer chromosome $FM7c$ which contains a Bar mutation as a phenotypic marker. The resulting genotype of these females is $y\ w\ N^{nd-p}\ Df(1)e(r)37-6/FM7c$. These females were used in eight different crosses with males each of which were wild-type but also carries a different autosomal transgene containing varying sequences of $e(r)$ (Fig. 1). These transgenes are marked with a w^+ gene as a phenotypic marker. $Tr[w^+\ e(r)^+]^{SS}$ males carry a transgene of approximately 6.1 kb from a SalI-SalI restriction sites that has been previously been shown to be sufficient for $e(r)$ expression. $Tr[w^+\ e(r)^+]^{cDNA}$ males carry a transgene from the 6.1 SalI-SalI restriction sites that excludes the two introns of $e(r)$. $Tr[w^+\ e(r)^+]^{ES}$ males carry a transgene of approximately 2.3 kb from an EcoRV-SalI restriction site. $Tr[w^+\ e(r)^+]^{SA}$ males carry a transgene from a SalI-AflII restriction sites that is missing the distal polyadenylation signal. $Tr[w^+\ e(r)^+]^{PS}$ males carry a transgene of approximately 3 kb from a PvuII-SalI restriction sites. Two different stocks carrying this transgene were used: PS-8 and PS-14. $Tr[w^+\ e(r)^+]^{HS}$ males carry a transgene of approximately 2.6 kb from a HinDIII-SalI restriction site.

Three different stocks of flies carrying this transgene were used: *HS*-2, *HS*-15, and *HS*-20. Males that picked up the X chromosome from the female containing the *Df(1)e(r)37-6* deletion and *notch* mutation, along with the transgene were counted along with number of females as a control. Crosses were then set up to obtain similar results in rescuing the N^{nd-p} *Df(1)e(r)37-6* lethality in females. Females, *y w* N^{nd-p} *Df(1)e(r) 37-6/FM7c* were crossed with males obtained from the previous crosses carrying the transgene, *y w* N^{nd-p} *Df(1)e(r)37-6, Tr[w⁺ e(r)⁺]*. Females progeny which were homozygous for the X chromosome carrying the *Df(1)e(r)37-6* deletion and N^{nd-p} mutation and also picked up the transgene were counted. These females' genotype is *y w* N^{nd-p} *Df(1)e(r)37-6/y w* N^{nd-p} *Df(1)e(r)37-6, Tr[w⁺ e(r)⁺]*.

Crosses were set up in an attempt to rescue the lethality of the *Df(1)e(r)27-1* deletion in *e(r)*, which spans 1,188 bp, combined with the deltex mutation, *dx*. Males heterozygous for *Tr[w⁺ e(r)⁺]* were crossed with *y w Df(1)e(r)27-1 dx/FM7c* females. Males which were hemizygous for the X chromosome containing the *Df(1)e(r)27-1* deletion and *dx* mutation along with their respective transgene, *Tr[w⁺ e(r)⁺]* were counted as well as females as a control.

Crosses were then set up to obtain similar results in rescuing the *dx Df(1)e(r)27-1* lethality in females. Females from the previous crosses, *y w dx Df(1)e(r)27-1/FM7c* were crossed with males that were obtained from the previous crosses carrying the transgene, *y w dx Df(1)e(r)27-1, Tr[w⁺ e(r)⁺]*. Female progeny which were homozygous for the X chromosome carrying the *Df(1)e(r)27-1* deletion, *dx* mutation and also the transgene were counted. These females' genotype is *y w dx Df(1)e(r)27-1/y w dx Df(1)e(r)27-1, Tr[w⁺ e(r)⁺]*.

Crosses were set up in an attempt to restore the viability that is seen in deficiencies of *e(r)*. Females heterozygous for the *Df(1)e(r)27-1* deletion, *y w Df(1)e(r)27-1/FM7c*, were crossed with y w *Df(1)e(r)27-1* males heterozygous for the transgene. The number of female progeny
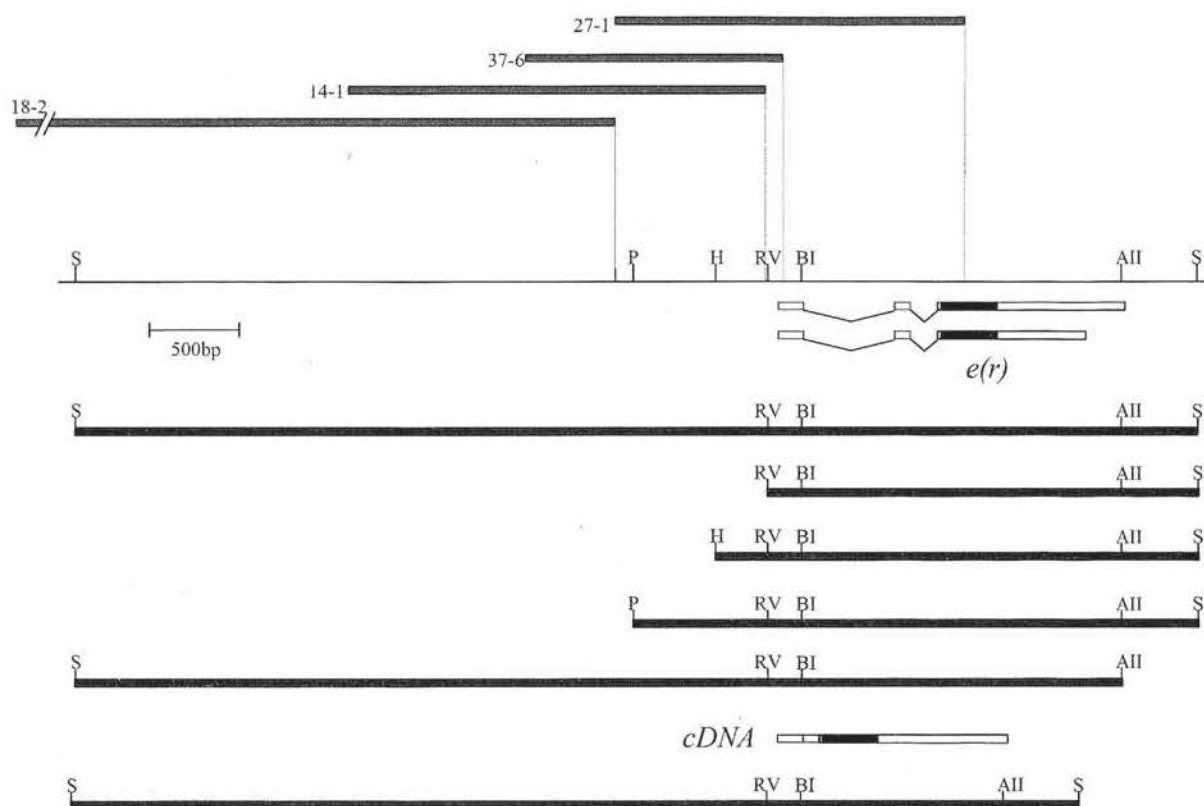
Figure 1. Genomic region containing *e(r)*, deficiencies generated and transgenes. The region is bound by two *SalI* sites (S) that define a 6.1 kb fragment that was previously shown to contain the *e(r)* gene. A *PvuII* (P) site splits the region in two. The short vertical line to the left of the *PvuII* site designates the position of the starting P-element, *P{SUPor-P, y⁺}KG01830*. Further downstream is a *HinDIII* (H) site and even further is a *EcoRV* (E) site. These restriction sites were used to make the *e(r)* transgenes $Tr[w^+ e(r)^+]^{PS}$, $Tr[w^+ e(r)^+]^{HS}$, and $Tr[w^+ e(r)^+]^{ES}$. The transgene $Tr[w^+ e(r)^+]^{SA}$ utilized an *AflII* site in the third *e(r)* exon. The region from this site to the downstream *SalI* site was removed for this transgene, which removes the female-specific polyadenylation signal. The $Tr[w^+ e(r)^+]^{cDNA}$ transgene was constructed from *e(r)* cDNA. Shaded bars at the top represent deficiencies used in this study; lengths and positions indicate the extent of each deficiency. Both *e(r)* transcripts are shown with the exons indicated as boxes and the amino acid coding region designated as black boxes.

heterozygous for the transgene and homozygous for the *Df(1)e(r)27-1* deletion was compared to

females that were homozygous for the deletion and did not obtain the transgene. Number of male

progeny hemizygous for the deletion and heterozygous for the transgene were compared to number of males hemizygous for the $Df(1)e(r)27-1$ deletion that did not obtain the transgene.

Crosses were set up to assess the effect of the different transgenes on the decrease in fertility that is seen in deficiencies of $e(r)$. Females homozygous for the $Df(1)e(r)27-1$ deletion and also heterozygous for the transgene; $Df(1)e(r)27-1/Df(1)e(r)27-1, Tr[w^+ e(r)^+]$ were crossed with w males. As a control, females homozygous for the $Df(1)e(r)27-1$ deletion with no transgene were also used. Total number of male and female progeny from these crosses for each transgene was compared to the control.

DNA Isolations for PCR

Single fly DNA isolations were prepared for amplification by polymerase chain reaction. A single fly was placed in a 1.5 ml tube and mashed with a pipette tip containing 50 μl of squishing buffer, without expelling any of the buffer. The squishing buffer contains 10 mM Tris-Cl pH 8, 1mM EDTA, 25 mM NaCl and 200 μg/ml of proteinase K. The buffer was then expelled from the pipette into the tube. The preps were then incubated in a 37°C water bath for 30 minutes. To inactivate the proteinase K, preps were heated in a heat block to 95°C for 1-2 minutes. Samples were then centrifuged for 5 minutes in a microfuge to pellet the debris. These samples were then stored in -20°C freezer for later use.

Multiple fly DNA isolations were also prepared for multiple reactions in deletion analysis by polymerase chain reaction. Frozen flies were ground in a homogenizer in homogenization buffer at a ratio of 100 flies/ml of buffer. The homogenization buffer consists of 10mM Tris pH 7.5, 60 mM NaCl, 10mM EDTA, and 5% sucrose. The buffer was then autoclaved. Fresh 0.15 mM Spermine HCl and 0.15 mM Spermidine HCl was added fresh to from a 20X solution. After

grinding, the homogenate was transferred to a 2 ml tube and the homogenizer was rinsed with an equal volume of proteinase buffer and transferred to the homogenate and mixed. The proteinase buffer consists of 0.2 M Tris pH 9.0, 30 mM EDTA, 2% SDS, and 5% sucrose. The buffer was then autoclaved. Proteinase K at a concentration of 200 µg/ml was added fresh from a 50X solution. The samples were then incubated in a 37°C water bath for 45-90 minutes. Two consecutive phenol extractions were then performed with an equal volume of neutralized phenol. After each phenol addition, samples were centrifuged for 5 minutes. Top aqueous layers were extracted after centrifugation. Samples were then extracted with an equal volume of chloroform. Two volumes of ethanol along with 1/20 the volume of 4 M NaCl was then added and mixed by inverting. Samples were placed in -20°C freezer overnight. The DNA was centrifuged for 5 minutes and supernatant poured off. Pellet was allowed to dry and then was dissolved in 1 ml of TE (10 mM Tris pH 8.0, 0.1 mM EDTA), 1/20 volume of 4 M NaCl, and 2 volumes of ethanol and reprecipitated. Samples were centrifuged again for 5 minutes and then rinsed with ethanol and allowed to dry. The DNA was re-dissolved in TE at a ratio of 1 fly/3 µl. Samples were stored in freezer until future use.

Deletion Stock Analysis Using Polymerase Chain Reaction

Polymerase chain reaction was performed in the deletion stock analyses using multiple primers (see Appendix C for primer names and sequences.) Total volume of the PCR reaction was 25 µl and consisted of 12 µl of sterile H_2O, 10 µl of Thermo Scientific Master Mix, 1 µl of 5' primer, 1µl of 3' primer and 1 µl of template DNA. Master Mix contains the PCR buffer, the Taq Polymerase, dNTPs, and Mg^{2+}. Some reactions were performed using an extra pair of primers as a control. In these cases 2 µl were subtracted from the volume of H_2O used, which

then became 10 µl. All amplified DNA samples were run in 1% agarose gels. Primary analysis was to determine if the control region was intact. Initial primers used for this determination were 14167 5 Prime 1 and 14167 3 Prime A; but later a more ideal 5' primer replaced this one; ER1 5 Prime 1 (single-fly DNA preps were used for these analyses.) A band of 549 base pairs (14167 5 Prime 1 and 14167 3 Prime A) would be evidence of an intact control region and an absence would lead to further analysis (ER1 5 Prime 1 and 14167 3 Prime A yields a band of 299 base pairs.) All stocks that were shown to have a deletion in the control region were further analyzed to determine if the coding region was also intact. Primers used for this determination were ER-P 5 Prime B and ER-P 3 Prime B which would yield a band of 401 base pairs. Those stocks which had an absence of the control region amplification yet had an intact coding region were further analyzed to localize these deletions.

Different combinations of primer pairs were used to identify deletions in stocks in which the P- element had excised. These primers were used to determine the endpoints of the deletion. Pairs of primers at the 5' end were initially used to amplify these regions. As these regions were confirmed to be intact, successive primer pairs were used to move further downstream until a fragment could not be amplified. These amplifications were repeated to verify that this region was not intact. This process was also done in order to localize the 3' breakpoint of the deletion. Primers that encompassed the deletion were then used to amplify the region containing the deletion. This compared to wild-type gave approximate size of the deletion. Amount of amplified product was determined by comparison to known concentrations of DNA in the bands of the hyperladder (Bioline). Bands in Hyperladder begin at 200 bp and followed by 400 bp, 600 bp, 800 bp, 1 kb, 1.5 kb, 2kb, 2.5 kb, 3 kb, 4 kb, 5 kb, 6 kb, 7 kb, 8 kb, 9 kb, and 10 kb.

DNA Purification

Purification of PCR fragments was performed using a QIAquick PCR purification kit (amount of DNA needed for sequencing was calculated using equation: #base pairs/5.0 = amount of PCR product in ng that is needed.) Contents and required materials included MinElute spin columns, Buffer PBI, Buffer PE, Buffer EB, and 2 ml collection tubes. Pure ethanol was added to Buffer PE before use. PCR reactions were mixed with Buffer PBI at a ratio of 1 to 5 volumes. Sample was then applied to a MinElute column, placed in a 2 ml collection tube and centrifuged for 1 minute. Flow-through was discarded and column was placed back into collection tube. To wash the sample, 750 µl of Buffer PE was added to the column and centrifuged for 1 minute. Flow-through was discarded and samples were centrifuged again for 1 minute at maximum speed. Column was then placed in a 1.5 ml microcentrifuge tube. To elute the DNA, 10 µl of Buffer EB was added to the center of the membrane in the column which was allowed to sit for a minute and then centrifuged for an additional minute.

DNA Sequencing

To prepare samples for DNA sequencing, 3 µl of purified DNA fragments were combined with 1 µl of primer and 14 µl of sterile H_2O, for a total volume of 18 µl. Samples were then sent to Cornell University's DNA Sequencing Facility at 147 Biotechnology Building, Ithaca, New York 14853. DNA sequences were determined using the Applied Biosystems Automated 3730xl DNA Analyzer, Big Dye Terminator chemistry and AmpliTaq-FS DNA Polymerase.

Bioinformatics Using EvoPrinter and TESS

In an effort to narrow down or possibly determine the specific transcription factors that bind to the regulatory regions (silencer and enhancer regions), the Internet tools EvoPrinter and TESS were utilized. Sequences that were shown to contain regulatory regions were entered into a EvoPrinter and compared to the sequence of different Drosophila species one at a time. The Drosophila species that were used are *D.simulans, D.sechelia, D.yakuba*, and *D.erecta*. If conservation in these regions existed among these species, then others such as *D.ananasse, D.pseudoobscura* and *D.persimillis* were also compared. The resulting sequences that EvoPrinter determined are conserved were then entered into TESS, which was used to determine the possible transcription factor binding sites and their respective transcription factors that are bound to these regions.

RESULTS

Lethality Rescue of *Notch* and *e(r)* Mutants

To define the limits of the *e(r)* gene, crosses were set up using several different transgene constructs (Fig. 1). These transgenes were utilized in an attempt to rescue a lethality caused by a deletion in *e(r)*, *Df(1)e(r)37-6*, and a mutation in *Notch*, N^{nd-p} (Table 1). The first crosses set up used wild-type males heterozygous for the transgene and females heterozygous for an X chromosome containing mutations in the *yellow* gene, the *white* gene, the *Df(1)e(r)37-6* deletion and N^{nd-p} mutation. The other X chromosome is a balancer chromosome *FM7c*, which carries a mutation *Bar*, that affects the shape of the eye as a phenotypic marker (*y w Nnd-p Df(1)e(r)37-6/FM7c*.) Males, which have normal eye color and shape and yellow body color, would indicate that these males had picked up the X chromosome containing the *e(r)* deletion and *Notch* mutation and also the transgene; this would also indicate that the lethality had been rescued. The transgene $Tr[w^+ \ e(r)^+]^{cDNA}$ which is the entire 6.1 *SalI* fragment excluding the introns, showed the ability to rescue the lethal combination yielding 88 males (Table X). The transgene $Tr[w^+ \ e(r)^+]^{SA}$, which is the entire 6.1 *SalI* fragment excluding the distal polyadenylation signal, or the sequence from the *AflII - SalI* restriction sites also proved to rescue the lethality, producing 73 males. Two different stocks with the transgene $Tr[w^+ \ e(r)^+]^{PS}$ (*PS-8* and *PS-14*) were shown to rescue the lethality, producing 74 and 125 males respectively. The transgene $Tr[w^+ \ e(r)^+]^{ES}$ was shown to rescue, producing 101 males. Flies carrying the transgene $Tr[w^+ \ e(r)^+]^{HS}$, unlike the other transgene did not produce any transgenic males. In other words $Tr[w^+ \ e(r)^+]^{HS}$ does not show the ability to rescue this lethality; this was reinforced by using three different *HS* transgenic stocks to verify the results (*HS-2, HS-15,* and *HS-20.*)

Table 1 - Attempt to Rescue Lethality of N^{nd-p} Mutation and $Df(1)e(r)37-6$ Deletion Combination in Males

Transgene $Tr[w^+ e(r)^+]^X$	Total Number of Females	Total Number of Males	Ratio: Males/Females
ES	386	101	0.26
SA	317	73	0.23
PS-8	305	74	0.24
PS-14	359	125	0.35
HS-2	254	1	0.003
HS-20	535	0	0.00
HS-15	547	0	0.00
cDNA	522	88	0.17

Females of the genotype $y \, w \, Nnd\text{-}p \, Df(1)e(r)37\text{-}6/FM7c$ were crossed to males carrying one of the above mentioned transgenes; $Tr[w^+ \, e(r)^+]^X$. Number of males and females were counted, with the males indicating whether the transgene in the cross was able to rescue the $Nnd\text{-}p \, Df(1)e(r)37\text{-}6$ lethality. Females were also counted to obtain a ratio that can be used as a measure of the ability of the transgene to restore $e(r)$ activity. The presence of the transgene was monitored by the presence of the w^+ gene on the transformation construct.

To further test these results, crosses were set up to determine if these same transgenes that rescued the lethality in males, would rescue the lethality in homozygous females (Table 2). Females, $y \, w \, N^{nd-p} \, Df(1)e(r)37\text{-}6/FM7c$ were crossed with males obtained from the previous crosses, $y \, w \, N^{nd-p} \, Df(1)e(r)37\text{-}6; \, Tr[w^+ \, e(r)^+]$. Females that picked up both of these X chromosomes and the transgene would prove the ability of the transgene to rescue the lethality. The phenotypes of these females would be yellow bodied, with wild-type eye color and shape; the genotype would be $y \, w \, N^{nd-p} \, Df(1)e(r)37\text{-}6/y \, w \, N^{nd-p} \, Df(1)e(r)37\text{-}6; \, Tr[w^+ \, e(r)^+]$. The crosses involving the transgene $Tr[w^+ \, e(r)^{+SS}]$, composed of the entire 6.1 SalI - SalI fragment which has been previously shown to be sufficient for $e(r)$ expression, was used as a control and showed rescue, producing 116 females. The transgene $Tr[w^+ \, e(r)^+]^{SA}$ also showed the ability to rescue, producing 102 females. The two different stocks of transgene $Tr[w^+ \, e(r)^+]^{PS}$ (PS-8 and PS-14) both showed rescue, producing 108 and 100 females respectively. In the last cross set up, the

transgene $Tr[w^+ \; e(r)^+]^{ES}$ was able to rescue the lethality, producing 68 females. Crosses using the transgene $Tr[w^+ \; e(r)^+]^{HS}$ could not be performed since there were no male progeny with the genotype $y \; w \; N^{nd-p} \; Df(1)e(r)37-6; \; Tr[w^+ \; e(r)^+]^{HS}$ that were produced in the previous set of crosses.

Table 2 - Attempt to Rescue Lethality of N^{nd-p} Mutation and $Df(1)e(r)37-6$ Deletion Combination in Females

Transgene $Tr[w^+ \; e(r)^+]^X$	Total Number of females	Total Number of Males	Ratio: Males/Females
PS-8	108	40	0.37
PS-14	100	49	0.49
SA	102	49	0.48
ES	68	22	0.32
SS	116	39	0.34

Males of the genotype, $y \; w \; N^{nd-p} \; Df(1)e(r)37-6; \; Tr[w^+ \; e(r)^+]$, which were obtained from the previous cross, were crossed with females used in the previous cross; $y \; w \; Nnd-p \; Df(1)e(r)37-6/FM7c$. Number of females and males were counted, with the number of females indicating the transgenes ability to rescue the $Nnd-p$ $Df(1)e(r)37-6$ lethality. High ratios of males:females indicates the ability of all used transgenes to rescue the lethality in females.

Lethality Rescue of *Deltex* and *e(r)* Mutants

The same transgenes were also used in an attempt to rescue a lethality resulting from a deletion in *e(r)*, *Df(1)e(r)27-1*, and a mutation in the gene *deltex, dx* (Table 3). The first crosses used females that were heterozygous for the *e(r)* deletion and *dx* mutation, and the balancer chromosome *FM7c*. The genotype of these females is *ec dx Df(1)e(r)27-1/FM7c*. These females were crossed to hemizygous males carrying a single copy of the transgene; $Tr[w^+ \; e(r)^+]$. Males that obtained the X chromosome containing the *dx* mutation and *e(r)* deletion, and also obtained the transgene, would phenotypically have normal eye shape and color but also have a nick in the tip of the wings. This wing phenotype is a result of the *dx* mutation. The transgene $Tr[w^+ \; e(r)^{+SS}]$ was used as a control, and showed the ability to rescue the *dx Df(1)e(r)27-1* lethality, producing

273 males. The transgene $Tr[w^+ e(r)^+]^{cDNA}$ showed the ability to rescue producing 122 males, which . The transgene $Tr[w^+ e(r)^+]^{SA}$, showed the ability to rescue producing 196 males. Two different stocks with the transgene $Tr[w^+ e(r)^+]^{PS}$, (PS-8 and PS-14) showed the ability to rescue, producing 139 and 209 males respectively. The transgene $Tr[w^+ e(r)^+]^{ES}$ showed the ability to rescue, producing 182 males. Flies carrying the transgene $Tr[w^+ e(r)^+]^{HS}$, showed low levels of rescue compared to the transgenes which rescued the N^{nd-p} Df(1)e(r)37-6 lethality. Three different stocks of $Tr[w^+ e(r)^+]^{HS}$ flies were used in crosses (HS-2, HS-15 and HS-20) producing only 16, 96 and 86 males respectively.

Table 3 - Attempt to Rescue Lethality of *ec dx* Mutation and *Df(1)e(r)27-1* Deletion Combination in Females

Transgene $Tr[w^+ e(r)^+]^X$	Total Number of Females	Total Number of Males	Ratio: Females/Males
ES	281	182	0.65
SA	271	196	0.72
PS-14	265	209	0.79
PS-8	179	139	0.78
cDNA	383	122	0.32
SS	333	273	0.82
HS-2	599	16	0.03
HS-15	341	96	0.28
HS-20	307	86	0.28

Females of the genotype *ec dx Df(1)e(r)27-1/FM7c* were crossed with males carrying one of the aforementioned transgenes. Number of males and females were counted, with the males indicating whether the transgene in the cross was able to rescue the *ec dx Df(1)e(r)27-1* lethality. Females were also counted to obtain a ratio that can be used as a measure of the ability of the transgene to restore *e(r)* activity.

An attempt was then made to rescue this lethality and produce homozygous females in the progeny whose genotype would be *ec dx Df(1)e(r)27-1/ ec dx Df(1)e(r)27-1* (Table 4). Females from the previous cross, *ec dx Df(1)e(r)27-1/FM7c* were crossed with males obtained in the progeny of the previous crosses that picked up a copy of the transgene, *ec dx Df(1)e(r)27-1;*

$Tr[w^+ e(r)^+]$. Females that picked up both X chromosomes containing the *deltex* mutation and

e(r) deletion along with the particular transgene would be proof of rescue.

Table 4 - Attempt to Rescue Lethality of *ec dx* Mutation and *Df(1)e(r)27-1* Deletion Combination in Males

Transgene $Tr[w^+ e(r)^+]^X$	Total Number of males	Total Number of Females	Ratio: Males/Females
ES	101	106	0.95
SA	76	85	0.89
PS-14	71	95	0.75

Males of the genotype *ec dx Df(1)e(r)27-1*; $Tr[w^+ e(r)^+]$, which were obtained from the previous cross, were crossed with females used in the previous cross; ec dx *Df(1)e(r)27-1/FM7c*. Number of females and males were counted, with the number of females indicating the transgenes ability to rescue the *ec dx Df(1)e(r)27-1* lethality. High ratios of males:females indicates the ability of all used transgenes to rescue the lethality in females.

Rescue of Low Viability in *e(r)* mutants

One of the phenotypes resulting from mutations in *e(r)* is a decrease in the viability of these flies. Viability is defined as being capable of living or having attained such form and development as to be normally capable of surviving to adult stage. Similar to the previous experiments, crosses were set up to determine the viability of the *Df(1)e(r)27-1* deletion as well as the ability of the previously used transgenes to rescue the decreased viability in these *e(r)* stocks. Number of progeny that obtain the transgene ($Tr[w^+ e(r)^+]/+$) were compared to progeny which did not obtain the transgene (+/+). Significant difference or similarity between these numbers is indicative of the ability of these transgenes to restore viability in these mutants. The crosses that were set up used females whose genotype was *y w Df(1)e(r)27-1/FM7c* crossed with males whose genotype was *y w Df(1)e(r)27-1*, along with a different transgene for every cross; $Tr[w^+ e(r)^+]$. The transgene $Tr[w^+ e(r)^{+SS}]$ was used as a control, and showed the ability to restore the decreased viability in these *Df(1)e(r)27-1* stocks (Table 5). The number of males in

the progeny, which obtained the transgene, were significantly higher, 233 compared to 37, as was the number of females which obtained the transgene, 258 compared to 14 that did not. Males carrying the transgene $Tr[w^+ \; e(r)^+ \;]^{SA}$ were also more abundant than those which did not obtain the transgene, 194 compared to 24. Females also showed a significant increase in number when obtaining the *SA* transgene. Males carrying $Tr[w^+ \; e(r)^+ \;]^{cDNA}$ also were higher in number than males with only the *Df(1)e(r)27-1* deletion; 113 versus 11. Females also showed increase in number when obtaining this transgene; 112 compared to 12. The transgene $Tr[w^+ \; e(r)^+ \;]^{PS}$ showed the ability to restore viability with the number of males carrying the transgene at 83 as opposed to males which did not pick up the transgene at only 4. Females also showed this dramatic difference with 115 carrying the transgene and only 17, which did not. Crosses involving the transgene $Tr[w^+ \; e(r)^+ \;]^{HS}$ did not show an ability to restore the normal viability to these *e(r)* mutants.

Two different stocks, which contained this transgene, were used in the viability assessment, *HS*-15 and *HS*-20. Neither of these stocks showed significant ability to restore viability as the other transgenes have. The number of males that obtained the *HS*-15 transgene was less than double the number of males in the progeny that did not, 72 vs. 48. Number of females in this cross shows a similar ratio; 89 which obtained the transgene compared to 68 that did not. The *HS*-20 crosses showed a slightly higher ability in restoring viability, however overall numbers were less. The number of transgenic males in the progeny was 44 compared to 24 with a wild-type background. Number of transgenic females in the progeny was 67, compared to 26 with a wild-type background. The last transgene used in these crosses in the attempt to restore viability was $Tr[w^+ \; e(r)^+ \;]^{ES}$. This transgene showed an approximate 10-fold increase in number of progeny that obtained the transgene. Number of transgenic males in the progeny was

113 vs. only 11 in the wild-type background. Number of transgenic females in the progeny was

112 vs. 20 in the wild-type background. These data show the ability of every transgene excluding

$Tr[w^+ e(r)^+]^{HS}$ to restore viability in stocks containing the $Df(1)e(r)27-1$ deficiency in $e(r)$.

Rescue of Low Fertility in *e(r)* Females

The last phenotype that has been seen in females with a mutation in *e(r)* is a decrease in

fertility, the ability to produce offspring. As the previous experiments have attempted, the goal

here is to determine whether the previously used transgenes are capable of rescuing the mutant

e(r) phenotype of decreased fertility. More specifically, what effect if any, do the transgenes

have in females homozygous for the $Df(1)e(r)27-1$ deletion? The crosses that were set up used

Table 5 - Transgenic Restoration of Viability in *Df(1)e(r) 27-1* Background

Transgene $Tr[w^+ e(r)^+]^X$	Total females		Female Ratio: (+/+)/(+/transgene)	Total Males		Male Ratio: (+/+)/(+/transgene)
	+/+	+/transgene		+/+	+/transgene	+/transgene
SS	14	258	0.05	37	233	0.16
SA	39	243	0.16	24	194	0.12
ES	43	239	0.18	36	192	0.19
HS-20	26	67	0.39	24	44	0.55
HS-15	68	89	0.76	48	72	0.67
cDNA	12	112	0.11	11	113	0.1
PS-14	17	115	0.15	4	83	0.05

Females of the genotype *y w Df(1)e(r)27-1/FM7c* were crossed with males of the genotype *y w Df(1)e(r)27-1 Tr[w⁺ e(r)⁺]* in order to test the ability of each transgene to restore the viability decrease via *Df(1)e(r)27-1*. Number of male and female progeny that obtained the transgene were compared to male and female progeny that did not. Ratios of each sex are showed to compare transgenes ability to restore viability. Low ratios show good ability to restore viability; high ratios show poor ability to restore viability.

females, which were homozygous for the $Df(1)e(r)27-1$ deletion and also carried one of the

transgenes fore mentioned. A cross was also set up with females, which did not carry any of the

transgenes being used here (Table 6). Females that were not carriers of any of the transgenes produced 162 females and only 15 males. The transgene $Tr[w^+ e(r)^+]^{SS}$, which contains the information sufficient for $e(r)$ expression, produced 1159 females and 943 males. These two sets of progeny are the comparison points for the remaining transgenes. The transgene $Tr[w^+ e(r)^+]^{SA}$ showed the ability to significantly restore fertility producing 2173 females and 1306 males. $Tr[w^+ e(r)^+]^{ES}$ showed similar abilities of restoring fertility producing 2052 females and 1463 males. Transgene $Tr[w^+ e(r)^+]^{cDNA}$ also restored fertility producing 1653 females and 1217 males, as did Transgene $Tr[w^+ e(r)^+]^{PS-14}$ producing 1849 females and 1425 males. Both of the HS transgenes; $Tr[w^+ e(r)^+]^{HS-15}$ and $Tr[w^+ e(r)^+]^{HS-20}$ showed numbers similar to the $Df(1)e(r)27-1$ deletion stock without any of the fore mentioned transgenes. These two were only able to produce 186 and 125 females, and 39 and 16 males respectively. These HS transgenes prove to be the only set that is unable to restore fertility in $e(r)$ deficient flies.

Table 6 - Effect of Transgenes on Decreased Fertility in *Df(1)e(r)27-1* Mutants

Transgene $Tr[w^+ e(r)^+]^X$	Total Females	Total Males	Ratio: Males/Females
NONE	162	15	0.09
SS	1159	943	0.81
SA	2173	1306	0.6
ES	2052	1463	0.71
HS-15	186	39	0.2
HS-20	125	16	0.13
cDNA	1653	1217	0.74
PS-14	1849	1425	0.77

Females of the genotype *Df(1)e(r)27-1/ Df(1)e(r)27-1,* also carrying one of the different transgenic constructs, $Tr[w^+ e(r)^+]$ were crossed with *w* males to test each transgenes ability to restore fertility in *Df(1)e(r)27-1/ Df(1)e(r)27-1* females.

Isolation of deficiencies in the regulatory region of *e(r)*

P-element mobilization has been shown to create localized deficiencies upon excision. To further refine the data obtained in the transgenic rescues, this approach was used to obtain flies, which contained deficiencies exclusively in the regulatory region of *e(r)* (see Appendix A for scheme summary). Any deficiencies found that extend into the coding region will be disregarded. These deletions will ideally help map the control sequences of *e(r)* further. Flies with potential deletions generated by the mobilization of the P-element, *P{SUPor-P, y+}* which was inserted 895 bp upstream of the transcriptional start site of *e(r)*, were primarily analyzed for a normal sized amplification of the transcriptional start site. At first, the primers 14167 5 Prime 1 and 14167 3 Prime A were used to amplify a sequence which crosses into the coding region, of 549 bp (Fig.2). The presence of this band signifies an intact coding region, and its absence, with a positive control, says that the P-element excision has created a deletion that enters this region. Initially, 155 different stocks were analyzed for a deficiency in the coding region of *e(r)*. A more ideal set of primers was utilized along with a pair for a positive control. The primers ER-P 5 Prime B and ER-P 3 Prime B were used to amplify a 401 bp region extending into the coding region, and the primers Notch 5 Prime B and Notch 3 Prime B were used to amplify a region of the *notch* gene (554 bp), which should always be intact in these stocks. Stocks in which this region was amplified were then saved for further analysis.

These stocks were then used along with the primers ER1 5 Prime 1 and 14167 3 Prime A in a PCR reaction to amplify a fragment spanning the regulatory region and transcriptional start site of 299 bp (Fig.3). These reactions also utilized the Notch primers as a positive control. Any reactions that showed amplification at 554 bp but no amplification of 299 bp would be saved for further analysis.
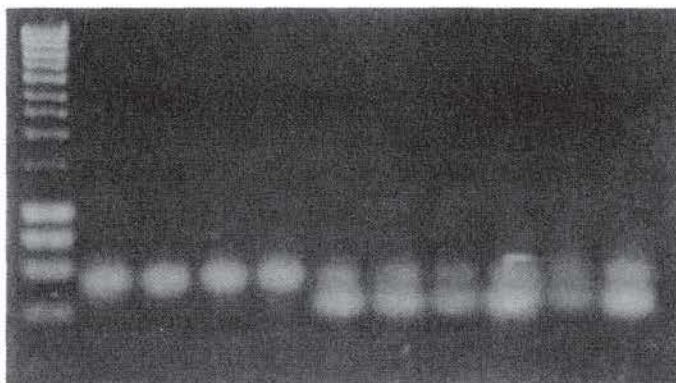
Figure 2. PCR assessing transcriptional start site of *e(r)* for possible deletions. Reaction samples of 20 µl were loaded in 0.9% agarose gels to amplify the region of the transcriptional start site of *e(r)*. Stocks that showed amplification were saved for further analysis.
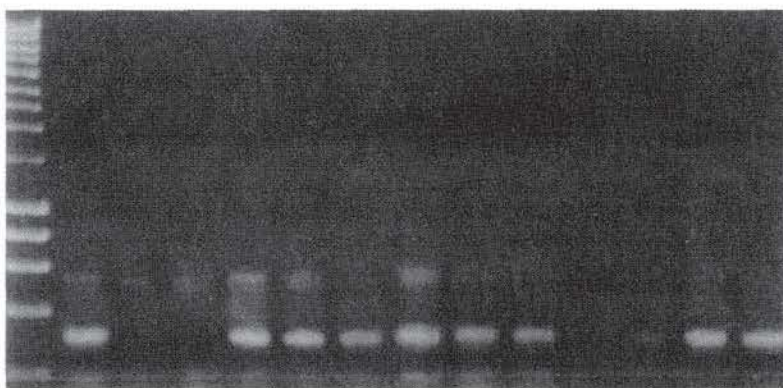
Figure 3. PCR assessing 5' regulatory region of *e(r)* for possible deletions. Reaction samples of 20 µl were loaded in 0.9% agarose gels to amplify the 5' regulatory region of *e(r)*. Stocks that did not show amplification were saved for further analysis.

Out of these two sets of reactions, six different stocks were successfully obtained in which the transcriptional start site was successfully amplified, but not the regulatory region. These stocks were noted as 1-2, 1-9, 16-15, 17-18, and 22-1. The next step in this process was to localize these deletions until their breakpoints were discovered so they could then be amplified and sequenced. The approach to localize the deletions used different sets of primers that begin upstream of where the P-element insertion site is located to localize the 5' breakpoint, and

primers downstream of the P-element site to localize the 3' breakpoint. However, an initial PCR reaction was performed to generally assess the 3' breakpoint of the six aforementioned deletion stocks. The 5' primer ER-P 5 Prime A and 3' primer Mut 1 3 Prime were used to possibly amplify a fragment of 444 bp. Mut 1 3 Prime is located just inside the coding region, if no amplification occurs, then the results will say that the deletions go into the priming site of ER-P 5 Prime A, and if there amplification then it will say that the 3' breakpoint of the deletion does not extend into the priming site for ER-P 3 Prime B. The results from this reaction showed amplification of stocks 1-9 and 22-1. Reactions were then set up for the six stocks using the primer pairs 5 prime S and 3 prime S (1451 bp), 5 prime R and 3 prime R (1453 bp), 5 prime P and 3 prime P (1412 bp), and 5 prime O and 3 prime O (1472 bp). The Notch primers were also used in the reaction as a positive control. Absence of these bands would indicate that the 5' end of the deletion extends at least into the region of the 3' primer site. All reactions produced successful amplifications showing that the deletions do not extend this far upstream. The next reactions used primers downstream of the CG15352 gene; 14167 5 Prime I and ER upstream 3 Prime F, which would amplify a fragment of 608 bp if present. All six stocks including the control were successfully amplified showing that the 5' breakpoint of the deletion does not extend into the ER upstream 3 Prime F site. The next pair of primers, further downstream of CG15352 were 14167 5 Prime H and ER upstream 3 Prime E which amplify a fragment of 703 bp, and again each sample was successfully amplified. The next pair of primers were 14167 5 Prime G and ER upstream 3 Prime D which amplify a fragment of 1139 bp. All stocks showed successful amplification. Approaching closer to the *e(r)* gene, the next primers used were 14167 5 Prime D and 14167 3 Prime D in an attempt to amplify a fragment of 544 bp, amplification was successful in all reactions. Next, the same 5 prime primer was used along with 14167 3

Prime C to amplify a region of 1040 bp, in which all samples were successfully amplified. Primer pairs 14167 5 Prime B and 14167 3 Prime B (558 bp), and 14167 5 Prime A and 14167 3 Prime A (512 bp) were used in separate reactions and amplified all six stocks. At this point all six of these stocks contain deletions which do not enter the 14167 3 Prime A site, which is 295 bp upstream of the P-element insertion site. Because the deletions at this point have been localized to a small genomic region, the next reaction was done in an attempt to encompass the deletion, amplifying and comparing it to the expected wild-type size fragment. The same 5' prime primer was used here, 14167 5 Prime 1, in two reactions with different 3' primers, Mut 1 3 Prime (1134 bp) and ER-P 3 Prime B (1448 bp). The reaction using Mut 1 3 Prime amplified fragments for 1-2 of approximately 400 bp and for 14-1 of approximately 400 bp. The reaction using ER-P 3 Prime B amplified fragments for 1-2 of approximately 700 bp, 14-1 of approximately 700 bp and 17-18 of approximately 300 bp (Fig. 4). These reactions were run again in an attempt to replicate these results, which was successful. Sufficient product must be produced in these reactions to send out for DNA sequencing. The equation used to determine this amount is # of bp/5= ng of DNA needed. For the smallest amplified fragments produced from the deletion stocks, the amount of PCR product needed is approximately 80 ng for both 1-2 and 14-1, and 60 ng for 17-18. The next step, in order to approximate the concentration of amplified fragments, 7 μl of the reaction product was run on a gel along with 1 μl of hyperladder. Since each band of the hyperladder is of known, different concentrations of DNA, it will give a comparison to the amplified bands produced and allow approximation of product. This initial comparison showed the product for 1-2 amplification to be approximately 8ng in 7 μl, for 14-1 approximately 4 ng/7 μl, and for 17-18 approximately 12 ng/7 μl.
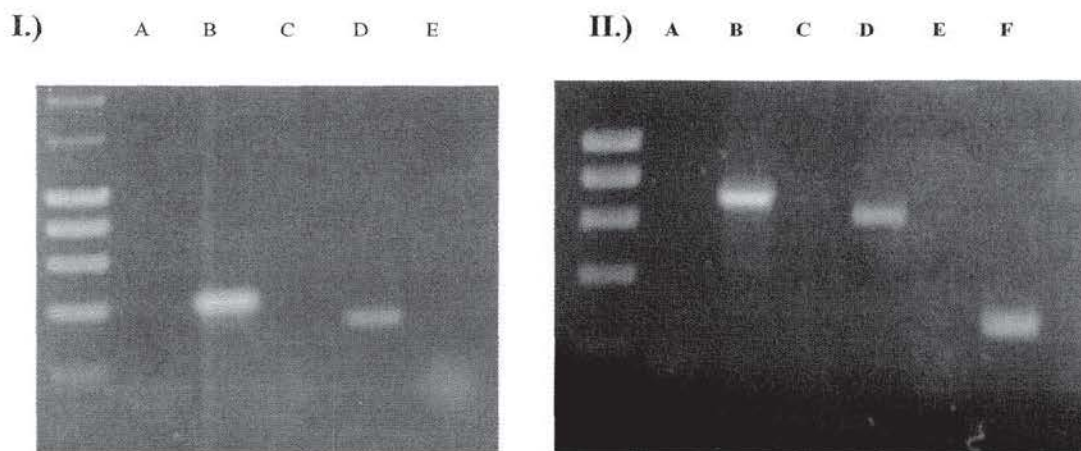
Figure 4. Amplification of regulatory region encompassing deletions. PCR reactions encompassing the deletions in each stock. Reaction samples of 20 µl were loaded in 0.9% agarose gels and compared to wild-type size of amplification. I.) PCR reactions using the primers14167 5 Prime 1 and Mut 1 3 Prime, which results in a fragment of 1134 bp in wild-type. II.) PCR reactions using the primers 14167 5 Prime 1 and ER-P 3 Prime B, which result in a fragment of 1448 bp in wild-type. Both gels were loaded with samples in the order of: molecular weight marker, control, 1-2, 1-9, 14-1, 16-15, 17-18 and 22-1. Gel I: lane A – negative control, lane B – 1-2, lane C – 1-9, lane D – 14-1, lane E – 16-15. Gel II: lane A – negative control, lane B – 1-2, lane C – 1-9, lane D – 14-1, lane E – 16-15, lane F – 17-18.

This amount was still not sufficient for sequencing. One more round of reactions were prepared using a 1 µl aliquot from the previous reaction products as template, from which two reactions were set up for each deletion stock. From these, 3 µl of product were loaded and ran in a 0.9% agarose gel along with 1 µl of hyperladder. Each sample that was loaded displayed a greater intensity and a more robust band than any fragment in the hyperladder. The most concentrated band in the ladder is 20 ng/µl, confirming that along with the previous reactions, enough product is present to send for sequencing (Fig. 5). However, these samples had to be purified to remove any remaining enzyme, primers, nucleotides etc., which may affect the sequencing process. A Qiagen DNA purification kit was used to purify the sample products. After the purification process, 0.5 µl were loaded onto a 1.2% agarose gel along with 1 µl of hyperladder to determine the product recovery in the purification process, and reassure that the amount is still sufficient for DNA sequencing (Fig. 6).
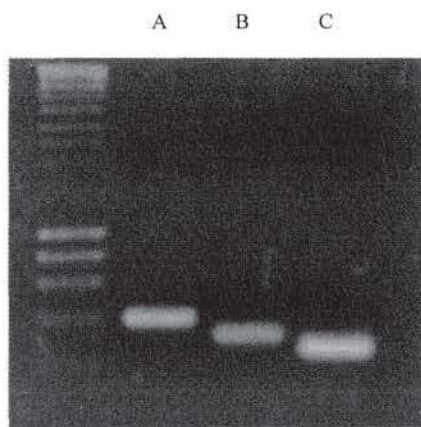
Figure 5. Determination of PCR product concentration. Using 1 µl of hyperladder as a comparison, 3 µl of PCR product was loaded and ran on a 0.9% agarose gel. The amplified products show greater intensity than any band in hyperladder. The most concentrated band in the ladder is 20 ng/µl. This is more than a sufficient amount for DNA sequencing. Lane A – 1-2, lane B – 14-1, lane C – 17-18.

The purified products show bands similar in intensity to the 20 ng/µl band in the hyperladder, which would make the product concentration approximately 40 ng/µl. This was still more than a sufficient amount to send out for sequencing. These three deletion stocks were the prepared for sequencing by mixing 3 µl of the amplified product with 14 µl of dH$_2$O, and 1 µl of primer (Mut 1 3 Prime for 1-2 and 14-1, and ER-P 3 Prime B for 17-18. Results for 17-18 sequencing were inconclusive and a sample using 14167 5 Prime 1 was sent for sequencing.) Sequencing results were received from Cornell University's Life Science Core Laboratories Center (Fig. 7). The sequenced fragments were compared to the wild-type sequence using the BLAST2 program. This is a tool that produces the alignment of two given sequences using BLAST engine for local alignment. It shows where the homology exists between two sequences. The P-element excision in the 1-2 stock removed 769 bp from the region. With the primers used, this would amplify a fragment of 365 bp. However, there remains sequence from the P-element itself, 39 bp to be precise, giving rise to an amplification of 404 bp.

Figure 6. Determination of PCR product concentration after purification. Using 1 µl of hyperladder as a comparison, 0.5 µl of product was loaded and ran on a 1.2% agarose gel. The amplified products show intensity similar to 20 ng/µl band in hyperladder. This is a sufficient amount of product for DNA sequencing. Lane A – 1-2, lane B – 14-1, lane C – 17-18.

The P-element excision in the 14-1 stock has removed 839 bp, this would produce a fragment of 295 bp. However there are 30 bp that remain in the region from the P-element producing a fragment of 325 bp. The 17-18 deletion sequencing results contradicted previous PCR results. The sequencing shows a 1201 bp deletion that extends into the coding region of *e(r)*. It was previously shown that this region was intact, and yet these results show the deletion extending to +302.

A.)

B.)

C.)

Figure 7. Chromatograms of Sequenced Deletion Stocks
A.) Chromatogram of 1-2 deletion. All results were obtained from DNA Sequencing Facility at Cornell University. Mut1 3 Prime primer was used for sequencing reaction. Deletion is 769 bp in length with a 39 bp relic of the P-element. The breakpoints of the deletion are -887 and -118.
B.) Chromatogram of 14-1 deletion. Mut1 3 Prime primer was used for sequencing reaction. Deletion is 839 bp in length, with a 30 bp relic of the P-element. Breakpoints of deletion are -887 and -48.
C.) Chromatogram of 17-18 deletion. 14167 5 Prime 1 primer was used for sequencing reaction. Deletion is 1201 bp in length and does extend into the coding region. The breakpoints of the deletion are −899 and +302.

Meanwhile the other three deletion stocks 1-9, 16-15, and 22-1 still needed to be localized and eventually amplified for sequencing as well. For the stocks 1-9 and 22-1, the primers 14167 5 Prime 2 and ER1 3 Prime 2 (expected fragment size of 933 bp) were used in an attempt to encompass these deletions. This 3' primer shares sequences with ER-P 5 Prime A,

which is known to be intact in these stocks. For the stock 16-15, the 3' primer used was ER

upstream 3 Prime (expected fragment size of 1,897 bp), which lies just downstream of the

EcoRV site. No amplification was achieved in these reactions except for the controls. The next

reactions set up were to again, further refine these deletions and localize their 3' breakpoint as

much as possible. The primers used were Mut 1 5 Prime and Mut 1 3 Prime, which amplify a

fragment of 701 bp. This fragment was successfully amplified for the 22-1 deletion stock and the

control, but not for 1-9 or 16-15 (Fig. 8). This showed that the 22-1 deletion's 3' end does not

extend past the Mut 1 5 Prime priming site. New 5' primers were then obtained located upstream

of *e(r)* to assist in the localization of these deletions. PCR reactions were prepared for 1-9 using

9-1 5 Prime A and Mut 1 3 Prime (607 bp) and also 9–1 5 Prime B and Mut 1 3 Prime (534 bp).

Reactions were prepared for 16-15 using primers 16-15 5 Prime A and ER-P 3 Prime B (660 bp)

and also 16-15 5 Prime B and ER-P 3 Prime B (579 bp). For 22-1 a single reaction was prepared

using 22-1 5 Prime and Mut 1 3 Prime (847 bp). For each of these reactions, the expected

fragment sizes were amplified (Fig. 9).
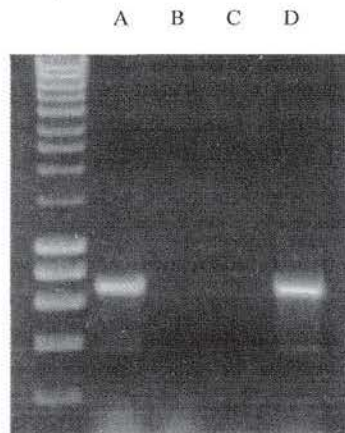
Figure 8. Assessing the 3' breakpoints in the stocks 1-9, 16-15 and 22-1. Reaction samples of 20 μl were loaded in 0.9% agarose gels to assess the 3' breakpoint of the remaining *e(r)* deletion stocks. Successful amplification confirms that deletion does not extend passed the Mut 1 5 Prime priming site. Lane A – positive control, lane B – 1-9, lane C – 16-15, lane D – 22-1.

From this data, PCR reactions were set up in an attempt to amplify a fragment encompassing the deletions. For 1-9 and 16-15, the primers 14167 5 Prime 2 and Mut 1 3 Prime were used to amplify the deletions and compare to the wild-type band of 1359 bp. For 22-1, the primers 14167 5 Prime 2 and 14167 3 Prime A were used, which amplify a fragment of 774 bp. No amplification was achieved in any of these reactions. These reactions were repeated several times, with the same result. More primer combinations were used multiple times, attempting to amplify these deletions with no success. At this juncture, it was thought best to halt any more
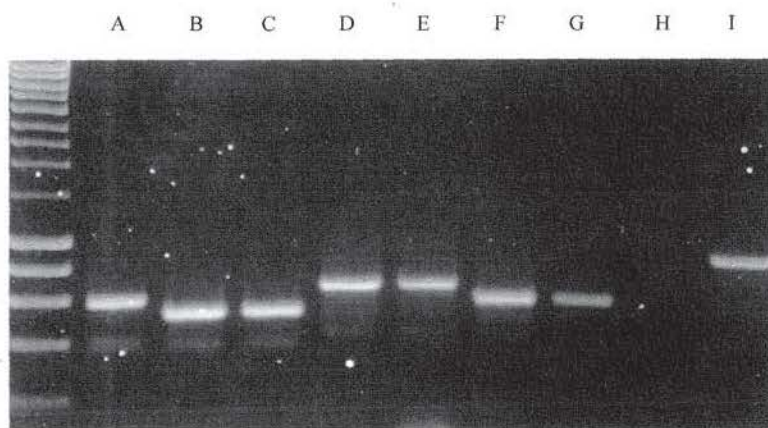
Figure 9. PCR to further localize 5' and 3' breakpoints for deletions for 1-9, 16-15 and 22-1. Reaction samples of 20 µl were loaded in 0.9% agarose gels to assess the 5' and 3' breakpoint of the remaining *e(r)* deletion stocks. Lane A – 1-9 (A primer), lane B – 1-9 control, lane C – 1-9 (B primer), lane D – 16-15 control (A primer), lane E – 16-15 (A primer), lane F – 16-15 control (B primer), lane G – 16-15 (B primer), lane H – 22-1 control, lane I – 22-1.

attempts at characterizing these deletions. However, there is some information that was achieved through these successive PCR reactions. There are known regions for the breakpoints of these deletions and information on where they do and do not extend (See Appendix B for deficiencies and breakpoints).

Conservation of Regulatory Sequences and Transcription Factor Binding Sites

Regulatory regions, although non-coding, are essential to gene expression and are conserved among closely related species. Sequences that are conserved in multiple species are more likely to be functional than non-functional and therefore tend to evolve at a slower rate than non-functional sequences[12]. With the recent sequencing of the 12 Drosophila species, genomic comparisons can be used to identify any possible conservation in both coding and non-coding sequence. EvoPrinter is a simple multigenomic comparative tool that allows one to rapidly identify multi-species conserved DNA sequences (MCSs) in the context of a single species of interest[13] (http://evoprinter.ninds.nih.gov/index11.html). After mapping sites in the regulatory region of a putative enhancer and silencer, these sequences were entered into EvoPrinter and compared among several Drosophila species. Through the deletion analysis, the region in which the putative silencer exists was refined to −118 to - 338. This sequence was entered into EvoPrinter with the initial comparison among the most closely related species: *D. simulans, D. sechelia, D. yakuba, and D. erecta*. The EvoPrint showed conservation at several locations and was run again, adding in *D. ananassae*, and then *D. pseudoobscura*. After adding *persimillis*, their were no longer any regions that were conserved among all of these species. The only remaining sequence that was conserved (see sequence in capital letters in Figure 10) was then entered into TESS. TESS, Transcription Element Search System is a web tool for predicting transcription factor binding sites in DNA sequences (http://www.cbil.upenn.edu/cgi-bin/tess/tess). After entering the conserved sequence, TESS shows known

D.melanogaster *e(r)* Genomic EvoPrint

aagctttagttAaTTTTATTTAAAAAAtctgaagactaaatatggaatgaattacgccAAATgAAATgAAAA

TatgCATTgaaaaacatgctttatacctaaaggcgtaatatacataaagtgtgcttttcaataaaaaagtgaatgaac

ataatatggaatcaaatacatatacAAAttGAaATGcAAGtatacattgatacaattaagtataaaaaat

Figure 10. EvoPrinter results from entering *e(r)* sequence between −118 - -338, the location of the putative silencer.
Capital letters represent bases in the D.melanogaster *e(r)* reference sequence that are conserved in *D.sechellia*, *D.simulans*, *D.yakuba* and *D.erecta* orthologous DNAs. Underlined sequence is predicted transcription factor binding site from TESS analysis.

transcription factor binding sites present, and the organism in which they exist. The only sequence that was entered, containing a known transcription factor binding site in Drosophila is TATTTAA, which is the binding site for TBP, TATA binding protein. TBP s a transcription factor that binds specifically to a DNA sequence called the TATA box that is typically found 25-30 base pairs upstream of the transcription start site in some eukaryotic gene promoters[14]. TBP, along with other associated factors, make up the TFIID complex, a general transcription factor that in turn makes up part of the RNA polymerase II pre-initiation complex[15]. The next sequence entered into EvoPrinter was the region where putative enhancer is located. Through the transgenic rescue experiments, this region was mapped to the region of −338 - -787. This sequence was compared to *sechelia, simulans, yakuba,* and *erecta*, which yielded several region of homology. After adding *ananassae* to the comparison, no similarities existed. The conserved sequence (see sequence in capital letters in figure 11) obtained from the EvoPrint was then entered into TESS.

D.melanogaster *e(r)* Genomic EvoPrint

CTGGTGgg<u>CACTC</u>gGATCttcGATTgCCGCTGTCTGGCAgCaCTGGGaacagtgagaacAG

TGGCcAGaCTGCGGAGCTgGCGATGgG<u>CGATA</u>GTTTTgtttcattaaaagtaacttttgactttatat

taaatacaatactaaaaaatgtaaaaatatgctttttaatatgcactttaagtgaaaatgattgtacattaagtttacttcataaagac

tttaatgaccataatttgttgatgggcgttaattttattctaacttaaataaacgtaatacctaatatttatatgtaataacaaaataaa

atgaacctaaataaataatgcaatccactttaatttgacattataattataatttttttataaatatataaaataaaaataataaataa

ataatttccgtagttattgtaatgtattactatgatgagtatatgttaaaag

Figure 11. EvoPrinter results from entering *e(r)* sequence between –338 - -787, the location of the putative enhancer.
Capital letters represent bases in the *D.melanogaster e(r)* reference sequence that are conserved in the *D.sechellia*, *D.simulans*, *D.erecta* and *D.yakuba* orthologous DNAs. Underlined sequences are predicted transcription factor binding sites from TESS analysis.

After entering the conserved sequence into TESS, two known transcription factor binding sites in Drosophila were located; they were sites for Zeste (CACTC), and BEAF-32 (CGATA). Zeste is a DNA-binding protein that has been shown to be a Trithorax group activator of the homeotic gene Ultrabithorax[16]. BEAF-32, Boundary element-associated factor of 32kD, is a transcription factor functioning in chromatin insulator sequence binding and protein binding. It is biologically involved in oogenesis, the establishment or maintenance of chromatin architecture, and regulation of transcription from RNA polymerase II promoter[17].

DISCUSSION

Defining the limits of *e(r)* with transgenes

The limits of the *e(r)* gene were previously mapped to a 6.1 kb fragment spanning from a *SalI* restriction site at -3836 to a downstream *SalI* restriction site at +2276. This information gave way to allow a further redefining of the genes limits. Using a set of transgenes of different lengths and sequences in an attempt to rescue different mutant lethal combinations, here, the limits have been further refined. In an attempt to rescue the lethality of the *Df(1)e(r)37-6* deficiency and *N^{nd-p}* combination in both males and females, the transgenes *Tr[w^+ e(r)^+]^{SA}*, *Tr[w^+ e(r)^+]^{cDNA}*, *Tr[w^+ e(r)^+]^{ES}*, *Tr[w^+ e(r)^+]^{PS}*, all showed the ability to rescue this lethality. The transgene *Tr[w^+ e(r)^+]^{HS}*, was the only construct not able to rescue, and this was shown using several different transgenic stocks. The same holds true for the attempts made to rescue lethality of the *Df(1)e(r)27-1* deficiency and *ec dx* mutation; *Tr[w^+ e(r)^+]^{HS}*, was the only construct not able to rescue. Also, for the attempts to restore viability and fertility in *Df(1)e(r)27-1* stocks all transgenic constructs showed the ability to rescue except for *Tr[w^+ e(r)^+]^{HS}*. The ability of the *Tr[w^+ e(r)^+]^{cDNA}* transgene to rescue shows that the two introns of *e(r)* do not contain transcriptional control sequences. The ability of *Tr[w^+ e(r)^+]^{SA}* to rescue shows that the female-specific polyadenylation sequence, and the rest of the sequence downstream of the *AflII* site are not necessary for *e(r)* function. As stated earlier, there are two *e(r)* transcripts, one 1.0 kb in length and the other 1.2 kb. The larger of the two is female specific and differs in its polyadenylation site. This transgene removes the female specific site while retaining the sequence for the shorter transcript. Although the female specific poly-A site has been removed, there does not seem to be an effect on the females, restoring viability and fertility. The larger

transcript is specific to the nurse cells, which contribute to oocyte development. One might predict that since the female specific polyadenylation site is absent that fertility would not be restored; however as shown here, it is. This shows that the shorter transcript is sufficient for the role of *e(r)* in oogenesis. The $Tr[w^+ \; e(r)^+]^{PS}$ transgene also showed the ability to rescue both lethal combinations and restore viability and fertility. The same can be said of the transgene containing the shortest sequence, $Tr[w^+ \; e(r)^+]^{ES}$. The $Tr[w^+ \; e(r)^+]^{HS}$ transgene, retaining more upstream sequence than $Tr[w^+ \; e(r)^+]^{ES}$, was shown not to rescue any of the lethal combinations nor restore viability or fertility. Because the $Tr[w^+ \; e(r)^+]^{HS}$ transgene does not show to be sufficient for normal *e(r)* activity, there must be something present in the sequence exclusive to this transgene that halts expression. However, the $Tr[w^+ \; e(r)^+]^{PS}$ also contains the same sequence as $Tr[w^+ \; e(r)^+]^{HS}$, along with additional sequence. This leads to the conclusion that there must be a silencer region located in between the *HindIII* and *EcoRV* restriction sites that hinders expression. If this silencer is also present on the transgene $Tr[w^+ \; e(r)^+]^{PS}$, then a dominant enhancer must exist between the *PvuII* and *HindIII* restriction sites. Through the use of these transgenic constructs, it has been shown that the sequences necessary for normal *e(r)* expression are mapped to a 2,618 bp fragment, which is comprised of the transcribed region of the gene along with 787 bp upstream from the transcriptional start site. Furthermore, a transcriptional enhancer has been mapped to a 5' region from −338 to −787 and a transcriptional silencer to a 5' region from −45 to −338.

Defining the control region of *e(r)* with deficiencies

Through the deficiency screening, six different stocks were found to contain deletions in the control region of *e(r)*. Out of these six, three were successfully sequenced and mapped. The

1-2 deletion removed 769 bp from the control region with 39 bp of the excised P-element

remaining. The deletion removes bases −887 to −118, surpassing the *HindIII* restriction site.
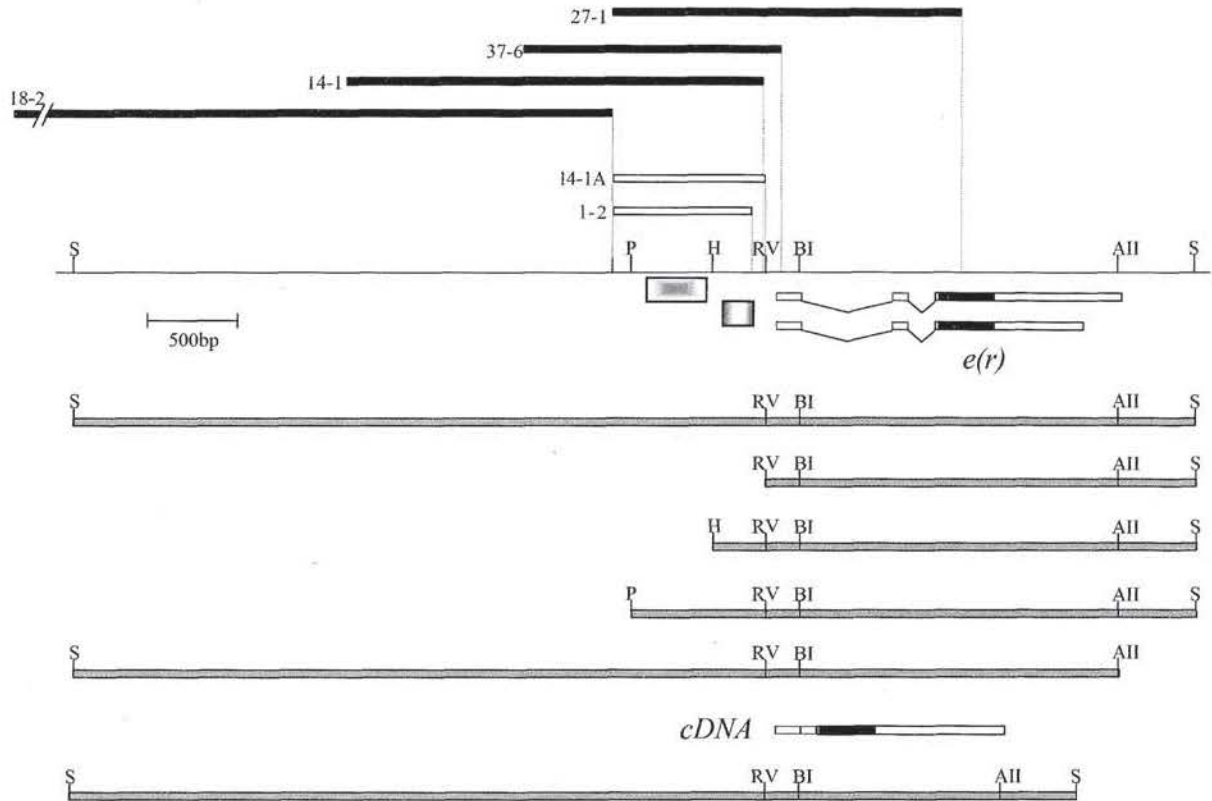
Figure 12. Genomic region containing *e(r)*, deficiencies generated and transgenes. The region containing *e(r)*, also highlighting the areas, which the putative enhancer and silencer are located, denoted by shaded boxes. The area between *Pvu*II and *Hin*DIII is the location of the enhancer. The area between *Hin*DIII and *Eco*RV is the location of the silencer. Also indicated, are the newly mapped deficiencies generated by imprecise P-element excision: 1-2, and 14-1A. Abbreviations: AII = *Afl*II, BI = *Bst*I, H = *Hin*DIII, P = *Pvu*II, RV = *Eco*RV, S = *Sal*I.

The 1-2 deletion stock shows good viability and fertility, which indicates probable normal *e(r)*

activity. If this stock does indeed have normal *e(r)* activity, then the location of the putative

silencer previously mapped from −45 to −338, can now be further refined and mapped to −118 to

−338. The 14-1 deletion removed 839 bp from the control region with 30 bp of the excised P-element remaining. Previous *e(r)* deficiency screens have already yielded a deletion noted as 14-1, so this new deficiency will be noted 14-1A. The deletion removes bases −887 to −48, which culminates adjacent to the *EcoRV* restriction site. This deletion in essence is the $Tr[w^+ e(r)^+]^{ES}$ transgene, which consists of *e(r)* sequence from the *EcoRV* site to the *SalI* site. This stock also shows good viability and fertility and thus, has probable normal *e(r)* activity. Although this deletion does not assist in further refining the mapped enhancer and silencer, it does support the data shown of the rescue and activity of the $Tr[w^+ e(r)^+]^{ES}$ transgene. The sequence data for the 17-18 stock, shows a deletion of 1201 bp in length and does extend into the coding region. The breakpoints of the deletion are −899 and +302. This contradicts the data obtained in the initial assessment of the coding region, which amplification showed that it was intact. Preliminary data shows that 17-18 flies have low viability and low fertility, which is consistent with low *e(r)* activity and the sequencing data. The three remaining stocks: 1-9. 16-15 and 22-1 were unable to be amplified successfully using primers encompassing the deletions. Breakpoints for these deletions have been identified to fullest capabilities via PCR. One possible reason for the inability to amplify these deletions is the occasional inversion caused by P-element excision. If the region that borders the site of P-element excision has been inverted then primers designed to bind to this site will not amplify because of a change in orientation of the sequence. Although this is a possibility and has been shown to occur, the probability of this to have happened in all three of these stocks is fairly low. From these three stocks, 1-9 and 22-1 show good viability and fertility and therefore probably have normal *e(r)* activity. The 16-15 stock shows low viability and fertility, which is consistent with low *e(r)* activity. Without the sequence data, these stocks will not assist in further mapping the control sequences for *e(r)*.

Sequence Conservation and Predicted Transcription Factor Binding Sites

With the use of EvoPrinter and TESS, conserved sequences among the 12 Drosophila species were picked out and analyzed for any sequences that are known transcription factor binding sites. For the region of the putative silencer, TBP or TATA Binding Protein was the single TF binding site present that exists in Drosophila. TBP binds to the TATA box in eukaryotic gene promoters and is part of the TFIID a complex that is involved in the initiation of transcription. This is not the type of transcription factor that one would assume to be involved in repression of gene expression, especially since it is highly involved in the initiation of gene expression. However, the binding site for TBP is what came from the results of the EvoPrinter and TESS work, not the actual transcription factor itself. If these results are correct and TBP does bind here and represses *e(r)* expression, it could possibly be a modified form or part of a complex that TBP is a part of but functions in blocking expression. On the other hand, it may be possible that another protein may have the ability to bind to this site and block expression or the binding of other enhancers that initiate *e(r)* expression. Because of the importance of TBP in expression of all genes as part of the RNA Polymerase pre-initiation complex, taking TBP out of the system itself would most likely affect expression of all genes, and be lethal to the organism. This makes testing the possibility of this being the actual silencer region more difficult than, and other approaches must be pondered. The putative transcription factors for the enhancer region, Zeste and BEAF-32 are good candidates and their interaction with *e(r)* can be tested very easily.

Future Work

The *e(r)* control region deletions found in these screens give some insight into the activity of *e(r)* in these stocks by looking at their viability and fertility. As stated earlier, the 1-2, 1-9, 14-1, and 22-1 deletions show good viability and fertility. This means that they probably have normal *e(r)* function and the deletions either leave the enhancer intact, or remove both the enhancer and silencer. The 16-15 and 17-18 stocks show low viability and fertility, meaning probable low *e(r)* activity. The 17-18 deletion actually enters the coding region, which is consistent with this data, and the 16-15 deletion probably removes the enhancer while leaving the silencer intact. These stocks have been combined with N^{nd-p}, and will be looked at to determine if these deletions are lethal in combination with this *notch* allele. A more extensive screen could also be done with stocks in which P-elements have been mobilized and deficiencies isolated which are strictly contained in the control region. Although an arduous process, information obtained from these deletions is highly valuable, and with enough of them can result in highly refined sequence data of the control regions for *e(r)*. Also, different transgenic constructs could be utilized in a similar manner as was done in this study, but using the information obtained here to help further refine the sequences that are necessary to control *e(r)* expression.

To further look at the proposed transcription factor binding sites obtained through TESS, these sites can be mutated through site-directed mutagenesis. By altering the sequence at these sites, their corresponding transcription factors will be incapable of binding. If these sites truly dictate the expression of *e(r),* then altering the putative enhancer should repress *e(r)* activity, and altering the putative silencer should increase *e(r)* activity. Also to test if the respective transcription factors are truly governing *e(r)* expression, mutants for these can be obtained to see

if they also affect *e(r)* activity. Again, TBP mutants are probably not the best approach, given

that they are probably not very fit, if they are viable at all. Zeste mutants however, are a possible

approach to see if they show any phenotypes similar to that of *e(r)* mutants. If, Zeste is the

transcription factor that binds to the enhancer of *e(r)*, than in its absence, expression should be

very low and viability and fertility should be affected. Zeste is suggested to be involved in the

expression of a wide variety of genes at different developmental stages. The *e(r)* and *Notch*

lethality utilized earlier, was shown to occur in the pupal stage (S.I. Tsubota, unpublished data),

while *Zeste* expression has also shown to be nearly ubiquitous in pupae[15]. Also, another *Notch*

allele N^5, has been shown to be lethal in combination with *Df(1)e(r)27-1*, and the initial data

show that these die as embryos (S.I. Tsubota, unpublished data). Both Zeste and ER are

maternally contributed and Zeste is also present in late stage embryos. These data are consistent

with the notion of Zeste possibly controlling *e(r)* expression during these stages. The other

possible transcription factor controlling *e(r)* expression is BEAF-32. Boundary Element

Associated Factor of 32 kD mutants could also be utilized to observe any possible *e(r)*

phenotypes. BEAF-32 has been shown to be involved in oogenesis, which *e(r)* is also involved.

Of the 12 alleles reported by FlyBase, two of the phenotypes are mentioned are associated with

the X chromosome, where *e(r)* is located, and also the nurse cells, which are polytenic germline

cells that contribute to the development of the oocyte. Levels of ER can be looked at in mutants

for these transcription factors. If they do enhance *e(r)* expression, then ER levels should be lower

in these mutants. Although the data for TBP does not have any clear association with acting as a

silencer, the other two predicted transcription factors; Zeste and BEAF-32 show good correlation

with *e(r)* function and potential in the possibility for controlling *e(r)* expression.

Appendix A. Scheme used to screen for deletions in control region of *e(r)*

1.) Isolate gDNA from stocks containing P-element insertion and mobilization

2.) Use PCR to amplify sequence entering coding region of *e(r)*

- If normal length of 549 bp, then keep for further assessment
- If shorter, then can eliminate from possible control region deletions

3.) Use PCR to amplify sequence downstream of P-element insertion site to transcriptional start site

- If amplification occurs, then region is intact, can eliminate from possible control region deletions
- If no amplification occurs, then will use further to define breakpoints and localize deletions

4.) Use PCR to localize deletions

- If successful amplification, then used pair of primers further downstream (closer to *e(r)* control region) to amplify region.
- If no amplification, deletion must extend into the region of 3' primer. This determines 5' breakpoint of deletion. Will repeat this procedure using primer pairs downstream of control region, and continue upstream until no further amplification is possible. This determines 3' breakpoint of deletion.

5.) After 5' and 3' breakpoint are determined, primers outside of these breakpoints (5' primer upstream of 5' breakpoint and 3' primer downstream of 3' breakpoint) are used to amplify the smallest fragment possible encompassing the deletion. These amplified fragments were then purified and sent for DNA sequencing.

6.) After obtaining sequence of deletion, used BLAST2 program to align sequence of deletion and wild-type *e(r)*. Two regions of the sequences align, the region of wild-type sequence that has no match in deletion stock is the sequence removed through P-element excision.

Appendix B. Sequences Deleted by P-element Excision and Breakpoints

Red = deleted sequence

1-2 Deletion

```
acaaaaagcacgtaaataatgttga**gacgagat**tttgcttttttgttgttgatttctttgccgat
cgctctgccgccgctcgactgcgctgcccgcgacaccgcgctcttccaattggtccaattgaga
tcaaacagctggtgggcactcggatcttcgattgccgctgtctggcagcactgggaacagtgag
aacagtggccagactgcggagctggcgatgggcgatagttttgtttcattaaaagtaacttttg
actttatattaaatacaatactaaaaaatgtaaaaatatgctttttaatatgcactttaagtga
aaatgattgtacattaagtttacttcataaagactttaatgaccataatttgttgatgggcgtt
aatttattctaacttaaataaacgtaatacctaatatttatatgtaataacaaaataaaatga
acctaaataaataatgcaatccactttaatttgacattataattataatttttttataaatata
taaaataaaaataataaataaataatttccgtagttattgtaatgtattactatgatgagtata
tgttaaaagctttagttaattttatttaaaaaatctgaagactaaatatggaatgaattacgcc
aaatgaaatgaaaatatgcattgaaaaacatgctttatacctaaaggcgtaatatacataaagt
gtgcttttcaataaaaaagtgaatgaacataatatggaatcaaatacatatacaaattgaaatg
caagtatacattgatacaattaagtataaaaaatcacaattgtcatttaatttggtaaatatta
ttaaacaatgtatatcaaataaaagatgttagtattaatcgatatctttcgattattttgtcac
ggcgaggcacgcagcgaagcgctcatctctagcttcagcctcctcttcaaaacccatagaactg
gcaaaataaaaaaaaacgaatcaagcgctgattttaccgaatctctttactgaatcagcatatt
atcatagtccatatagtccggcgttcgaacgggtaagttgaacactaaacgggagacatgggaa
ttatttaaaattcgagaaaaatcggtgaaaaggaagaaaaacaagtgattgcatacagttagtt
acacacacgcggcggccgttactccttgtacatgtgtgtgcgtgtctgttcgcgcttccgaatg
tttgtgttttagtgaagcggcataagcgtgcattatcttgtacaaatgtacgcatttctccatt
gccaacattcgcttcacacacaaattact
```

14-1 Deletion

```
acaaaaagcacgtaaataatgttga**gacgagat**tttgcttttttgttgttgatttctttgccgat
cgctctgccgccgctcgactgcgctgcccgcgacaccgcgctcttccaattggtccaattgaga
tcaaacagctggtgggcactcggatcttcgattgccgctgtctggcagcactgggaacagtgag
aacagtggccagactgcggagctggcgatgggcgatagttttgtttcattaaaagtaacttttg
actttatattaaatacaatactaaaaaatgtaaaaatatgctttttaatatgcactttaagtga
aaatgattgtacattaagtttacttcataaagactttaatgaccataatttgttgatgggcgtt
aatttattctaacttaaataaacgtaatacctaatatttatatgtaataacaaaataaaatga
acctaaataaataatgcaatccactttaatttgacattataattataatttttttataaatata
taaaataaaaataataaataaataatttccgtagttattgtaatgtattactatgatgagtata
tgttaaaagctttagttaattttatttaaaaaatctgaagactaaatatggaatgaattacgcc
aaatgaaatgaaaatatgcattgaaaaacatgctttatacctaaaggcgtaatatacataaagt
gtgcttttcaataaaaaagtgaatgaacataatatggaatcaaatacatatacaaattgaaatg
caagtatacattgatacaattaagtataaaaaatcacaattgtcatttaatttggtaaatatta
ttaaacaatgtatatcaaataaaagatgttagtattaatcgatatctttcgattattttgtcac
ggcgaggcacgcagcgaagcgctcatctctagcttcagcctcctcttcaaaacccatagaactg
gcaaaataaaaaaaaacgaatcaagcgctgattttaccgaatctctttactgaatcagcatatt
atcatagtccatatagtccggcgttcgaacgggtaagttgaacactaaacgggagacatgggaa
```

```
ttatttaaaattcgagaaaaatcggtgaaaaggaagaaaaacaagtgattgcatacagttagtt
acacacacgcggcggccgttactccttgtacatgtgtgtgcgtgtctgttcgcgcttccgaatg
tttgtgttttagtgaagcggcataagcgtgcattatcttgtacaaatgtacgcatttctccatt
gccaacattcgcttcacacacaaattact
```

17-18 Deletion

```
acaaaaagcacgtaaataatgttga**gacgagat**tttgcttttttgttgttgatttctttgccgat
cgctctgccgccgctcgactgcgctgcccgcgacaccgcgctcttccaattggtccaattgaga
tcaaacagctggtgggcactcggatcttcgattgccgctgtctggcagcactgggaacagtgag
aacagtggccagactgcggagctggcgatgggcgatagttttgtttcattaaaagtaacttttg
actttatattaaatacaatactaaaaaatgtaaaaatatgctttttaatatgcactttaagtga
aaatgattgtacattaagtttacttcataaagactttaatgaccataatttgttgatggcgtt
aattttattctaacttaaataaacgtaatacctaatatttatatgtaataacaaaataaaatga
acctaaataaataatgcaatccactttaatttgacattataattataatttttttataaatata
taaaataaaaataataaataaataatttccgtagttattgtaatgtattactatgatgagtata
tgttaaaagctttagttaattttatttaaaaaatctgaagactaaatatggaatgaattacgcc
aaatgaaatgaaaatatgcattgaaaaacatgctttatacctaaaggcgtaatatacataaagt
gtgcttttcaataaaaaagtgaatgaacataatatggaatcaaatacatatacaaattgaaatg
caagtatacattgatacaattaagtataaaaaatcacaattgtcatttaatttggtaaatatta
ttaaacaatgtatatcaaataaaagatgttagtattaatcgatatctttcgattattttgtcac
ggcgaggcacgcagcgaagcgctcatctctagcttcagcctcctcttcaaaacccatagaactg
gcaaaataaaaaaaaacgaatcaagcgctgattttaccgaatctctttactgaatcagcatatt
atcatagtccatatagtccggcgttcgaacgggtaagttgaacactaaacgggagacatggaa
ttatttaaaattcgagaaaaatcggtgaaaaggaagaaaaacaagtgattgcatacagttagtt
acacacacgcggcggccgttactccttgtacatgtgtgtgcgtgtctgttcgcgcttccgaatg
tttgtgttttagtgaagcggcataagcgtgcattatcttgtacaaatgtacgcatttctccatt
gccaacattcgcttcacacacaaattact
```

Red = deleted sequence

Blue = deletion does not extend into this site

1-9 Deletion

```
aaaaagcacgtaaataatgttga**gacgagat**tttgctttttgttgttgatttctttgccgatcg
ctctgccgccgctcgactgcgctgcccgcgacaccgcgctcttccaattggtccaattgagatc
aaacagctggtgggcactcggatcttcgattgccgctgtctggcagcactgggaacagtgagaa
cagtggccagactgcggagctggcgatgggcgatagttttgtttcattaaaagtaacttttgac
tttatattaaatacaatactaaaaaatgtaaaaatatgctttttaatatgcactttaagtgaaa
atgattgtacattaagtttacttcataaagactttaatgaccataatttgttgatggcgttaa
ttttattctaacttaaataaacgtaatacctaatatttatatgtaataacaaaataaaatgaac
ctaaataaataatgcaatccactttaatttgacattataattataatttttttataaatatata
aaataaaaataataaataaataatttccgtagttattgtaatgtattactatgatgagtatatg
ttaaaagctttagttaattttatttaaaaaatctgaagactaaatatggaatgaattacgccaa
atgaaatgaaaatatgcattgaaaaacatgctttatacctaaaggcgtaatatacataaagtgt
gcttttcaataaaaaagtgaatgaacataatatggaatcaaatacatatacaaattgaaatgca
```

```
agtatacattgatacaattaagtataaaaaatcacaattgtcatttaatttggtaaatattatt
aaacaatgtatatcaaataaaagatgttagtattaatcgatatctttcgattattttgtcacgg
cgaggcacgcagcgaagcgctcatct
```

Known 5' Breakpoint: - 556
Known 3' Breakpoint: - 458

22-1 Deletion

```
ctgctgcggctgcgccttcacttttcactaaaacaaaaagcacgtaaataatgttgagacgaga
ttttgctttttgttgttgatttctttgccgatcgctctgccgccgctcgactgcgctgcccgcg
acaccgcgctcttccaattggtccaattgagatcaaacagctggtgggcactcggatcttcgat
tgccgctgtctggcagcactgggaacagtgagaacagtggccagactgcggagctggcgatggg
cgatagttttgtttcattaaaagtaacttttgactttatattaaatacaatactaaaaaatgta
aaaatatgctttttaatatgcactttaagtgaaaatgattgtacattaagtttacttcataaag
actttaatgaccataatttgttgatgggcgttaatttattctaacttaaataaacgtaatacc
taatatttatatgtaataacaaaataaaatgaacctaaataaataatgcaatccactttaattt
gacattataattataatttttttataaatatataaaataaaaataataaataaataatttccgt
agttattgtaatgtattactatgatgagtatatgttaaaagctttagttaatttttatttaaaaa
atctgaagactaaatatggaatgaattacgccaaatgaaatgaaaatatgcattgaaaaacatg
ctttatacctaaaggcgtaatatacataaagtgtgcttttcaataaaaaagtgaatgaacataa
tatggaatcaaatacatatacaaattgaaatgcaagtatacattgatacaattaagtataaaaa
atcacaattgtcatttaatttggtaaatattattaaacaatgtatatcaaataaaagatgttag
tattaatcgatatctttcgattattttgtcacggcgaggcacgcagcgaagcgctcatct
```

Known 5' breakpoint: -734
Known 3' breakpoint: -698

16-15 Deletion

```
ctgctgcggctgcgccttcacttttcactaaaacaaaaagcacgtaaataatgttgagacgaga
ttttgctttttgttgttgatttctttgccgatcgctctgccgccgctcgactgcgctgcccgcg
acaccgcgctcttccaattggtccaattgagatcaaacagctggtgggcactcggatcttcgat
tgccgctgtctggcagcactgggaacagtgagaacagtggccagactgcggagctggcgatggg
cgatagttttgtttcattaaaagtaacttttgactttatattaaatacaatactaaaaaatgta
aaaatatgctttttaatatgcactttaagtgaaaatgattgtacattaagtttacttcataaag
actttaatgaccataatttgttgatgggcgttaatttattctaacttaaataaacgtaatacc
taatatttatatgtaataacaaaataaaatgaacctaaataaataatgcaatccactttaattt
gacattataattataatttttttataaatatataaaataaaaataataaataaataatttccgt
agttattgtaatgtattactatgatgagtatatgttaaaagctttagttaatttttatttaaaaa
atctgaagactaaatatggaatgaattacgccaaatgaaatgaaaatatgcattgaaaaacatg
ctttatacctaaaggcgtaatatacataaagtgtgcttttcaataaaaaagtgaatgaacataa
tatggaatcaaatacatatacaaattgaaatgcaagtatacattgatacaattaagtataaaaa
```

```
atcacaattgtcatttaatttggtaaatattattaaacaatgtatatcaaataaaagatgttag
tattaatcgatatctttcgattattttgtcacggcgaggcacgcagcgaagcgctcatct
```

Known 5' breakpoint: −294
Known 3' breakpoint: −197

Appendix C. Region showing *e(r)* and the two flanking coding regions - 5' to 3' for *e(r)*

```
ataaaacaaacaaacaaaaaccagaaacgtgtgtttgctgatgaattgcttttcgcggag
catgtggacgacggtcattgccaagggcgtgatacttttgccggcaactgtgacgaaaac
agcggcaattaaaaaataaaaggaaaaatacgaaatatttatgtattgtacatacgtata
taaaaaaaaaaccaaaaaaaaggtgtcaactcagacaaccttcaatggcacacaataaaa
atcggcgaccaaagaaaccagaaacaattggccaaaaaaaaaaccaagcggaagaaacg
agaaaatcaaattcaatctcaatggcaacgaatttgtgatgctcttttaaaatgaatggta
taaatacatgtgttctttaacgtatattattaccttattaataactagaatattaactct
atttatctgtcagttagcaatttgtagttcgcagaaaaaaaactctcaaaaaccattgc
taaatgctctaaacctttaaaaaaaaacccaacattttcgcactaaaaaaaactttaaaa
ctttaatgtacactgattcctatcagttgaatgcccgtttcgttataaagtttttttttt
atgttgggccaaaaggcaacgaccaa<b><u>gtcgac</u></b>gagttttttttcgaggcggcaaattcgt
aatgcagaataaataacgcataaatacaaatccacgataacagccaaatgcggtagcacg
acaaaataaaacaataatcatagcccagtagaaatgcgaaacaattacgccaaatgtgga
gaaaccaaaaaacgagcaagaaaaaaaaacagtcgtaaaacgtagcagcaaaaaaaaac
aacagaactagaactctgtagactaacagatactccttgttctgttcaacctggctggtt
tgtaagaatacaaattttttggatttttttgtgtaacaaaacttgcaaaaagagcacaa
aatttaaacaaaattaccaaatttaaaaccaaatcgtttggcatttaaaaaacaagttta
agtttcatttcattatttcgagtgctcaaaatactgattatttttacacaaaaaatagg
gaaaacaattttttggccaaaattcatatttacgattggttttttctgtataacttggctt
aaaattggcagagagccaaaagaactaccgttttttactcctaattgacaatactaacta
tccatatccgttttttacggactttgtaaaattaattttttgaccaaaatttcgcattttt
tgtaaggggtaacatcataaaaaatgaaaaaattgaaaaaatttacatttccaacttttt
gaatacagttcgattaaaaatttaaa<b><i>tacaaactcagcgatgtatgg</i></b>cattccataatca
gacaaatattttttcaagttgtggccaaaaaaaatagatatttttatgacaaaaaaatagtaaa
aacgggttttgtcaaaaattgaaatattttttgattggcatttttaatcataacttggctaa
aaatgatcataaagctgaaagaataactgttttgggcagctaattaccaatgccaaacag
ccctatcactttttgacgtatttgaaaaaattaatttttttaagaaattttcgcattttttt
gtaagggggtaacatcatcaaaatttggaaaaaaattgaaaaaattaaaattccaatttttt
gaacacagtttaattggaaatttaattacgagctcaacgaggtatgacattccatattaa
aagcaattaaaccaatggaattgtacagatatacagatttaagaatgattgtttaa<b><i>aaca</i></b>
<b><i>tgagtgctgtttaactgtc</i></b>attataaatatgaaactatctggttgttaaaaaatttatat
ttctcatccg<b><i>ggtagtgggaaattagcgac</i></b>gcaccttaaaacaaataacaatacgcctac
aaagtcgtttataatagatacacagcttttgcacacttttgacgataaaaacgatgatgt
ttcaaacacatcaggggcgcccacgttcttatcagttgagaatgttctttacttatcggc
gccgtaaagcaaacacttttcaaaaacgagcgttaaacaagacaaattattacagatgct
tagatgcgataattgtaacctaattgtgattaattgtaactaaatgattaaaattaaat
aaaaatcacagtttgaatgcacaaaattgaaataaagtcgatttctccatttacaaaaac
ttgatatctccatttggaaattctaattggtttgctctcctgccttttctgcttccactc
tctctctcacttccaattggaattggggactaacgaaatggaatcaagttcaaggccacg
atcgattgattgccatctctgtc<b><i>gca</i></b><u>cttgttggatgtgactc</u>acct<b><u>GCTGATTGGAATG</u></b>
<b><u>AAGCTGCGCTGCTCCTCCTCTGCCGGCGTTGCCTCCTCATCGACGTTGGCAGAGGCAGCG</u></b>
<b><u>ACCTCTGACGTCACGGGGCGGTAAATCAAAGGCTTCACGTTGCGTTCGAAATCGTCGAGG</u></b>
<b><u>AAAACGACATCGTGGACGGGAATAATGCGGTGATCGGTGACGTTGAAGTGGAGATCGCGA</u></b>
<b><u>AAGGGGCCGTTGAAGAGGCAGCGCAACGGTGACGTCACCAGCACAGCATCTTCCGGAGGT</u></b>
<b><u>AA</u></b>ctgtaaaaaatgtacaatatttgttgtattttaacgtatatcaaacatttactgatat
```

Sal I

14167 5prime

14167 5prime C

14167 3prime

14167 5prime

14167 3prime

agtttttatattaaatcaaatacaccgcaacaaatctcgtgcaaacatcagataagatatt
gatacgatacgataattagtataatatttaaacaatgcgcagtcaggagaaatatagaat
ggcttatcaattttaaaataaaactaacggaaacagtaatattgtacatatttcagtttc
aactcaaagcataagttt*ctctgtagccaagacaccatt*tggtaagttaagaaaactatt | 14167 5prime
aagtggacgtttttttttttttac**CTTGATTTTGCCATTGGTCGCCGAGAACAT**tgagattt
cgagatttcagcgataagagatgagcaacaacacaaatcaaatagaatgcacaagaaaca
cagatacaacacccacacatgaatgatcgtgcaacggtgtgcgcagaacagtgaagagga
agcagagagacgaagcagaggcgacaaattggacacgcgtgcgtcgc*gttggttggatt* | 14167 5prime 3
*tgtgttt*tcttatttgactattgccttttcggtgctactttaacgggtttccccgtttca | 14167 5prime 2
cagcctttatca*tcttattttcgcagggcttt*tttgcttttcaatttactgcctgggttt | 14167 3prime A
catttaacttcgctgctcttctacttcattcattatttgatttgtttacgttgctatttc
ccaatttgtttgcttcttttcacttcattt*tctttcaattaggacgcaca*caaacacaca | 14167 5prime 1
ggaaaaaacaaaaaaaaaaaaagagagacacgaacgccacttgcacaattgaaaata*aca | 14167 5prime I
acaaattaatgggaacg*cgcgaaacacgggctgctgcggctgcgccttcacttttcacta
aaacaaaaagcacgtaaataatgttga**gacgagat**tttgctttttgttgttgatttcttt | 14167 P elemen
gccgatcgctctgccgccgctcgactgcgctgcccgcgacaccgcgctcttccaattggt
ccaattgagatcaaa*cagctg*gtgggcactcggatcttcgattgccgctgtctggcagca | *Pvu*II
ctgggaacagtgagaacagtggccagactgcggagctggcgatgggcgatagttttgttt
cattaaaagtaacttttgactttatattaaatacaatactaaaaaatgtaaaaatatgct
ttttaatatgcactttaagtgaaaatgattgtacattaagtttacttcataaagacttta
atgac*cataatttgttgatgggcgtt*aatttattctaacttaaataaacgtaataccta | Mut1 5 prime
atatttatatgtaataacaaaataaaatgaacctaaataaataat*gcaatccactttaat | 14167 3prime A
ttgaca*ttataattataattttttataaatatataaaataaaaataataaataaataat
ttccgtagttattgtaatgtattactatgatgagtatatgttaa*aagctt*tagttaattt | *Hin*DIII
tatttaaaaaatctgaagactaaatat*ggaatgaattacgccaaatga*aatgaaaatatg | ER-P 5prime A
cattgaaaaaca*tgctttatacctaaaggcgt*aatatacataaagtgtgcttttcaataa | 14167 3prime 3
aaaagtgaatgaacataatatggaatcaaatacatatacaaattgaaatgcaagtataca
ttgatacaattaagtataaaaaatcacaattgtcatttaatttggtaaatattattaaac
aatgtatatcaaataaaagatgttagtattaatc*gatatc*tttcgattattttgtcacgg | *Eco*RV
cgaggcacgcagcgaagcgctc**atctctagcttcagcctcctcttcaaaacccatagaac** | 1st *e(r)* exon.
tggcaaaataaaaaaaacgaatcaaagcgctgattttaccgaatctctttactgaatcag | ER-P 5prime B
catattatcatagtccatatagtccggcgt*tcgaacgggtaagttgaacac*taaacggga | Mut1 3 prime
gacatgggaattatttaaaattcgagaaaaatcggtgaaaggaagaaaaacaagtgatt
gcatacagttagttacacacacgcggcggccgttactccttgtacatgtgtgtgcgtgtc
tgttcgcgcttccgaatgtttgtgttttagtgaagcggcataagcgtgcattatcttgta
caaatgtacgcatttctccattgccaacattcgcttcacacacaaattactcgttttcca
tgctggaagtgccatcttagtctatttcccagttgcaaattgga*aaatcaaagagcgttc* | ER-P 3prime B
*tgctg*gccagcatgtgtggtgcgcgtgtattggtgttgcggcgagagcgcgagcgagatg
gggcgatgcgttgcatgcggcaacagcatatctcttccatctctttcgtttttgttcttt
tcctttttttttttgtcttttttttttttgctgacagctctctatgcttctaataaatatt
tatgcacaatttttgca**gagctggcgaagtaaacttcaacttcgagagtctcggaatctt** | 2nd *e(r)* exon
tcggactttgctatagaaaaaacgcttgttaaaaaaaaggatgtaagtgatgaagatgaa | 14167 3prime
acgtaatgtaatgaaacgcacactagtaacatacatacatacatatcgcttgagtacata
g*atcaccgccaataacattaat*gttccattttttgc*cttgcagatcaccATGTCGC*ACACC | ER 5 prime
ATCCTATTGGTACAGCCGGGTGCTCGTCCGGAGACCCGCACTTACTGTGACTACGAGAGC | 3rd *e(r)* exon.
GTCAACGAGTGCATGGAGGGCGTGTGCAAGATCTACGAGGAGCACCTGAAGCGCCGCAAT | *e(r)* coding
CCAAACACTCCGACCATTACGTATGACATTAGCCAGCTATTCGATTTCATCGACACAATG

<u>GTGGACATTAGCTGCATGGTTTACCAGAAGAGCACCAATACCTATGCGCCCTACAATAAG</u>
<u>GATTGGATCAAGGAGAAGATCTATGTGCTGCTCCGTCAGGCGGCATTTAGTTCCAATACC</u>
<u>TAA</u>agggacggcgatgcgacaagatacatacatataaaaccaaataacacgatgatgcat
gcatcaatgcgtgc**aacactaacgacaacaacacgttgatgtatgcaa**_caatgcgtgcaa_
tagcaact**acaacaacaacaacaatacgaggatacatgcaacagcgcttggaacagcatc**
cgcaacggccacaacaccataacaccaataacaccaacaagatgaggaagaagaagaaga
tgatgatggagtggtggggcatcgattccattggagacttgctgctaccgactagatcgc
caaatgacagcaacttttttttgcaacataacgcaatgcaaaaaaggaaaaggaacacaa

ER 3 prime

agagaacgataaactgtataacttagaaatggaaaacttactttaaacgtactaatatat
atatatatatataaatatatatataggcgtatatatatatatactgtatatacaccca
actgacaac***tgtgtgtgtttttgtgtgtt***tcaatgttttttttgtgcggcagccacaggtg
acggttgttcttatcgatagcgaactgtatgtttgtttgtttgtaaaaccgaatggcaaa
tttgcaagtgtgcattttttt***ttttgtatgttttgccgtgt***gtgctagcttaagtctttgc

ER downstrean 5D

aatctatcgaataataaactaacacatcaggacacactgagcatatataaacatatttgt

ER 3prime C

aaactactgattccccccccccccctttatcatttgaacttagttaaacaacaacttttа
gcccgctgattttgcacaccgatgggctttgtggcgttttaacggcttgttttaagcaaa
acagaacctttggtcgcagatgcgaaatgaccttgagatgctggttgttgtgtgtgtggt
gagggggggagggggtagggggtttgtgggttgtaccccgccagatattcaatcggataacg
aacttggaccgccgctgaatcgcttggccgcactgcattgaacttggcgtgcgatattta
attagctccgcgacagcattttgttaaatatatatatatatatatatatgtatata
tgcaca***tttagt**<u>***gtcgac***</u>***ttgtggat***gctgcttactagcgttaactagcaataaacatag

ER 3prime D

_Sal_I

tgtttatagtttacccattatatttgctgaatattaatgatcactagatttctcattatt
cagacaaatatatttgtttgtgaaaaaagcatttgctacagcaaactgtacagtttatta
caaaaatatgtatatatattttcaaaaaaaaaaaaaaaagaatttctcctatttaaagta
gtttgcttgccacgcaactccgcgcttttttacagggtatgcaatagcgctacagctgcgt
cgaatatggtgcatagttttttttttttagaatatttttcaaaactggttttaaacactgc
attgtcgccacaccgccccgcaccgcccactttcaatgtcagttagccagcctcgctttt
tggccagaaagcaaagcgaatgggatgcagatgaagtgcatcctgcatgagatgcatttt
tcggtcatatagtgttgcacttgtctagttttgttttttttctttcttt**tttatgtaaa**
atactattcttaaccggctaaacaataagcgttaacagctaagtggattgagcatggggg
gggggggagaaatatatattgtggacctggctaatggcgcatctgccgctgtgcctccttc
tccacctcatactcgtactccagctgcagtttgggcttcgagcggctgc_gtcgac_**tgttg**
cccgccggcgatttcttgtccttcttcgctttgctggcggttttcttggctggcactttg
gtctcctgttagaatgcaattgttcgttttaagttcggtttggcaattggaatttagttt
tgtttcgttttagttttctttttttttggcaattttctggagctgttagttagcgcgcact
gaatggaaaccgacgcaactgaatgaaaagcggcgaacggcgctcggaaagctttattct
ccaagctccaagcggggcagaactcgtttgctatttctaattatggtaatttcggaagaag
gtggcaggatttaagcaattgccaactatttacatactcagaattttcaatattaaaaga
aattcattgaaaaaggatttggaaatcgttaaatgctccagaaaactgccaacaaaataa
aaaagaaaagcctgatacatgaccagaaacactac**ctcaatatcatctgcatccgcatcg**
tcatcctcgtcggactcctcgaaatcgctgcccacctcgacctcttcccgcttgccggaa
tcgctgtcgtagtcgtcctcgtcatcatcatcatcatcatcatcctcctcctcatcg
tccgtatctgcctcgacgaactcctcgtcgagcagggatctctgtacttcctcagcttcg
cgtgaatcgtcgccgtccacctccatttcgcgctccagctgctcctcattctcgtcttcg
tcgtctagctcctcctcatcatcctcgtccacgcgttccgcctccagagccttgttgaaa
gcggtctgcgaaaagttgtagatgtcgcgataggtgccattcttcagacgatccatcagc
gccttctcgatgtgattgtcgatcttggcggccaccaaggccttctcctcgcgccgcg

References

1) Orphanides G., Reinberg D. A Unified Theory of Gene Expression. Cell 2002; 108: 439-451.

2) Bornstein P., McKay J., Liska DJ., Apone S., Devarayalu S. Interactions between the promoter and first intron are involved in transcriptional control of alpha 1(I) collagen gene expression. Molecular and Cellular Biology 1988; 8: 4851-4857.

3) Tsubota SI, Schedl P. Hybrid dysgenesis-induced revertants of insertions at the 5' end of the *rudimentary* gene in *Drosophila melanogaster*: transposon-induced control mutations. Genetics 1986; 114:165-182.

4) Adams MD. The Genome Sequence of Drosophila melanogaster. Science 2000, 287:2185

5) Lindsley DL, Zimm GG. The Genome of Drosophila melanogaster. 1992, Academic Press, Inc.

6) WAN, C., W. TEMPEL, Z. J. LIU, B. C. WANG and R. H. ROSE, 2005 Structure of the conserved transcriptional repressor enhancer of rudimentary homolog. Biochemistry 2005 **44**: 5017-5023.

7) Phillips A. Genetic and Biochemical Studies of *Enhancer of Rudimentary* to Elucidate Protein Function. Ph.D. Thesis, Saint Louis University; 2005.

8) Brody, TB., Meet the Interactive Fly; The Interactive Fly [edition 53], sponsored by Society for Developmental Biology (http://www.sdbonline.org/); August 24, 2008; Accessed January 27, 2009 @ http://www.sdbonline.org/fly/aimain/1aaintro.htm.

9) Greenspan, RJ., Fly Pushing. The Theory and Practice of Drosophila Genetics. New York. Cold Spring Harbor Laboratory Press. 1997.

10) Lodish H., Berk A., Zipursky LS., Matsudaira P., Baltimore D., Darnell J., Molecular Cell Biology. New York. W. H. Freeman. 2000.

11) Matsuno K, Diederich RJ, Go MJ, Blaumueller CM, Artavanis-Tsakonas S. Deltex acts as a positive regulator of Notch signaling through interactions with the Notch ankyrin repeats. Development. 1995 Aug;121(8):2633-44.

12) Maeder ML., Polansky BJ., Robson BE., Eastman DA. Phylogenetic Footprinting Analysis in the Upstream Regulatory Regions of the Drosophila *Enhancer of split* Genes. Genetics 2007; 177:1377-1394.

13) Odenwald et al., Introduction to EvoPrinter. 2005. PNAS 102: 14700-14705. Accessed May 8, 2009 @ http://evoprinter.ninds.nih.gov/aboutevop.htm

14) Kornberg RD., The molecular basis of eukaryotic transcription. PNAS 2007. 104 (32): 12955-12961

15) Lee TI, Young RA. Transcription of eukaryotic protein-coding genes. Annu Rev Genet. 2000; 34:77-137.

16) Hur MW., Laney JD., Jeon SH., Ali J., Biggin MD. Zeste maintains repression of Ubx transgenes: support for a new model of Polycomb repression. Development. 2002 Mar;129(6):1339-43

17) FlyBase. http://flybase.bio.indiana.edu/reports/FBgn0015602.html, Gene Report Dmel\BEAF-32. Accessed May 10, 2009.

18) Pirotta V., Bickel S., Mariani C. Developmental Expression of the Drosophila *zeste* gene and localization of zeste protein on polytene chromosomes. Genes and Development 1988; 2: 1839-1850.