Genetic and molecular characterization of two proximal heterochromatic genes in *Drosophila melanogaster* and their homologues in *Drosophila virilis*

by

Sandra Rimma Schulze

B.Sc., University of British Columbia, 1985

THESIS SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in the

Department of Molecular Biology and Biochemistry

SIMON FRASER UNIVERSITY

July 2003

Approval

Name: Sandra Rimma Schulze
Degree: Doctor of Philosophy

Title of Thesis: Genetic and molecular characterization of two proximal heterochromatic genes in *Drosophila melanogaster* and their homologues in *Drosophila virilis*

Examining Committee:

Chair: Dr. W.R. Richards, Professor

Dr. B.M. Honda, Senior Supervisor
Department of Molecular Biology and Biochemistry, S.F.U.

Dr. B. Brandhorst, Supervisor
Department of Molecular Biology and Biochemistry, S.F.U.

Dr. N. Harden, Supervisor
Department of Molecular Biology and Biochemistry, S.F.U.

Dr. D.A.R. Sinclair, Supervisor
Department of Molecular Biology and Biochemistry, S.F.U.

Dr. Lynne Quarmby, Internal Examiner
Department of Biological Sciences, S.F.U.

Dr. John Locke, External Examiner
Department of Biological Sciences, University of Alberta

Date Approved: July 30, 2003

PARTIAL COPYRIGHT LICENSE

I hereby grant to Simon Fraser University the right to lend my thesis, project or extended essay (the title of which is shown below) to users of the Simon Fraser University Library, and to make partial or single copies only for such users or in response to a request from the library of any other university, or other educational institution, on its own behalf or for one of its users. I further agree that permission for multiple copying of this work for scholarly purposes may be granted by me or the Dean of Graduate Studies. It is understood that copying or publication of this work for financial gain shall not be allowed without my written permission.

Title of Thesis/Project/Extended Essay:

Genetic and molecular characterization of two proximal heterochromatic genes in *Drosophila melanogaster* and their homologues in *Drosophila virilis*

Author: _____
(signature)

_____
(name)

6 August '03
(date)

Abstract

Heterochromatin comprises a considerable portion of most eukaryotic genomes. It is gene poor, possesses transcriptional silencing properties, and consists primarily of repetitive sequences. It is therefore exceedingly difficult to characterize using standard molecular methods, and has remained largely uncharacterized in those genomes sequenced to date. One third of the *Drosophila melanogaster* genome is heterochromatic, and since heterochromatin can silence gene expression, the presence of active genes in this region is paradoxical. By identifying and mapping these virtually inaccessible genes, our laboratory is contributing to the completion of the *Drosophila* genome project, and learning how chromatin structure affects gene expression.

I have focused on two genes located deep within the heterochromatin of the third chromosome's left arm. In the first part of this study, I outline the methods used to characterize these genes. They are separated by ~10kb of DNA, and present a study in contrast. *Dbp80* is a huge gene, spanning more than 140kb of genomic DNA, due to the expansion of repetitive sequences in its many introns, consistent with what is already known about heterochromatic genes. Although its homolog in other species plays an essential role in mRNA export, *Dbp80* is identified by no known lethal complementation group in *D. melanogaster*. *RpL15*, however, does correspond to a lethal complementation group (*lethal 2*), and encodes a large subunit ribosomal protein. This gene is uncharacteristically small, occupying less than 2kb of genomic DNA. Both genes appear to be positively regulated by Heterochromatin Protein 1, a chromosomal protein that normally silences gene expression, and thought to be important in maintaining heterochromatin structure. In the second part of this work, I describe the cloning of both genes from the related species *Drosophila virilis*. The contrasting nature of these genes is reflected by their evolutionary history. In *D.virilis, vRpL15* appears to have conserved both its gene organization and heterochromatic location, whereas *vDbp80* is a small euchromatic gene, approximately $1/90^{th}$ the size of its *D.melanogaster* homologue. It is flanked by a large and likely active retrotransposon, which may help to explain the kinds of intrachromosomal rearrangements that caused it to relocate into a heterochromatic environment.

Dedication

This thesis is dedicated to my parents Fritz and Ludmilla Schulze.

Acknowledgements

I should like to thank my supervisor Dr. Barry Honda for providing me with the opportunity to work in this fascinating field, and for preparing the beautiful polytene *in situs* which comprise an important part of the data in Chapter Four of this thesis. Barry also provided me with transgenic lines of the smaller of the two genomic constructs described in Chapter Three. I am also grateful to Drs. Nick Harden and Bruce Brandhorst, both members of my committee, who patiently sat through many meetings, providing encouragement, support and commiseration in equal measure, and without judgment or reservation. Members of the Honda lab, past and present, have been sources of ideas and indispensable aid, true friends and co-workers. In particular, had it not been for Kathleen Fitzpatrick and Elizabeth Silva, I truly doubt I ever would have stayed the course.

Dr. Don Sinclair, the fourth member of my thesis committee, deserves special mention. Apart from the sheer luck of having a brilliant geneticist working about three feet away from me on a daily basis, Don instilled in me that most critical combination of qualities required of a professional scientist: inspiration and tenacity. I don't know how I could have done any of the work in this thesis without him.

One of the most agreeable aspects of working in science is the necessary dependence on a world-wide borderless community. The following people generously gave up their time and resources in helping with this research: Dr. Horst Kress in Berlin for his expertise in reading the chromosomes of the related dipteran *Drosophila virilis;* Dr. Joel Eissenberg for his advice and *Su(var)* stocks, which comprise a full section in Chapter Three; Dr. Ling Hong of the Berkeley Drosophila Genome Project, who generously sent me an aliquot of his brand-new *D.virilis* cDNA library before it was released to the public, and Drs. Lori Wallrath and Pamela Geyer, who together provided a fresh and stimulating perspective after being presented with this research in close to its current form, in my first public seminar last year.

Finally I want to thank my parents, who suffered through all my academic and financial agonies, with equanimity and unfailing love. Words cannot describe my appreciation.

Table of Contents

List of Tables

List of Figures

List of Abbreviations & Definitions

Abbreviation /Name	Description
[w+]	Wild-type expression from a *white*$^{+}$ transgene
3L	Left arm of third chromosome in *D.melanogaster*. Homologous to Muller's element D.
Act-5C-GAL4	A GAL4 driver on the second chromosome under the control of an actin sensitive promoter (Flybase Stock # BL 4414).
ARM-GAL4	A GAL4 driver on the second chromosome under the control of a Armadillo sensitive promoter (Flybase Stock # BL 1560).
ArmU	Database of genomic sequences from Release 2 of the *D.melanogaster* sequencing project which cannot be mapped to any cytological position and therefore are called "unlocalized".
Chromodomain	"Chromosome organization modifier": a protein sequence motif (~41 amino acids) found in a variety of chromosomally associated proteins.
BDGP	Berkeley *Drosophila* Genome Project (http://www.fruitfly.org/).
BLAST	Basic Local Alignment Search Tools: a set of sequence comparison algorithms.
Dbp80	DEAD-box protein 80. DEAD stands for an amino acid motif (Asp-Glu-Ala-Asp).
EST	Expressed Sequence Tag.
Flybase	Repository of *Drosophila* gene descriptions and stocks in Bloomington Indiana (http://flybase.bio.indiana.edu:82/).
HP1	Heterochromatin Protein 1.
GADFLY	Genome Annotation Database: The Flybase (q.v.) annotation database.
Heteroallelic	See transheterozygote.
HSGAL4-CyO	A GAL4 driver on the second chromosome under the control of a heat-shock sensitive promoter (Flybase Stock # BL 2077).
In(3L)C90	A third chromosome inversion which breaks in proximal 3L heterochromatin (h51) and distal 3L euchromatin (62D).
Moonrat	A cis-dominant allele of *hedgehog*, which is suppressed in a *verthandi* (i.e., *lethal 3*) mutant background.
Muller's element D	Cytological definition of chromosome arm homology in *Drosophila*. D refers to chromosome 3 in *D.virilis* and 3L in *D.melanogaster*.
PcG	Polycomb Group of genes (described in Chapter One).
PEV	Position Effect Variegation (described in Chapter One).
RpL15	Large subunit ribosomal protein 15
SNAP25	A large gene that maps to the middle-distal portion of 3L heterochromatin
Su(var)2-5	Suppressor of variegation 2-5 (encodes HP1)
Su(var)3-9	Suppressor of variegation 3-9 (encodes a methyl transferase)
Trans-Heterozygote	A genotype heterozygous for two different alleles; whether from the same gene or different genes is made clear from context. Sometimes used synonymously with *heteroallelic* (q.v.) though this refers specifically to heterozygotes between different alleles from the same gene.
trx-G	Trithorax Group of genes (described in Chapter One).
verthandi	A *trx-G* gene which maps to proximal 3L heterochromatin. AKA *lethal 3*

CHAPTER ONE: General Introduction

Why study heterochromatic genes?

A genome is the sum of all the genetic information carried by a cell or an organism, but that information is packaged into a complex of nucleic acid and protein called chromatin, and chromatin is itself organized into different environments. In eukaryotes, chromatin is broadly divided into two contrasting compartments, which have been defined cytologically as euchromatic and heterochromatic (Heitz 1928). In interphase cells, euchromatin appears loosely compacted and diffusely staining, while heterochromatin is densely compacted and stains darkly. Heterochromatin retains this compaction throughout the cell cycle, in contrast to euchromatin, which condenses during the mitotic phase. In addition to these cytological distinctions, heterochromatin and euchromatin also differ in their sequence content. Euchromatin consists principally of single copy sequences and contains most of the active genes, whereas heterochromatin is comprised of both middle repetitive (Pimpinelli et al. 1995), and highly repetitive (Lohe et al. 1993) sequences, and is largely gene poor (Weiler and Wakimoto 1995). Its repetitive nature makes heterochromatin intractable to molecular analysis, and therefore there are few examples of completely characterized heterochromatic genes in the scientific literature. In addition, heterochromatin forms a transcriptionally repressive domain, which presents something of a paradox for the genes contained therein. Clearly they have evolved to function in an environment which normally silences gene expression. An in-depth study of these genes should therefore lead to a better understanding of the effect of chromatin structure has on gene expression in general.

Chromatin Structure and Gene Expression

The paradigm of gene expression was established in the 1960's with the elegant genetics of the lac operon in bacteria, in which defined regulatory sequences upstream of a coding region bound trans-acting factors that determined whether a gene was off or on. In eukaryotes however, gene expression occurs in the context of chromatin structure, and a new paradigm is in the process of being established, called the Histone Code Hypothesis (Jenuwein and Allis 2001). It is becoming increasingly clear that the fundamental unit of

chromatin – the nucleosome – plays a critical role in how genes are regulated. Nucleosomes are composed of histones, which have been shown to be subject to a wide range of post-translational modifications, including acetylation, methylation and phosphorylation (Richards and Elgin 2002). Trans-acting factors still play a critical role, but rather than recognizing a defined sequence, they are reading a pattern of histone modifications, which can result in a range of compacted states in the chromatin fibre, rendering the underlying genes more or less open to the transcriptional machinery. This is known as chromatin remodeling, and the trans-acting factors involved often work in multimeric complexes and in combinatorial fashion, permitting a huge range of variation in gene expression (Struhl 1999).

The contrasting cytological appearance of heterochromatin and euchromatin correlates with different patterns of chromatin remodeling, so that euchromatin is more open and accessible to transcriptional regulators, while heterochromatin is not, and in fact exhibits transcriptional silencing properties. This repressive tendency is best demonstrated by the phenomenon of position effect variegation (PEV), which was originally observed in the fruit fly *Drosophila melanogaster* (Muller 1930). PEV takes place when a euchromatic gene is relocated (usually by irradiating the chromosomes to induce breakages and translocations) near or within a block of heterochromatin. The newly translocated euchromatic gene will be selectively shut down in a proportion of the progeny cells as development proceeds, resulting in patches of adult tissue in which the gene is expressed and patches where it is not. This variation in expression is known as variegation, and the regulation is said to be epigenetic, since the actual sequence of the euchromatic gene has not changed, only its location. Screens have been carried out looking for modifiers of this effect (Sinclair et al. 1983, Wustmann et al. 1989), and when these modifiers were eventually cloned and characterized, many of them turned out to be components of the chromatin remodeling complexes described above (Schotta et al. 2003). For example the *Suppressor of variegation 2-5* gene (*Su(var)2-5)* encodes Heterochromatin Protein 1 (HP1) – a protein which binds modified histones as well as itself, causing the chromatin fibre to become even more compacted and transcriptionally repressed (Eissenberg and Elgin 2000). HP1 belongs to a class of chromatin components that appear to have a

general involvement in chromatin remodeling, and shares a specific protein motif (the chromo domain) with Polycomb, the first member of the *Polycomb Group* (*PcG*) of homeotic gene regulators to be identified. These genes encode products that act in a combinatorial fashion to prevent ectopic expression of specific developmental genes (for example, the genes of the *bithorax* and *Antennapedia* complexes). When *PcG* genes are mutated, their developmental targets are mis-expressed, resulting in the whole scale transformation of one set of segmental body patterns into another. This phenomenon was originally called *homeosis* (Bateson 1894) and so these *PcG* targets are called homeotic genes. Another group of genes, which comprise the *trithorax* Group (*trxG*), appear to operate in the opposite direction, maintaining prescribed homeotic gene expression through development. These two groups of genes therefore have antagonistic functions specifically affecting homeotic gene expression in a developmental context. But the fact that PcG proteins and HP1 exhibit some structural homology suggests that their silencing mechanisms might also overlap, especially since both mechanisms are epigenetic. This issue is still debated (Sass and Henikoff 1998), but there is genetic evidence that some modifiers of PEV do encode members of the *PcG* and *trxG* families (Dorn et al. 1993, Sinclair et al. 1998). Since PEV represents a constitutive rather than a developmental phenomenon, these results suggest that the two silencing mechanisms may have features in common.

HP1 clearly has a role in maintaining heterochromatin as a transcriptionally repressive environment, within which it would seem very unlikely that genes could function normally. But they do, and in addition, they also exhibit PEV, but of a reciprocal kind – a heterochromatic gene will shut down when translocated into euchromatin, and factors that enhance or suppress PEV for euchromatic genes will have the opposite effect on heterochromatic genes (Eberl et al. 1993, Howe et al. 1995, Elgin 1996, Clegg et al. 1998). So heterochromatic genes present something of a paradox, the resolution of which may help to explain the effect of chromatin structure on gene expression in general.

Heterochromatic sequences in *Drosophila*

As has already been described above, one of the principal discoveries concerning the regulation of gene expression in a chromatin context was made using the fruit fly *Drosophila melanogaster*. This organism has been a favoured model for geneticists for almost a century, and as such provides an enormously rich biological resource. It is easy to rear in the lab and manipulate genetically, yet it is a highly complex animal with an almost inexhaustible supply of mutable characteristics. *Drosophila* heterochromatin is organized into pericentric (around the centromeres) and telomeric (at the chromosome termini) blocks on the major autosomes and the X chromosome, and displays the typical cytological appearance described above. In addition, the Y chromosome is completely heterochromatic, and the tiny "dot" fourth chromosome appears to possess interspersed heterochromatic and euchromatic domains (Sun et al. 2000), reminiscent of the pericentric euchromatic-heterochromatic junctions of the larger chromosomes. *Drosophila* was one of the first model organisms to have its genome sequenced. The first release was published in March 2000 (Adams et al. 2000) and has since undergone three revisions. Currently (Celniker et al. 2002, Hoskins et al. 2002) the genome size is estimated to be 176 megabases (Mb), roughly a third of which is heterochromatic. 116.8 Mb have been assembled into large contigs spanning all six euchromatic chromosome arms and the remaining 20.7 Mb is assumed to be heterochromatic, leaving 38.5 Mb yet to be sequenced. Gene models are predicted using a combination of *ab initio* gene prediction computer programmes and evidence from cDNA sequences and BLAST homologies. Release 3 predicts 13,676 protein-coding genes of which 13,379 are in euchromatin, (98%), leaving 297 (2%) in heterochromatin. In fact this number may be closer to 400, since proximal euchromatic scaffolds overlap with distal heterochromatic domains on all chromosome arms, and these regions of overlap do contain genes. It is not yet clear whether the transition from proximal (towards the centromere) euchromatin into distal heterochromatin is abrupt, graded or mosaic, and what effect this may have on the genes in this region.

Heterochromatic genes in *Drosophila*

It has long been assumed that roughly a quarter to a third of all genes will mutate to a given phenotype (Miklos and Rubin 1996), which suggests that one might expect to identify by classical genetic means minimally 75-100 genes in heterochromatin. In fact only 32 genes in *Drosophila* centric heterochromatin have been identified to date: (Hoskins et al. 2002), and a mere handful have been characterized both molecularly and genetically: (Hilliker 1976, Devlin et al. 1990a,b, Biggs et al. 1994, Risinger et al. 1997, Hanai et al. 1998, Warren et al. 2000, Tulin at al. 2002), as shown in Table 1.1. Heterochromatic genes cover a wide range of biological classes and include essential functions. They tend to be very large, due to the expansion of repetitive sequences in their introns. This leads to a discrepancy in size estimation, since gene finding algorithms often miss-call exons as different genes when they are separated by large introns. Cloning individual exons and assembling the sequence can be a laborious process, since repetitive DNA is very unstable in bacterial cells, and cannot easily be physically mapped to any specific location. Clearly the cloning and characterization of heterochromatic genes pose particular problems for both molecular and bioinformatic analysis. Since the goal of any sequencing project is the complete sequence of all the genetic information in a genome, no project can be said to be complete until this difficult area is mapped. This region remains problematic in all genomes that have been sequenced to date (Mardis et al. 2002).

GENE	FUNCTION	MAP	GENE SIZE (GADFLY)*	REFERENCE
light (CG18028)	Vacuolar assembly	2L, 40D3-4	15,970bp	Devlin et al 1990
rolled (CG12559)	MAP kinase	2R, h41	50,341bp	Hoskins et al.2002
Drad21 (CG17436)	Cohesin subunit	?	22,039bp	Warren et.al. 2000
Parp (CG40411)	Poly (ADP-ribose) polymerase	3R, 81F	>150,000bp	Tulin et.al. 2002
Snap-25	Soluble NSF-attachment protein	3L het	222,192bp	Risinger et.al. 1997

Table 1.1: Selected heterochromatic genes in *D.melanogaster*
Examples of heterochromatic genes that have been characterized in some detail. Some of these genes are still annotated as fragments in the genome project database *(GADFLY = Genome Annotation Database).

Methods used to study functional units in heterochromatin

Cytological methods have been successfully employed to study heterochromatin in *Drosophila*; in particular specific sequences - even highly repetitive ones - can be physically mapped to the heterochromatic regions of mitotic chromosomes by variable stringency *in situ* hybridization. Normally physical chromosome mapping would make use of the giant and highly polytenized chromosomes that come from *Drosophila* salivary gland nuclei. However, these nuclei are in a perpetual interphase state, where the heterochromatin of all the chromosomes remains under replicated, and coagulates into an undifferentiated mass known as the chromocentre. Instead, highly condensed metaphase chromosomes ("mitotics") derived from brain tissue (neuroblasts) are used – here the chromosomes remain distinct, and the heterochromatic regions acquire a specific and reproducible banding pattern in the presence of certain dyes, which can be used for mapping. *Drosophila* heterochromatin has thus been divided into 61 regions: h1-h61. Mitotic chromosomes pose formidable technical difficulties due to their small size and highly condensed structure. However they have been efficiently used to localize specific satellite sequences (Lohe et al. 1993) and transposons (Pimpinelli et al. 1995), as shown in Figure 1.1. They are currently being used by the Berkeley *Drosophila* Genome Project (BDGP) to localize large sequence contigs to heterochromatin (Hoskins et al. 2002).

Classical genetics has always been the most effective approach when attempting to map genes in heterochromatin, since protein-coding sequences may be mutable to a visible or lethal phenotype. But even classical methods are problematic: deficiencies near the centromere often disrupt chromosome segregation causing non-recoverable cell lethals. Therefore a specialized genetic method is required in order to generate suitable deficiencies near the centromere in heterochromatin. In this method, chromosomes are irradiated to produce compound left and right arms, and irradiated again to restore the natural configuration (Baldwin and Suzuki 1971). During this process, deficiencies surrounding the centromere can be recovered. One major problem with this scheme is the potential for generating complex rearrangements, including duplication of genetic material, which can confound subsequent analysis. Nevertheless the procedure has been

used successfully to create a set of deficiencies in heterochromatin which have since been employed in a number of mutagenesis screens, resulting in a collection of lethal complementation groups. These have subsequently formed the basis for more detailed studies of heterochromatic loci. For example, a number of lethal complementation groups on the second chromosome were identified by this method (Hilliker and Holm 1975, Hilliker, 1976) and have been subsequently cloned, of which the *light* gene was among the first to be molecularly characterized. Initial findings in the Honda lab revealed the unusual presence of middle repetitive DNA both within and around *light* protein coding sequences (Devlin et al. 1990a).

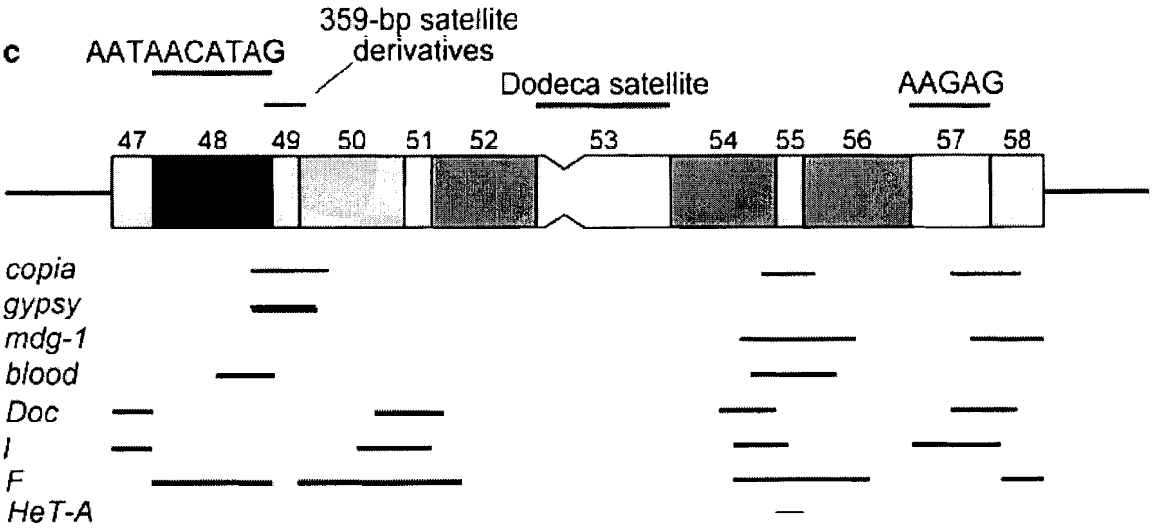


Figure 1.1: Schematic of the cytological map of 3rd chromosome heterochromatin

Positions of a subset of satellite and transposable element sequences are shown. Black-gray-white blocks represent different levels of fluorescence which result from DAPI staining techniques. Region h48 fluoresces brightly, h52 shows moderate fluorescence and the remaining cytological divisions exhibit low fluorescence. (Figure reproduced with permission from Patrizio Dimitri).

Previous work identifying genes in the heterochromatin of chromosome 3 in *Drosophila melanogaster*

The aforementioned methods used to characterize heterochromatic loci on the second chromosome have since been extended to the third chromosome (Marchant and Holm 1988 a,b). A large number of compound-reattachment deficiencies were generated, and used to organize and position at least 12 lethal complementation groups. Subsequent research carried out in the Honda lab has focused on the heterochromatin of 3L, which appears to contain the greater number of mutable genes. Through new screens (Schulze et al. 2001) and collaborations with other researchers (Kennison and Tamkun 1988, Vilinsky et al. 2002), a wide variety of new deficiencies using both EMS and radiation has been generated. This has permitted a finer scale mapping of the lethal complementation groups originally defined by Marchant and Holm (1988b), leading to a new estimate for both their number and relative order. Figure 1.2 depicts a purely genetic map of 3L heterochromatin; note that the distances between loci are inferred from breakpoint frequency, and may differ substantially from the cytological arrangement of the genes.

The next step in attempting to identify lethal complementation groups makes use of a transposable element mutagenesis (hybrid dysgenesis) screen. Transposable elements have revolutionized the process of cloning genes, since the mutations they produce are revertible (by causing the transposable element to excise from the gene), and can be used to rescue flanking DNA, possibly containing coding sequences of the gene in question. Two proximal genes in 3L heterochromatin were tagged in the Holm Lab by natural P elements using the *Birmingham 2* strain (Robertson et al. 1988), which carries a number of defective P elements on the second chromosome. These defective elements cannot transpose by themselves, but when crossed into a background containing an active transposase source, they comprise a powerful mutagenic potential. The screen employed to tag 3L heterochromatic lethal complementation groups is outlined in Chapter Two.

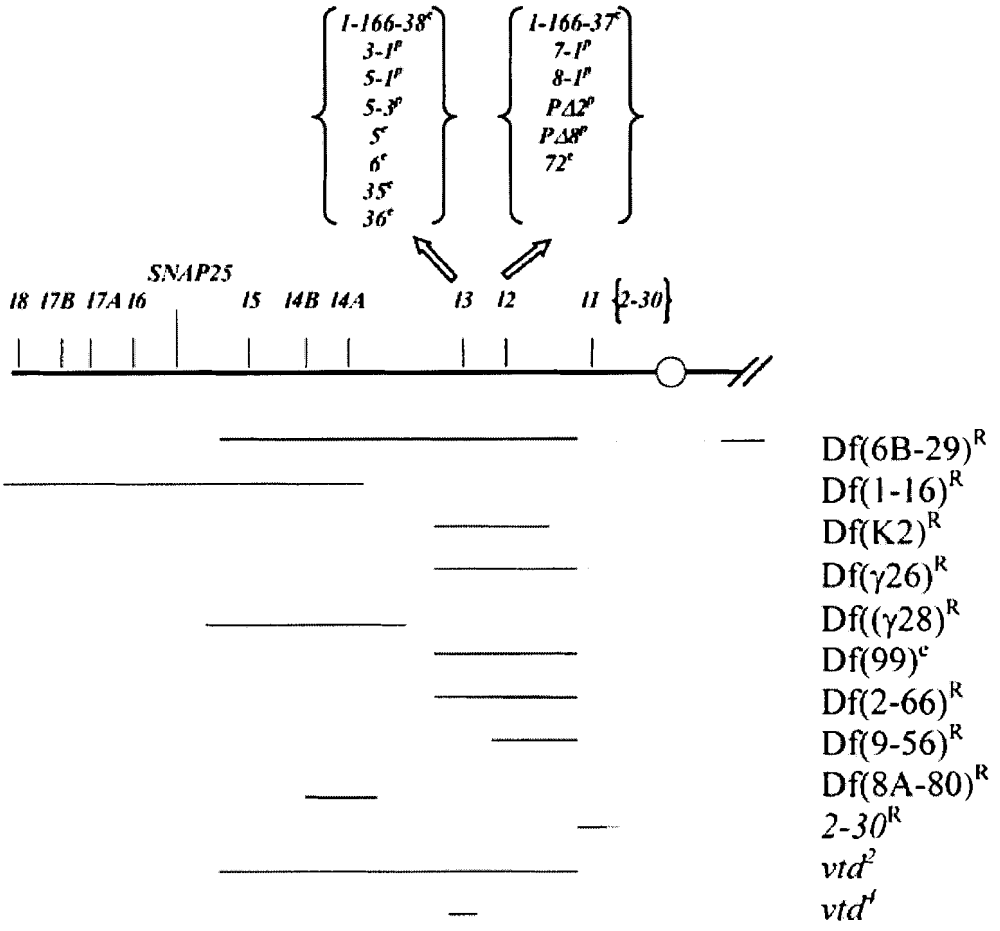


Figure 1.2: Genetic map of 3L heterochromatin

The heavy line at the top of the figure represents the heterochromatic portion of the left arm of chromosome 3. The circle on the right represents the centromere, and the lethal complementation groups (genes) are numbered *lethal 1 (l1)* through *lethal8 (l8)*. Different alleles for the first three genes are listed and enclosed in brackets. A superscript "e" indicates an EMS (chemically) induced mutation; a superscript "p" indicates a transposon-tagged (P element) mutation. The genes are mapped relative to chromosomal deficiencies for the region, which are listed to the right of this figure. Superscript "R" indicates a deficiency was generated by radiation, and "e" by EMS. Note that lines corresponding to the deficiencies demark the regions of DNA *missing* from the respective deficiency chromosome. Note also that *lethal 1* is not completely removed by any deficiency (indicated by a dotted line extending into the *lethal 1* region). Finally, note that the distance between the lethal complementation groups reflects breakpoint frequency, which does not represent actual cytological distance.

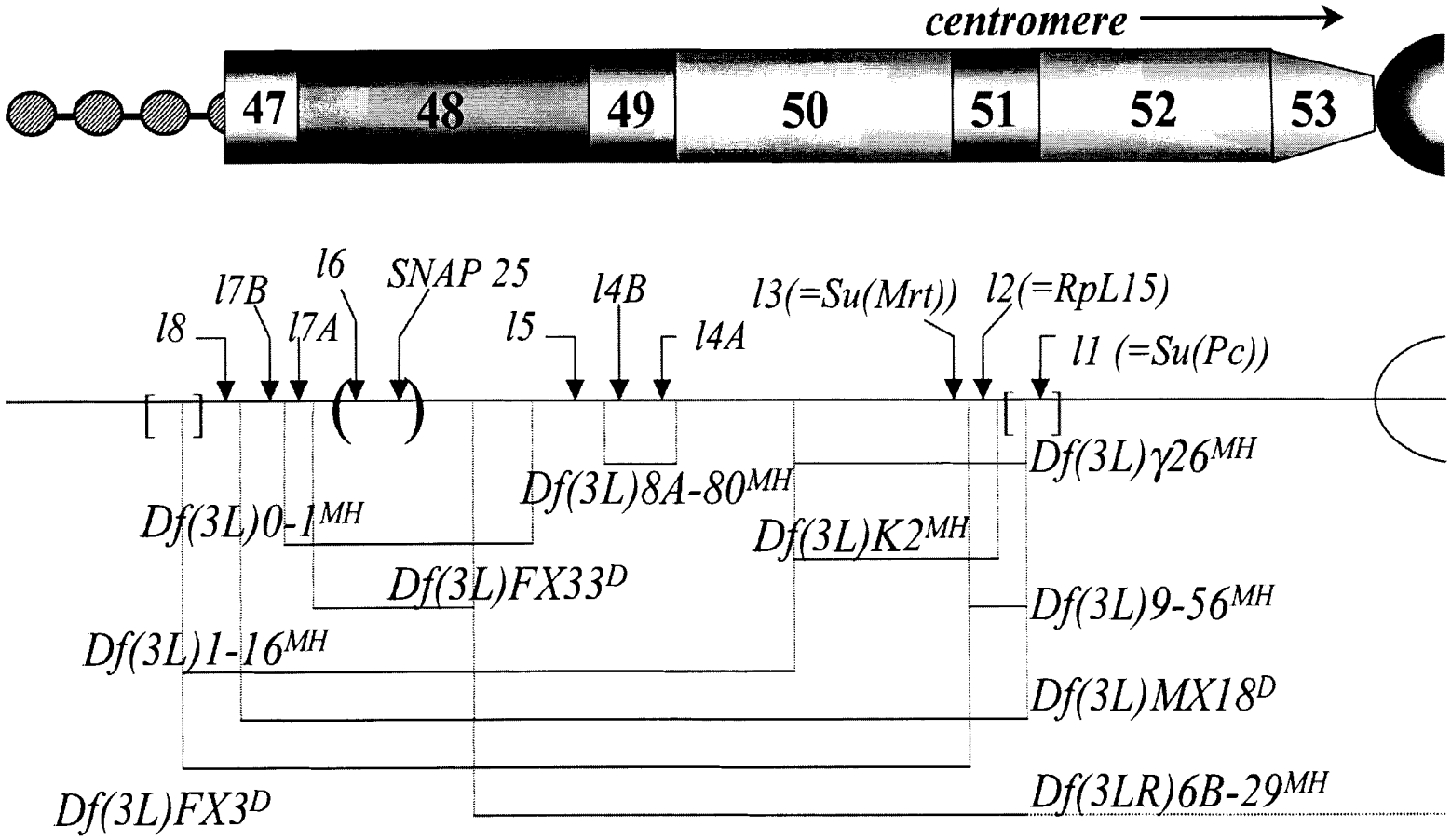
Cytological organization of genes in 3L heterochromatin

An attempt to describe the cytological organization of 3L heterochromatin was recently carried out by a group of researchers from Russia and Italy, who have mapped many of the compound-reattachment deficiencies to mitotic chromosomes (Koryakov et al. 2002). This served two useful purposes: firstly it distinguished those deficiencies likely to contain complex rearrangements (like duplications – for example *Df(3L)2-30*) and secondly, since the position of lethal complementation groups is known relative to these deficiencies, a clearer picture of how heterochromatic genes are physically organized was obtained. Their initial findings placed the 3L heterochromatic genes into three large groups, but a finer resolution can be inferred, by comparing the cytological regions removed by overlapping deficiencies, with the lethal complementation groups they contain. The result is shown in Figure 1.3, which represents an attempt to align the cytological and genetic data for this region.

This particular arrangement must remain hypothetical until the lethal complementation groups have been identified and physically mapped to the chromosomes, but it is supported by some inferential argument. For instance: the three most proximal genes map within the proximal borders of *Df(3L)1-16* and *Df(3L)2-66*, while the middle group of genes lies in the overlap between *Df(3L)1-16* and *Df(3LR)6B-29*. According to Koryakov et al. (2002), *Df(3L)1-16* and *Df(3L)2-66* overlap cytologically in distal h50, but *Df(3L)2-66* only removes the first three proximal genes. This suggests these genes are clustered within the narrowly defined cytological region h51. In addition, Koryakov et.al. have mapped the heterochromatic breakpoint of a particular inversion ($In(3R)hb^{bs23}$) to region h49, and this same breakpoint has generated a lethal allele of *lethal 4A*, placing this gene within another narrowly defined cytological region. A similar argument stands for the heterochromatic breakpoint of *In(3L)C90*, which is lethal in combination with existing lesions in *lethal 1*, placing this gene in the cytological region h51. Further detailed genetic and cytological mapping will undoubtedly refine this current picture.

Figure 1.3: Alignment of cytological and genetic maps of 3L heterochromatin.

The gray-shaded bar on the top of the figure represents heterochromatic regions h47-53, corresponding to the cytological banding pattern which results when mitotic chromosomes are treated with specific dyes. Underneath is a schematic of the genetic map of the same region. Positioning of genes relative to cytological divisions is based on published work that correlates the absence of cytological divisions with specific deficiencies for the region (Koryakov et al. 2002). Square brackets indicate regions where deficiency breakpoints are difficult/impossible to obtain; round brackets indicate regions where gene order has yet to be established.


centromere
47
48
49
50
51
52
53
l8
l7B
l7A
l6
SNAP 25
l5
l4B
l4A
l3(=Su(Mrt))
l2(=RpL15)
l1 (=Su(Pc))
Df(3L)γ26MH
Df(3L)8A-80MH
Df(3L)0-1MH
Df(3L)K2MH
Df(3L)FX33D
Df(3L)9-56MH
Df(3L)1-16MH
Df(3L)MX18D
Df(3L)FX3D
Df(3LR)6B-29MH

What cannot yet be portrayed in this map is the number of genes that do not mutate to a given phenotype. If the ratio of essential to non-essential genes is maintained in heterochromatin (Miklos and Rubin 1996), a large number of these genes must exist.

In the work that follows, I have established the identity of one of the proximal lethal complementation groups, and discovered another gene within 10kb which may not be essential. I have mapped and characterized both of these genes, in keeping with the overall objective of our lab to contribute to a complete map of this difficult region. In addition, I have cloned and characterized homologues for both genes in a related *Dipteran* species for which no genome project exists – *Drosophila virilis* – and present a comparative study which may help to explain the origin of heterochromatic genes and chromosome evolution.

CHAPTER TWO: Materials & Methods

Genetics

Drosophila culture conditions, stocks and strains

Flies were grown on standard cornmeal-sucrose medium with either Tegosept or proprionic acid as a mold inhibitor. Stocks were routinely maintained and crosses performed at room temperature (22°C), unless otherwise indicated. For embryo collections, apple juice agar plates were made according to Ashburner (1989).

Descriptions of most mutations, special chromosomes, and deficiencies used in this work can be found at the flybase website (http://flybase.bio.indiana.edu:82/). The genes, and extent of various third chromosome heterochromatic deficiencies, are shown in Figure 1.2 and 1.3 (Marchant and Holm 1988a,b, Schulze et al. 2001). To simplify nomenclature, 3L genes *l(3)80Fj* through to *a* (proximal to distal), are numbered outwards from the centromere i.e. *j = lethal 1 or 11,* through to *a=l8. Df(3LR)6B*-29, *Df(3L)1-16* and *Df(3L)9-56* are three overlapping third chromosome deficiencies bearing the recessive markers *ri* and p^p. *Df(3L)6B-29* lacks the 3L genes *l1,2,3,4A,4B* and *5*, and the 3R genes *l(3)81Fa. Df(3L)1-16* removes *l4A-l8. Df(3L)9-56* removes *l1* and *l2. Df(3L)99* was identified as an EMS-induced mutation that was lethal in combination with *Df(3L)K2* (see below), and fails to complement *l2, l3, l4A* and *l4B. Df(3L)FX3* was isolated in a screen for lesions in *SNAP25* (Vilinsky et al. 2002) and removes *l3-l8.* Strains bearing three previously reported *verthandi (vtd)* alleles, isolated as suppressors of *Pc* (Kennison and Tamkun 1988), were also used in the present study: *ru h th vtd2* e^s */TM3, Sb*; *vtd3/TM6C, Sb*; and *vtd4/TM3, Sb. vtd6* to *vtd11* were isolated as EMS-induced dominant suppressors of a *hedgehog (hh)* gain of function allele, *Moonrat (Mrt)* (Felsenfeld and Kennison 1995, Schulze et al. 2001).

Mutagenesis Screens

Screen for radiation-induced lethal mutations

ru h st p^p ss e^s/ ru h st p^p ss e^s males were treated with 3000-3700 rads of gamma radiation from a ^{60}Co source and mated to *Ly/TM3, Sb Ser* females. Single F1 *ru h st p^p ss e^s */TM3, Sb Ser* males (where * indicates a mutagenized third chromosome) were then crossed to *Df(3L)6B-29/TM3,Ser* or Df(3L)*1-16/TM3,Ser* females (3 females per vial), and stocks of putative lethal mutations were established from *ru h st p^p ss e^s */TM3, Ser* flies (Marchant and Holm 1988a,b)

Screen for P-induced mutations:

Birm 2/Birm 2; ry^{506}/ ry^{506} females were crossed *en masse* to *CyO/Sp*; *P*[*ry+ Δ2.3*] *Sb/TM6, Ubx* males. Their *Birm 2/Sp*; *Sb Δ2.3/* ry^{506} and *Birm 2/CyO*; *Sb Δ2.3/* ry^{506} male progeny were crossed *en masse* to *Ly/TM3* females and *+/Birm 2*; ry^{506}*/TM3*, *+/CyO*; ry^{506}/TM3 and *+/Sp*; ry^{506}*/TM3* male progeny were crossed singly to either *Df(3L)6B-29/TM3* or *Df(3L)1-16/TM3* females. (Schulze et al. 2001).

Screen for EMS-induced mutations in the proximal region of 3L heterochromatin:

ri p^p/ ri p^p males were fed 50mM EMS and mated *en masse* to *w-/w-*; *TM3, Sb e/TM6B, Tb e* females. Individual *ri p^p */TM3, Sb e* or *ri p^p */TM6B, Tb e* F1 males were then crossed to three to five *Df(3L)K2,* e^s */In(3L)DcxF, D* or *Df(3L) γ26/TM3, Sb* females. F2 progeny resulting from these individual cultures were examined for putative lethal mutations, and *ri p^p */TM3* or *ri p^p */DcxF* flies from appropriate cultures were used to establish stocks of lethals (Schulze et al. 2001).

Removing background P elements.

Background chromosomes were replaced in the P mutant strains *3-1* for *lethal 3*, and *PΔ2, PΔ8,* and *8-1* for *lethal 2.* P mutant males (marked with ry^{506}) were crossed to the doubly balanced, reciprocally translocated strain ap^{Xa}*/CyO;TM3.* F1 *Cyo;TM3* progeny

were then crossed to ap^{Xa}/*CyO;TM3* again, and the F2 *CyO, Sb+* males were crossed to a *Df(3L)K2*p^{p}, *e/TM3Sb,Ser, e,* and the P mutant chromosome recovered over TM3 in the F3 generation.

Southern analysis initially showed that *8-1, PΔ2* and *PΔ8* contained few if any background P elements after the background chromosomes were replaced. However the same analysis revealed that *3-1* still contained many background P elements, and so the euchromatic arms of the third chromosome from this stock were crossed off. *3-1/TM3* males were crossed to homozygous virgins from the multiply marked third chromosome strain of *rucuca* (*ru h th st ri* p^{p} *cu sr e ca).* The F1 *3-1/rucuca* females were then crossed to males from another multiply marked third chromosome stock *ru h st ry e/ru h st ry e,* and F2 males carrying the recombinant chromosome were individually crossed to *Df(3L)99/TM3* females. Lines producing no wild type flies (*ru h st (3-1) ry e /Df(3L)99)* were established from the *ru h st (3-1) ry /TM3* sibs. Separate sets of males from these stocks were then crossed to *st in Ki eg*^{2} p^{p} e^{s} */ st in Ki eg*^{2} p^{p} e^{s} females, and the F1 *ru h th st (3-1) ry / / st in Ki eg*^{2} p^{p} e^{s} females were crossed to *Df(3L)99/TM3Sb* males. Any F2 progeny that were *Ki,* p^{p} e^{s}*/TM3Sb* or p^{p} e^{s}*/TM3Sb* were individually tested over *Df(3L)99/TM3Ser.* Only one stock, *3-1*^{29}*/TM3Sb* was lethal, and so a stock from this line was established, and tested by Southern analysis.

Generation of lethal 2 P excision lines

This experiment was performed to generate excisions from both *PΔ8* and *8-1*, and took advantage of the fact that *PΔ8/PΔ2* and *8-1/ PΔ8* transheterozygotes can survive to adulthood. The screen was designed to capture independent P excision events that were completely lethal in either of these combinations. +/Y; *PΔ8* ry^{506} */TM3Sb* males were crossed to *yw/yw; KiΔ2-3/ KiΔ2-3* females, and approximately 50 F1 *yw/Y; KiΔ2-3/ PΔ8* ry^{506} males were individually crossed to +/+; *Df(3L)99/TM3Ser* females. Twenty F2 *PΔ8* ry^{506}** /TM3 Ser* males were individually crossed to *PΔ2* ry^{506} */TM3Sb* females, and stocks were established from lines which produced no *PΔ8* ry^{506}**/ PΔ2* ry^{506} offspring. The same

experiment was carried out in parallel starting with *8-1* ry^{506}*/TM3*, selecting for lethality of *8-1* ry^{506}*/ PΔ8* ry^{506} in the F3 generation.

Notch interaction assay

This experiment was designed firstly to confirm earlier work (Hart et al. 1993) demonstrating an interaction between ribosomal protein genes and genes involved in wing morphogenesis; secondly to create a sensitized genetic background in which to study the effects of reduced *Su(var)* dose on a heterochromatic gene. A weak *Notch* allele, N^{55e11}, was employed (http://flybase.bio.indiana.edu:82/). N^{55e11}*/FM7* virgins were initially crossed to various deficiencies (*Df(3L)γ26, Df(3L)99, Df(3L)K2, Df(3L)9-56, Df(3L)6B-29)* and mutant alleles (*PΔ2, PΔ8, 8-1, 7-1, 1-166-37* and *72)* for *lethal 2,* as well as deficiencies and alleles outside *lethal 2* (*Df(3L)8A-80, Df(3L)2-30,35, 1-16-0)* for controls. The N^{55e11}*/+; Df* or *m/* + F1 progeny (where *Df* denotes a deficiency, and *m* a mutant allele) were examined for enhancement of the notched wing margin and thickened vein phenotype characteristic of the N^{55e11} allele. Subsequently, a *Su(var)2-*5^{01}*/CyRoi; 72/TM3Sb* stock was constructed, and males were crossed to N^{55e11}*/FM7* virgins. The F1 N^{55e11}*/+ or Y; Su(var)2-*5^{01}*/+; 72/+* progeny were compared with their N^{55e11}*/+ or Y; CyRoi/+; 72/+* siblings, and to progeny from the first part of the experiment (where the *Su(var)* was omitted.) A stock of a *Su(var)2-*5^{01}*/CyRoi; 1-16-0/TM3Sb* was used in the same set of experiments as a control. Since the N^{55e11} allele shows variable penetrance, a large number of wings from different individuals were examined, and representative samples showing a subjectively characterized "average" degree of severity were selected for photography.

Isolation and segregation analysis of transgenic flies

Injection procedures are described elsewhere (see below). Injected embryos were collected on apple juice agar plates, and those which survived to hatch were transferred to vials containing standard cornmeal-sucrose medium with either Tegosept or proprionic acid as a mold inhibitor. Adult survivors were individually crossed to the same strain which was injected (*iso-yw* for the 3.5kb HIII genomic constructs, w^{1118} for all others),

and F1 flies with a [*w+*] eye colour were outcrossed to the doubly balanced *w/w,Y; ap^{Xa}/CyO, TM3* stock, and the progeny from subsequent generations was scored for segregation. If the [*w+*] transgene was hemizygous in males but both homozygous and heterozygous in females, the transgene was presumed to be on the X chromosome. This was confirmed by establishing attached-X stocks where possible. If the transgene was observed to be homozygous in the presence of *Sb*, it was presumed to be on the second chromosome, and if homozygous in the presence of *CyO,* on the third. Transgenes that followed none of these patterns were presumed to be on the fourth chromosome.

Germ line rescue experiments

Four levels of *lethal 2* rescue were tested: lethality, viability, fertility and the *Minute* phenotype. Stocks were constructed which contained both the transgene and either a deficiency for *lethal 2* (*Df(3L)9-56 ri p^p/TM3 ri p^p e* or *Df(3L)K2 e//TM3 ri p^p e*) or a mutant allele (*72 ri p^p/TM3 ri p^p e, PΔ8 ry^{506}/TM3, 8-1 ry^{506}/TM3*). Individuals from these stocks were either directly crossed to other *lethal2* deficiencies or alleles and then subjected to heat shock (pCaSpeR lines – see below), or if they were pUAST lines, they were first crossed to a driver strain (*w/w,Y, HSGAL4/CyO, w/w,Y; armadillo GAL4/armadillo GAL4* or *yw/yw,Y; ACT5-C/CyO*) and males heterozygous for transgene, driver and *lethal 2* lesion were then crossed to females from *w- lethal 2* deficiency or mutant stock. In many cases, shared markers (*ri)* were employed to score for the rescued genotype. In those cases where there were no shared markers between the mutant chromosomes (for instance, *Df(3L)K2/72)* the ratio of *Sb:Sb+* was recorded in the progeny females. X-linked transgenes were almost exclusively employed, and the male progeny therefore served as an internal control. The same set of experiments was also performed using different combinations of P alleles, particularly *PΔ2/ PΔ8,* since this combination produces a significant number of transheterozygote escapers, and thus serves as a good test for rescuing viability, fertility and the *Minute* phenotype. In these crosses, the final generation was taken out of a w- background, and progeny were scored for ry^{506}/ ry^{506}

Analysis of lethal 2 gene expression in a Su(var) mutant background.

The mutations in *Su(var)2-5* are described in Lu et al. 2000, but in summary, *yw/Y; Su(var(2-5)^{04} /CyO, y+* males were crossed to *yw/yw*; *Su(var)2-5^{149}/CyO, y+* females, and the *yw/yw*; *Su(var(2-5)^{04} / Su(var)2-5^{149}* progeny L3 larvae (no functional HP1 dose) were selected on the basis of mouth hook colour (dark brown as opposed to black). Their *yw; Su(var)2-5^{x}/CyO* siblings (one functional HP1 dose) were also collected. *yw/,Y,yw* larvae were collected separately (2 functional HP1 doses). The larvae were weighed, snap frozen in liquid nitrogen, and stored at -70°C until needed. Both *Su(var)2-5* mutants used in this experiment encode truncated proteins.

Cytology

Chromosome preparation

3rd instar larvae (wandering phase) were grown on standard bottle media (see above) at 18°C in uncrowded conditions. Salivary glands were dissected in 0.1% Triton X-100 in PBS pH 7.5, and fixed for 10-30 seconds in a drop of 50 ul tritonX-100 10%, 400ul PBS, 50ul 37% p-formaldehyde, on siliconized coverslips. The glands were then transferred to a drop of 50ul 37% p-formaldehyde, 200ul H2O, 250ul 100% acetic acid on another siliconized coverslip for 2-3 minutes. The glands on coverslips were then taken up with poly-L-lysine treated slides and squashed. Suitable chromosome spreads were selected after examination under phase contrast, and the slides frozen in liquid nitrogen. The cover slips were removed with a razor blade and the slides washed twice for 15 minutes each in PBS. Slides were stored for no more than one week, in 100% methanol.

Immunostaining

Slides were washed 2X15 minutes in PBS, and blocked for 1 hour at room temperature in blocking solution (BSA 3%; 0.2% (w/v) NP40; 0.2% (w/v) Tween 20; 10% non fat dry milk). 20 µl of affinity purified primary antibodies (i.e rabbit polyclonal antibodies; dilutions 1:50 to 1: 500 in blocking solution needed to be adjusted for each individual primary antibody) were added to each slide, covered with a coverslip and incubated for

1hour at room temperature in a humid chamber. The slides were then washed for 15 minutes in PBS, 300mM NaCl, 0.2% NP40, 0.2% Tween20-80; 15 minutes in PBS, 400mM NaCl, 0.2% NP40, 0.2% Tween 20-80, and then rinsed in PBS. The coverslips were removed, and 20 µl of diluted secondary antibody (i.e. either a fluorescent labeled like Cy3-Anti-Rabbit IgG (Fc) (Dianova), or Anti-Rabbit IgG (Fc) HRP Conjugate, Promega Kat. Nr.: W4011, 1:100 dilution) + 2% normal (goat) serum in blocking solution were then added, and the slides were covered with coverslip and incubated for 40 minutes at room temperature in humid chamber. After rinsing in PBS, the slides were washed for 15 minutes in PBS, 300mM NaCl, 0.2% NP40, 0.2% Tween20-80; 15 minutes in PBS, 400mM NaCl, 0.2% NP40, 0.2% Tween20-80, and rinsed again in PBS.

DAPI staining

Slides were stained for 10 minutes in 1 ug/ml in DAPI/PBS (DAPI is 4'6' Diamidino-2-phenylindole), and then washed for 5 minutes in PBS. The chromosomes were mounted in 40ul Mowiol (made as follows: add 2.4 g Mowiol to 6 g Glycerol and 6 ml H2O; mix for 3 hours, add 12 ml 0.2M Tris-Cl pH 8.5 and incubate 10 minutes at 50°C with mixing; insoluble material is pelleted by centrifugation for 15 minutes at 5000xg; add DABCO (Diazabicyclo(2.2.2.) octane: Merck #803456) to final 2.5% to the solution as anti-bleaching agent.) The chromosomes were then analysed under a fluorescent microscope.

Enzymatic reaction

100µl of 0.5 mg/ml DAB-solution (Diaminobenzidinetetrachloride; Sigma # D5637) plus 0.01% H2O2 (Merck # 7210) were added. Staining was followed under phase contrast. The reaction was stopped by dipping slides in PBS, and washing for 10 minutes in PBS. The chromosomes were stained for 10-20 seconds in Giemsa ((Merck # 9204; 1:130 dilution in 10 mM sodium phosphate buffer pH 6.8), rinsed in H2O, mounted in 99.5% glycerol and immediately examined under the microscope.

Microscopy and Photography

Cuticle phenotypes were examined in embryos according to Wieschaus and Nüsslein-Volhard (Chapter 6 in Roberts D.B. 1998). Wings were dissected from adult flies and mounted in aquamount. Bright field images were acquired using an Olympus VANOX microscope, or a Zeiss Axioplan microscope, with Northern Eclipse (version 5.0) software.

Molecular Biology

Bacterial and phage strains, vectors and culture conditions

The *E.coli* bacterial strains XL1-Blue and DH5α were used to propagate plasmid DNA (pBluescript, pFLc1, pOT2A, pOT7B, pTOPO 2.1, pBR322, pUAST, and pCaSpeR) by electroporation (XL1-Blue) and heat shock (DH5α) transformation. Bacterial cultures were routinely grown in 2YT medium (Sambrook 1989) with appropriate concentrations of antibiotics where necessary (100ug/mL ampicillin or 25ug/mL chloramphenicol). The *E.coli* strain LE392 was used to propagate genomic libraries in phage λEMBL3, and C600 to propagate cDNA libraries in phage λgt11. Phage cultures were routinely grown in NZYM medium (Sambrook 1989). Culture growth was always in a shaking incubator at 37°C.

Library and clone sources

The *Drosophila virilis* cDNAs were isolated from a mixed embryonic plasmid library constructed for the Berkeley *Drosophila* genome project, and kindly provided by Ling Hong. The *Drosophila melanogaster* cDNAs used in this work were ordered from Resgen and therefore originally came from the following EST libraries: RE01373 (Riken embryo) for *RpL15*, and LD10689 (Ling Hong, 22 hr embryo) for *Dbp80*. The *RpL15* cDNA was also isolated from a phage λgt11 library, which was kindly provided by John Tamkun's laboratory (Tamkun 1991).). The genomic regions for *vRpL15* and *vDbp80* were subcloned from a phage λEMBL3 *D. virilis* genomic library (Thummel 1993), and

the genomic region for *mRpL15* (which was subsequently used to make the transgene constructs) was subcloned from a λEMBL3 *D. melangaster* genomic library (*ibid.*)

Isolation of Plasmid and Phage DNA

Plasmid DNA for general use (restriction digests, generation of probes, subcloning of genomic DNA and stock maintenance) was isolated using standard alkaline lysis procedures as described in Sambrook (1989). Plasmids with inserts that were to be sequenced were further purified by PEG precipitation (2.5N NaCl in 20% polyethylene glycol) and resuspended in deionized, distilled autoclaved water. Germline transformation constructs were purified using the Qiagen endotoxin-free plasmid maxi-prep kit (cat. no. 12362). Phage DNA was isolated following the protocol for plate lysates from the Rubin Manual (1986). Where necessary, quantitation was performed using an Ultrospec III (Pharmacia).

Isolation of genomic DNA

Isolation of Adult DNA

Drosophila genomic DNA was isolated using the method outlined by T. Jowett (Chapter 11 from Roberts D.B. 1998). The procedure outlined was designed and used to extract DNA from a sample of 200 flies, but was often scaled down for smaller numbers, so the whole procedure could be carried out in 1.5mL eppendorf tubes. In this variation, up to 30 flies were homogenized in 750ul Fly Lysis Buffer (100mM Tris-HCl pH 8, 50mM NaCl, 50mM EDTA, 1%SDS, 0.15mM spermine, 0.5mM spermidine – made up in advance, not added just prior to use as described in Roberts.). 2ul of a 10mg/mL stock of Proteinase K were added, and the preps were incubated in a 37°C water bath for up to four hours. The preps were then phenol extracted once, phenol:chloroform extracted twice, and chloroform extracted once (equal volume). After ethanol precipitation, the pellets were washed in 70% ethanol, dried in a vacuum dessicator and resuspended in 70ul of deionized, distilled autoclaved water or TE8. RNAse was added to a concentration of 100ug/mL, and the samples were quantitated by spectroscopy, and run

on an agarose gel to check integrity. This DNA was used for Southern analysis, PCR and the generation of size-selected plasmid libraries.

Isolation of DNA from single embryos.

This procedure was adapted from Kevin O'Hare's method (Hatton and O'Hare, 1999), designed to extract DNA from single embryos homozygous for specific mutations or deficiencies. Individual embryos were collected from apple juice agar plates, placed in 0.6mL ependorf tubes and homogenized in 13ul of embryo lysis buffer. 2ul of a 20ug/ul stock of proteinase K were added, and the preps were incubated in a 37°C water bath for 30 minutes. Thereafter the preps were heated to 99°C in a PCR machine for 10 minutes, and the DNA was used immediately in PCR reactions, both for mapping specific coding regions under specific deficiencies, and for generating inserts for sequencing mutant alleles.

Restriction digests

All restriction digests were carried out with enzymes purchased from Invitrogen or NEB, according to their protocols. For plasmid digests, 200-1000 ng of DNA was digested with 1-3 units of enzyme at 37°C, for minimally 1 ½ hours. For genomic DNA digests used in Southern analysis, 10-20ug of DNA was digested in a volume of 300ul, using up to 120 units of enzyme (Invitrogen high concentration enzymes, 40u/ul), and these digests were carried out for minimally 4 hours. In all cases, reactions were stopped at 65°C for ~20 minutes, and the genomic digests were precipitated and resuspended in a volume suitable for gel electrophoresis (usually 20 ul).

Isolation of RNA

Total RNA was isolated from various *Drosophila* developmental stages using Trizol™ reagent (Invitrogen), with the following variations from the published protocol: up to 50 adults, or 13-15 L3 larvae or pupae were snap frozen in liquid nitrogen, and homogenized in 500ul Trizol™, using 1.5mL ependorf tubes. After centrifuging for 10 minutes at

12,0000g, the cleared homogenate was transferred to a new tube, and extracted with chloroform. The RNA was then precipitated with isopropanol (1/2 volume) and washed with 70% ethanol made with DEPC treated water. The pellets were dried and resuspended in DEPC treated water and quantitated by spectrophotometry. This RNA was stored at -70°C until needed. It was used for Northern analysis, or extraction of poly(A)+ RNA, which was accomplished using the Oligotex^{R} mRNA minikit (Qiagen).

Agarose Gel Electrophoresis

DNA gels

DNA samples were loaded on 0.7%-1.5% agarose gels depending on the size of the bands to be resolved. Genomic digests were generally run in 0.7% gels overnight. Ethidium bromide was added to a final concentration of 500ng/ul for visualization of the DNA. All gels were run in 0.5X TBE (Sambrook 1989). Gels were photographed with a UVP documentation system.

RNA gels

30ug of total RNA per sample, and 3.5ul of RNA ladder (Sigma) were mixed for loading as described in Sambrook (1989) and electrophoresed in a 1% agarose formaldehyde gel, prerun at 70 volts for 20 minutes before loading, in 1X MOPS. The electrophoresis chamber was treated beforehand by soaking in 0.2%SDS overnight, and all glassware used to make the MOPS solution was baked at 200°C for at least 6 hours. Gels were photographed with a UVP documentation system.

Isolation of DNA from agarose gels

DNA was digested with restriction enzymes and electrophoresed through agarose gels in clean 0.5X TBE as described above. If the digestion products were to be used for cloning, no pictures under short wave UV light (302nm) were taken. The band(s) were excised from the gel under long wave UV light (360 nm), cut into pieces, and extracted from the agarose according to the GFX PCR DNA and Gel band purification kit (cat. no. 27-9602-

01). For construction of size-selected libraries, restriction digested genomic DNA of a specific size range was cut from the agarose gel, placed into clamped spectropore 0.2% sodium bicarbonate/1mMEDTA treated dialysis tubing and electroeluted within an electrophoresis chamber (Sambrook 1989). The buffer and DNA were then removed from the tubing, the DNA was phenol-chloroform extracted, butanol precipitated, washed, dried and resuspended in TE8.

Cloning

Subcloning from phage

Genomic inserts from λEMBL3 (*D.virilis* and *D.melanogaster* genomic libraries) were cut with Sal I to liberate the inserts. The restriction digest products were then diluted, run on an agarose gel, transferred to nylon membrane by Southern blotting (see below), and hybridized with relevant cDNA probes (see below) in order to establish which bands contained coding regions. These bands were then extracted (see above) and cut again with Eco RI. The digests were split into two halves – one half was used in a ligation reaction with Eco RI cut pBluescript, and the other half in a reaction with EcoRI/SalI cut pBluescript. Small SalI fragments were also gel extracted and cloned into Sal I-cut pBluescript. All single enzyme-digested plasmids were dephosphorylated with Shrimp Alkaline Phosphatase (Boehringer Mannheim) according to the manufacturer's instructions. Ligations were carried out in a volume of 15-20ul, using 1-2 units of T4 DNA ligase (Invitrogen), and 5uM ATP at 15°C overnight (using a PCR machine). Transformations, plasmid preparations and diagnosis were carried out as described above. The same procedure was used to subclone the *RpL15* cDNA from the λgt11 library yielding a single EcoRI fragment which contained the entire cDNA. This cDNA unfortunately appeared to contain a small rearrangement that made it unsuitable for most experiments except for generating probes.

Cloning PCR products

PCR amplified DNA was gel extracted as described and cloned into the T-tailed vector pTOPO 2.1 using the TOPO TA cloning kit (Invitrogen cat. no. 45-0046). The protocol was followed in general, except that the One-Shot™ competent cells were sometimes replaced with either XL1-Blu or DH5α, and 2YT medium was often used in place of SOC.

Generation of a size selected library

Genomic DNA was digested with high concentration Eco RI and electrophesed as described above, and the size selected gel extracted DNA was ligated *en masse* to Calf Intestinal Phosphatase (Invitrogen) treated EcoRI-cut pBluescript. The concentrations of T4 ligase and ATP were the same as for subcloning (see above) but scaled up for larger volumes.

Constructs for germ line injection

The 3.5kb HIII genomic RpL15 fragment was end-filled with Klenow (Invitrogen) according to the manufacturer's instructions, and blunt ligated into Stu I cut pCaSpeR, which was dephosphorylated using Shrimp Alkaline Phosphatase (Boehringer Mannheim).

The 2.2kb BglII-Hind III *RpL15* genomic fragment was subcloned from the 3.5kb HIII pCaSpeR construct, and ligated into appropriately cut pUAST. The RE01373 *RpL15* cDNA was cut out of pFlc-1 using BamH1 and EcoRI. The fragment was then Klenow end-filled (see above) and blunt ligated into StuI cut SAP-treated pCaSpeR. Bam HI/EcoRI cut RE01373 was also ligated into BglII/EcoRI cut pUAST.

The constructs (and helper construct pII25.7 *wings clipped*) were purified for injection using an endotoxin-free kit (Qiagen – see above). The injection mix consisted of 600ng of construct plasmid and 400ng of helper plasmid, 1X PBS (autoclaved and filter sterilized), 2ul glycerol in a total volume of 20ul. This was centrifuged for approximately

20 minutes at top speed, and the upper 18ul was transferred to a fresh tube and used for injection.

Embryo injection

iso-yw or w^{1118} flies were collected in egg-lay bottles and timed to lay synchronously over several days at room temperature. Each round of injections took no more than 40 minutes to prevent the accumulation of cellularized embryos. Eggs were collected and dechorionated: either manually, or for 30 seconds in 50% bleach, then lined up on double sided tape which was attached to a glass cover slip. Prepared eggs were dehydrated briefly (~20-40 seconds depending on the day's humidity) with a hair-dryer, and then immersed in 50:50 light:heavy halocarbon oil (Lab Scientific). The coverslips with eggs were then mounted onto the stage of a Leitz laborflux microscope fitted with a micromanipulator. The injection controller was an Eppendorf Model 5242, which uses pressurized nitrogen. Injected embryos were placed on an apple-juice agar egg lay plate for recovery at 18°C. Newly hatched L1 larvae were transferred to food vials and maintained at room temperature.

Southern Analysis

DNA transfer

Restriction digests and electrophoresis conditions were carried out as described above. For genomic Southerns, 10-20 ug of DNA was loaded per lane; for clone blots, 10-100ng was loaded per lane. After the gels were photographed, the marker lane was removed, the gels measured, and were then treated as described Sambrook (1989), with the following variations. Denaturing was carried out in a solution of 1.5N NaCl, 0.5N NaOH, for 20-30 minutes on a shaking table, followed by two neutralizing washes (0.5M Tris-HCl pH7.5, 1.5N NaCl) of 20 minutes each. The gels were rinsed in deionized distilled water between each wash, and the depurination step was omitted. Transfer to nylon membrane (Hybond N+ Amersham) was carried out in 10X SSC overnight, and the DNA was covalently linked to the membrane by exposure to UV light using a Stratalinker made by Stratagene.

Probe labeling and hybridization conditions

Probes were generally digested and purified away from vector sequences as described above (GFX). 20-100ng was labeled with ^{32}P, following instructions from a Boehringer Mannheim random-primed DNA labeling kit (cat. no.1-004-760). For intraspecific Southerns, blots were prehybridized at 68.5°C for at least 2-3 hours in FSB-7% SDS (50mM Sodium pyrophosphate, 100mM Sodium dihydrogen phosphate, 7% SDS), with 100 ug/mL herring testis or salmon sperm DNA used as a blocking agent. Hybridization took place overnight in the same buffer at the same temperature, and the blots were washed 2-3X in FSB-1%SDS, also at 68.5°C. For interspecific southerns, prehybridization and hybridization were carried out at 55-58°C in a buffer composed of 2X Denhardt's (made from a 50X stock of 1% Ficoll (Type 400), 1% Polyvinylpyrrolidone, 1% Bovine Serum Albumin), 6X SSC and 0.5%SDS. The wash solution was the same except the Denhardt's was omitted. After washing, the blots were drained, wrapped in Saran, and exposed to X-ray film, (Amersham Hyperfilm™ MP) in a light-tight cassette with an intensifying screen, usually at -70°C.

Library Screening

Phage libraries

Phage libraries were screened in three stages: for the primary screen plates were almost confluent, and the secondary and tertiary screens were plated at low titre to ensure unique isolates. Phage were plated in NZYM Top agarose (Sambrook 1989), grown overnight and then transferred to 4°C for at least two hours before transfer to filters (lifts). In all cases, Hybond-N+ (Amersham) circular filters were used for plate lifts, and treated with the same denaturing and neutralizing solutions described above for Southern analysis. The solutions were pipetted onto Saran, and the filters were placed plaque-side up for 3-5 minutes for each of two denaturing washes, and 7 minutes for a single neutralizing wash. The filters were rinsed briefly in 2X SSC, drained on whatman filter paper and irradiated with UV light using a Stratagene stratalinker. Probe preparations, hybridization and wash conditions were identical to those described above for Southern analysis. Positive plaques

were picked from the plates, placed in 1 mL of SM (Sambrook 1989), and left at 4°C to elute overnight. These plugs were then titered and plated at appropriate dilutions the next day for subsequent screens, or phage preps (see above).

Plasmid libraries

Plasmid libraries were screened in three stages, with the primary screen plated with liquid culture to near-confluence, the secondary plates patched, and the tertiary plates streaked to ensure unique isolates. Colonies were grown overnight on 2YT agar plates with the appropriate antibiotic, then transferred to 4°C for at least two hours before treatment with Hybond nylon filters (lifts). Lifts were carried out exactly as described for phage plates, with some variations. Immediately after lifting, the filters from the colony plates were transferred colony –side up to fresh plates (with appropriate antibiotic) and both sets of plates were left to recover at 37°C for at least 4 hours. The filters were then treated with denaturing and neutralizing solutions, but the wash in 2X SSC was more thorough than for the phage lifts: the bacterial colonies which had grown on the filters during recovery were actively scrubbed off, and the filters were then rinsed a second time in 2XSSC. Thereafter, UV treatment, hybridization and wash conditions were exactly as described above for Southern analysis.

Northern Analysis

30ug of total RNA or 10ug polyA+ RNA was loaded per lane and electrophoresed as has already been described above. After the gels were photographed, the marker lane was removed, the gels measured, and were then treated as described in Sambrook (1989), with the following variations. Gels were washed two times in 10X SSC for 30 minutes each, and transfer to Hybond-N+ (Amersham) filters was carried out overnight also in 10X SSC. The Northern blots were disassembled the next day, and the RNA was covalently linked to the membrane by exposure to UV light using a Stratalinker made by Stratagene. Probe preparation, hybridization and wash conditions were exactly as described for Southern analysis. Gene expression levels were measured by exposing labeled Northern blots to a Storage Phosphor screen (Amersham) which was scanned by a Typhoon 9410

phosphoimager and analyzed with ImageQuant 5.2 software, using the "volume analysis" option for quantitation (www.mdyn.com)

PCR amplification

Taq polymerase from Invitrogen, or Vent polymerase from NEB were used to amplify DNA for cloning, probe preparation and general diagnostic tests. Templates were either cloned or genomic and conditions for amplification were generally as described in the manufacturer's protocols, with some minor changes. Reactions were almost always 25ul, but if the volume increased, all amounts were scaled up accordingly. For cloned template, 10-100 ng was used per reaction, and for genomic, 100-500ng was used.

Reaction mix

For single band amplification with Taq, the reaction mix was made as follows: 2.5ul 10X buffer (supplied with enzyme), 1ul 25mM Magnesium, 2ul 2.5mM NTPs (company), 1ul each of 10uM primer working stocks, 1-2.5units Taq polymerase, and deionized distilled autoclaved water to a final volume of 25ul.

For multiplexing, the reaction mix was made as follows: 2.5ul 10X Buffer (supplied with enzyme), 1ul 25mM Magnesium, 4ul 2.5mM NTPs (company), 1ul each of 10uM primer working stocks, 1-2.5units Taq polymerase, and deionized distilled autoclaved water to a final volume of 25ul.

For amplification with Vent polymerase, the reaction mix was as follows: 2.5ul 10X Buffer (supplied with enzyme), no magnesium, 2ul 2.5mM NTPs (company), 1ul each of 10uM primer working stocks, 1-2.5units Vent polymerase, and deionized distilled autoclaved water to a final volume of 25ul.

Cycling profiles

All PCR reactions were carried out using a PCR Sprint (Thermo Hybaid) machine.

For diagnostic PCR used to confirm identity and orientation of plasmid inserts using Taq polymerase, the cycling profile was as follows: an initial "hot start" of 2 minutes at 94°C, followed by 28-32 cycles of: 94°C/30s (denaturing); X°C/40s (annealing); 70°C/50s (extension), followed by a final extension of 70°C for 5-8 minutes. X depended on the melting temperature of the primers employed, and both the length and temperature for annealing and extending varied depending on the individual experiment. In general, shorter annealing times were employed for greater specificity, and longer extension times for longer inserts (approximately 60 seconds added per kilobase of DNA.)

For PCR amplification from genomic templates, the cycling conditions were the same as described above, except that the variations favoured higher specificity by using higher annealing temperatures and shorter annealing times.

For Vent amplification, the conditions were as described above, except the extension temperature was raised to 75°C, and the extension time was lengthened to accommodate the enzyme's proof-reading capacity.

DNA Sequencing

Templates which required sequencing were almost always cloned first, and purified as described above (Plasmid DNA preparation). All samples were prepared for automated sequencing as follows: 100ng of template per kilobase of DNA (including vector), primer added to an approximate final concentration of 300 picomoles, and deionized distilled water to a final volume of 12ul. Samples were then sent to the University of Calgary Core DNA and Protein services (www.ucalgary.ca/~dnalab) for sequencing.

Sequence analysis software and Bioinformatics tools

Sequence assembly (*D.virilis genomic subclones)* and integrity checks (germline transformation constructs) were carried out using the BLAST algorithm for two sequences (www.ncbi.nlm.nih.gov/BLAST/). Identifying *D.virilis* coding and non-coding regions by homology to sequences from those model organisms that have already been

sequenced was carried out using BLASTN and TBLASTN. All the default BLAST settings were used, except that low complexity sequence was not masked.
Multiple protein or nucleic acid alignments were made first by using CLUSTALW (Thompson et al. 1994) to generate the alignments, then BOXSHADE (Boxshade version 3.3.1, by Kay Hofmann and Michael D. Baron) for colouring conserved regions.
Examinations for conserved non-coding DNA motifs were carried out using DiALIGN (http://bibiserv.techfak.uni-bielefeld.de/dialign/submission.html: Morgenstern,1999) and TFSITESCAN (www.ifti.org/cgi-bin/ifti/Tfsitescan.pl. Ghosh, 2000). Putative promoter consensus sequences were confirmed using (BDGP neural network based promoter prediction software http://www.fruitfly.org/). AT:GC content was measured using NASTATS (http://workbench.sdsc.edu/.). Primer sequences were designed using Oligo 4.1 Primer analysis software (National Biosciences Inc.). Restriction enzyme analysis was performed with TACG (http://biotools.umassmed.edu/bioapps/rsites.html).
Non-coding DNA sequences used in seaches for transcription factor binding sites, or conservation between species, were randomized using RANDSEQ software (Pearson 1990).

CLUSTALW, BOXSHADE, NATSTATS, RANDSEQ are all available through http://workbench.sdsc.edu/.

CHAPTER THREE: Identification and characterization of two *Drosophila melanogaster* genes in proximal 3L heterochromatin

INTRODUCTION

There are three lethal complementation groups that have been identified in proximal 3L heterochromatin of *Drosophila melanogaster* (Marchant and Holm 1988b). *lethal 1* is closest to the centromere, and behaves genetically as a *Suppressor of Polycomb* (Kennison and Tamkun 1988), placing it in the *trithorax* Group of transcriptional regulators (please see Chapter One for a description of this group). *lethal 2* alleles exhibit a complex pattern of intra-allelic complementation, with escapers surviving to adulthood and displaying a classical *Minute* phenotype characteristic of defects in the regulation of protein synthesis. *lethal 3* is the most distal of this group, and like *lethal 1* appears to behave like a *trithorax* group gene, in this case regulating important developmental genes like *hedgehog,* since mutant alleles will suppress the cis-dominant mutation of *hedgehog* called *Moonrat* (Schulze at al. 2001). In keeping with the nomenclature for *trx-G* genes taken from the mythology of various cultures depicting fate, *lethal 3* has been given the name *verthandi* (Norse: goddess of Fate).

One goal of our laboratory has been to contribute to a complete molecular genetic map of *Drosophila* heterochromatin, by cloning and characterizing lethal complementation groups previously identified in the heterochromatin of the third chromosome (Marchant and Holm 1988a,b). Heterochromatin remains largely uncharted, and has presented problems both for bioinformaticists and biologists (Hoskins et al. 2002).

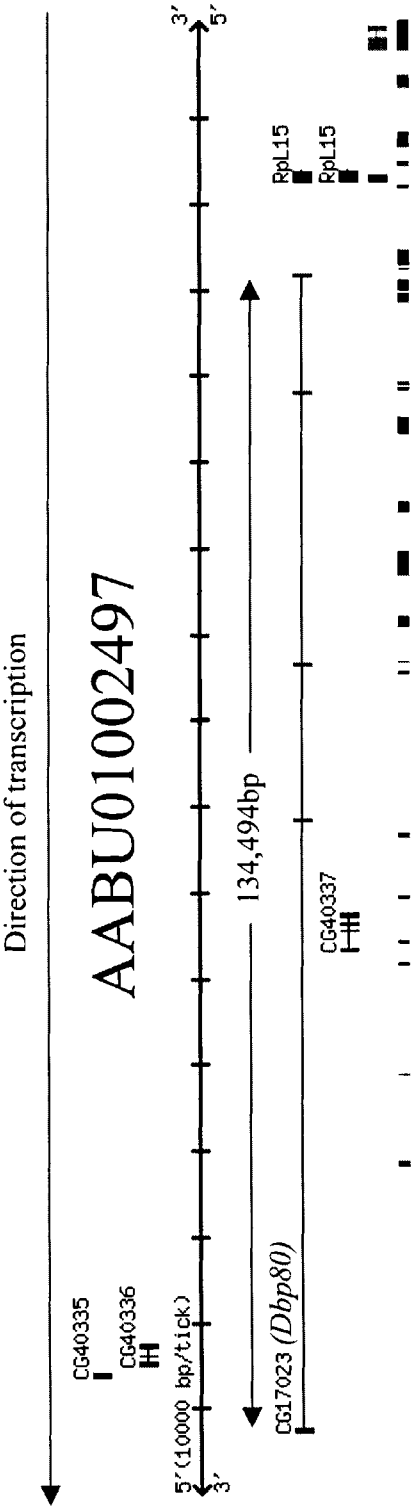
I chose to focus on two proximal genes, *lethal 2* and *lethal 3,* initially by carrying out a thorough genetic analysis, and then by applying molecular methods to establish their identities. Both of these genes have been tagged by natural albeit internally deleted P elements from the *Birmingham 2* strain (see Chapter One). Transposon tagging has proved to be a powerful way to clone genes, by using molecular methods to retrieve the element and flanking DNA. Unfortunately these natural P elements have no engineered sequences (plasmid rescue, markers for inverse PCR etc) to facilitate this process. In addition there is evidence which suggests that tranposons can insert a great distance from the genes they affect (Robert et al. 2001). Also, transposable elements will insert into other transposable element sequences, which comprise a significant proportion of

Drosophila heterochromatin, including sequences in and around genes. When this happens, only repetitive flank is rescued, and cannot therefore be used to identify the gene. Since rescuing flank from the natural P elements in *lethal 2* and *lethal 3* was likely to be problematic, a genomics approach was employed in parallel. This involved selecting potential candidate (predicted) genes from the Berkeley *Drosophila* Genome Project database (www.fruitfly.org/). The Drosophila genome sequence has undergone three releases since March 2000, and in each release the sequence has been organized into scaffolds which could be physically mapped to the euchromatic chromosome arms, and scaffolds which could not be localized, and so reside in a separate database called armU (for Unlocalized). Currently these armU scaffolds are in the process of being more fully mapped. A number of ESTs (expressed sequence tags) map to these armU scaffolds, and thus provide good candidates for *lethal 2* and *3*.

There were (and are) no obvious candidates for *lethal 3*, but two possibilities were considered for *lethal 2* based on the products they encode. *Dbp80* belongs to a large family of DExH box helicases which are involved in almost every stage of RNA metabolism, and *RpL15* encodes a large subunit ribosomal protein. RNA helicases play an important role in dosage compensation, the mechanism by which a male fruit fly upregulates the transcriptional activity of his single X chromosome, in order to balance the expression from the female X chromosome pair. Defects in this pathway are characterized by sex-specific lethality – a phenotype also exhibited by certain transheterozygous combinations of *lethal 2* alleles. Ribosomal protein genes comprise a large group, of which many are essential. Mutations in some ribosomal protein genes lead to a syndrome of defects called a *Minute* phenotype, (fine bristles, rough eyes, wing vein defects and reduced or gapped sex combs), which, as has already been mentioned, *lethal 2* transheterozygote escapers display. Both genes share a scaffold (Figure 3.1), which suggests they reside in the same vicinity on the chromosome. *Dbp80* is a huge gene, over 140 kb, while *RpL15* is rather small for a heterochromatic gene, fully contained within a genomic fragment of less than 2 kb.

Figure 3.1: Schematic of BDGP scaffold AABU01002497, containing *Dbp80* and *RpL15*

A number of genetic elements are present in this region, but the two relevant genes are *Dbp80* (~140kb long) and *RpL15* (less than 2kb long). These two genes were considered as potential candidates for *lethal 2* since they encoded products which have been implicated in processes which when defective produce phenotypes characteristic of *lethal 2* heteroallelic combinations: sex-specific lethality (DEAD-box helicases) and *Minutes* (ribosomal proteins). The other genetic elements shown in this picture appeared only in the most recent BDGP release, and may be transposable elements mis-called as genes. Note also the figure is shown 3' to 5': in reality, *RpL15* is upstream of *Dbp80*; both genes are transcribed in the same direction. Figure courtesy of Flybase.


Direction of transcription
AABU01002497
CG40335
CG40336
5'(10000 bp/tick)
3'
3'
5'
134,494bp
CG17023 (Dbp80)
CG40337
RpL15
RpL15
KEY:
Gene
cDNA(dark) EST(lt)
Repeat

Both *Dbp80* and *RpL15* map to the same genetic region as *lethal 2* in 3L heterochromatin, but only *RpL15* corresponds to the *lethal 2* complementation group. *Dbp80* corresponds to neither *l1* nor *l3* and therefore may not be essential.

It was of interest to discover if these two contrasting genes shared any of the unusual properties characteristic of heterochromatic gene function. For instance, recent studies have suggested that heterochromatic genes require Heterochromatin Protein 1 for their expression (Lu et al. 2000, and please see Chapter One for a fuller description of chromatin structure and gene expression). HP1 encodes a protein that recognizes and binds histone H3 that has been methylated on lysine 9 (Bannister et al. 2001, Lachner et al. 2001). Since it also binds to itself, it serves as a crosslinking protein, which can further compact the chromatin fibre, rendering it less accessible to the transcriptional machinery. As such, it is considered a transcriptional repressor, which has been shown to silence transgene arrays (Dorer and Henikoff, 1994). But heterochromatic genes thrive in an otherwise repressive environment, and both the heterochromatic genes *light* and *rolled* have exhibited a dependence upon HP1 for their proper expression (Lu et al. 2000). Therefore, the levels of *Dbp80* and *RpL15* expression were measured in a genetic background in which HP1 dose had been reduced, and these results were compared with those described for *light* and *rolled.*

In sum, the results of the present analysis indicate that *lethal 2* likely encodes *RpL15*, and *Dbp80* corresponds to neither *lethal 1* nor *lethal 3.* This would suggest that in some places at least, the density of genes deep within heterochromatin can resemble that of euchromatin (one gene per 9kb: Adams et al. 2000. Also, Tulin et al. 2002, Hoskins et al. 2002), as can the ratio of essential to non essential genes. In addition, both genes exhibit compromised expression in a genetic background in which HP1 expression is reduced, in keeping with the results observed for the heterochromatic genes *light* and *rolled* (Lu et al. 2000). Finally, attempts were made to rescue lesions in *lethal 2* with germline transformation constructs containing cDNAs or genomic inserts for *RpL15*. These rescue experiments have been partially successful, and underscore the extreme dose sensitivity of this gene.

RESULTS

3.1. *lethal 3* analysis

lethal 3 (verthandi) is the farthest from the centromere of the proximal group (Figure 1.3). It was originally identified by an EMS allele called *1-166-38*, and three P alleles, *5-1, 3-1* and *5-3* (Marchant and Holm 1988b, Schulze et al. 2001), of which the latter two were revertible. A new EMS screen was performed to generate more alleles. Complementation and lethal phase analysis were carried out, and collaboration with Jim Kennison (Schulze et al. 2001) established a biological function for *lethal 3* as a member of the *trithorax* group of transcriptional activators. Attempts were made to clone the flank from the revertible P alleles, which were not successful. There are no potential candidates yet evident from the BDGP.

3.1.1: Genetic analysis of lethal 3

Table 3.1a lists the names and types of *lethal 3* alleles generated both previously and during the course of the present work. The new alleles, *35, 36, γ26-5* and *γ26-6* were all generated in an EMS screen using *Df(3L)γ26* and *Df(3L)K2*. This screen is outlined in Chapter Two. It was carried out in order to generate more alleles for proximal 3L heterochromatic genes, and also to attempt to saturate the region with respect to the induction of lethal mutations. Table 3.1b presents a complementation table for pair wise combinations of all *lethal 3* alleles. They fail to complement each other completely, and exhibit an embryonic lethal phase, with no cuticle defects. From Jim Kennison's work (Schulze et al. 2001) it was demonstrated that mutations in *lethal 3* suppress a cis-dominant mutant allele of *hedgehog* called *Moonrat*. This suggests that the wild type function of *lethal 3* is to positively regulate developmental gene expression, placing it into the *trithorax*-Group of transcriptional activators. As such, it has been called *verthandi*, in keeping with the nomenclature of these genes in *Drosophila* (see above, and Kennison and Tamkun 1988).

lethal 3 allele	Mutagen	Reference
3-1	Natural P element	Schulze et.al.
5-1	Natural P element	Schulze et.al.
5-3	Natural P element	Schulze et.al.
1-166-38	EMS	Marchant & Holm
35	EMS	Schulze et.al.
36	EMS	Schulze et.al.
γ26-5	EMS	Schulze et.al.
γ26-6	EMS	Schulze et.al.

Table 3.1(a): List and sources of *lethal 3* alleles.

lethal 3	*1-166-38*	*3-1*	*5-1*	*5-3*	*35*	*36*	*γ26(5)*	*γ26(6)*	*vtd4*
1-166-38		0/598	0/310	0/483	0/97	0/206	0/248	0/231	0/307
3-1			0/448	0/445	0/121	0/277	0/319	0/62	0/429
5-1				0/185	0/94	0/103	0/257	0/59	0/373
5-3					0/168	0/397	0/114	0/67	0/265
35						0/393	0/175	0/271	0/188
36							0/112	0/212	0/327
γ26(5)								0/168	0/171
γ26(6)									0/125

Table 3.1(b): Complementation data for pair wise combinations of *lethal 3* alleles.

3.1.2: Molecular analysis of lethal 3

Three P alleles of *lethal 3* were generated by the screen outlined in Chapter One, using the *Birmingham 2* strain as a source of P elements (Robertson et al.1988). This strain has 17 natural but internally deleted P elements on the second chromosome, which can be mobilized in the presence of a transposase source. The advantage of this kind of screen is that for every experiment, there is a large number of mutagenic "bullets" (in the form of P elements). The disadvantage of this strain is depicted in Figure 3.2: the background contains a large number of P elements which do not tag the gene of interest. For this kind of screen, new mutant strains must be "cleaned up", which means first replacing all the background chromosomes, and then crossing off the euchromatic arms of the third chromosome (see Chapter Two).

A variety of molecular methods were employed to try and clone this flank. The first step involved cutting *3-1* genomic DNA with a restriction enzyme and purifying the DNA from the band corresponding to the P element signal in the Southern. This DNA was then used in inverse PCR experiments and also to generate a plasmid library. Neither case was successful, and at the same time, there were no candidate ESTs available from the BDGP. As a consequence I chose to direct my efforts towards the characterization of *lethal 2*.

3.2. *lethal 2* analysis

lethal 2 is the second closest lethal complementation group to the centromere in the proximal group (Figure 1.3). It was originally identified by a single EMS allele, *1-166-37* (Marchant and Holm 1988b), and four P alleles: *PΔ2, PΔ8, 8-1*, and *7-1* (Schulze et al. 2001), all of which can be reverted. A new allele, *72*, was isolated in the EMS screen described above and in Chapter Two, and all pair wise combinations were tested for complementation. Initial attempts to clone the flank from the P alleles were unsuccessful, however there were two promising candidates from the BDGP. These were selected based on a correlation between the products they encode, and the type of mutant

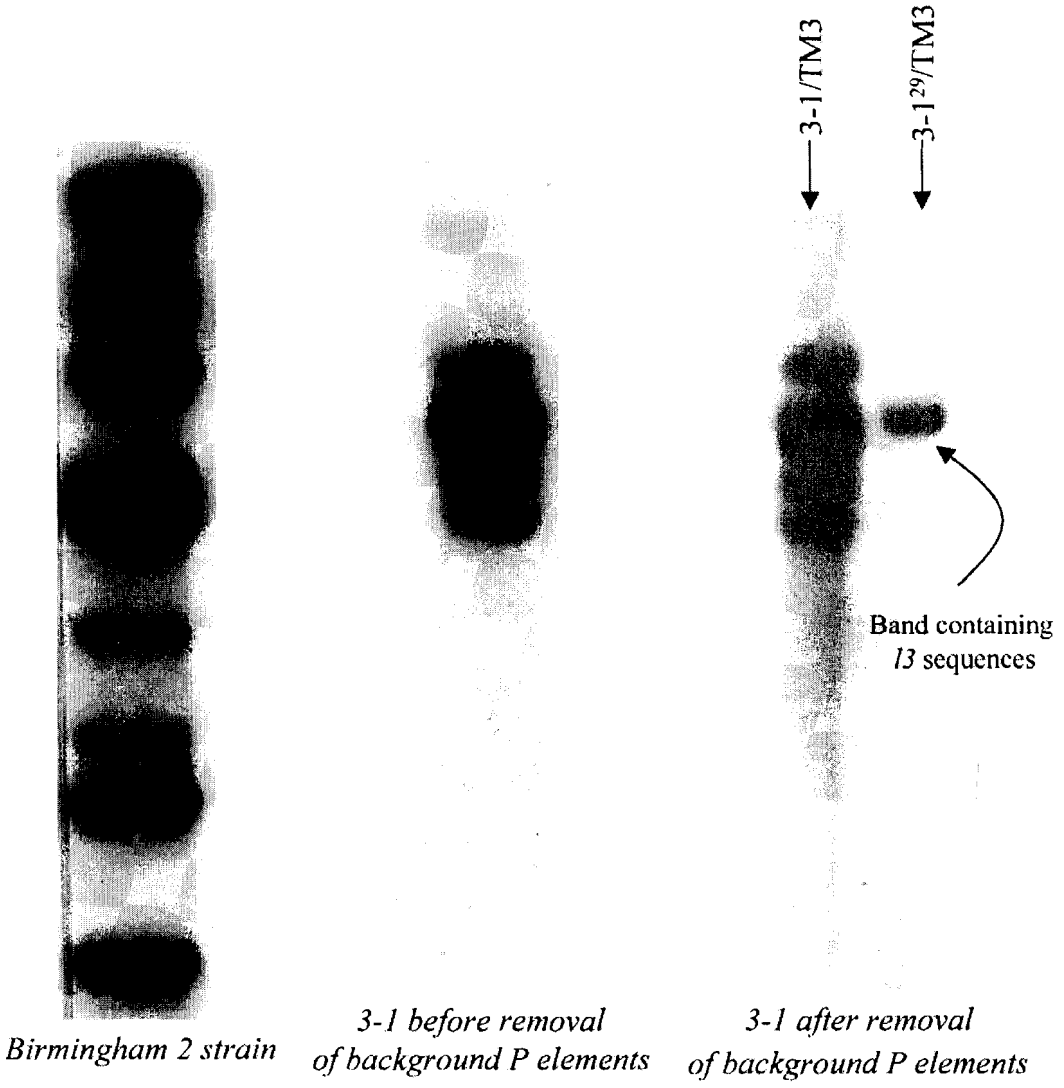


Figure 3.2: *lethal 3* P mutant Southerns
Genomic DNA cut with EcoRI and probed with the first third of the P element sequence. The *Birmingham 2* strain, used as a mutator, is shown on the left. The middle panel shows the *lethal 3* mutant strain *3-1* before background P elements were removed, and the panel on the right shows a comparison of *3-1* before and after the removal of background elements.

phenotypes exhibited by various heteroallelic combinations of *lethal 2* mutants. Their characterization suggests that *lethal 2* encodes the ribosomal protein RpL15. Both genomic and cDNA transgene constructs were made, and a number of independent lines of transgenic animals were established. The effect of reducing HP1 dose on both *lethal 2* candidates was also assayed. Finally total RNA from *lethal 2* mutants was used in a quantitative Northern analysis to measure the effect on *RpL15* transcription.

3.2.1: Genetic analysis of lethal 2

Table 3.2a lists the names and types of *lethal 2* alleles generated both previously and during the course of the present work. Table 3.2b presents a complementation table for pair wise combinations of all *lethal 2* alleles. The pattern of complementation for mutations in this gene is complex, with specific transheterozygotes displaying appreciable levels of survival. Under certain conditions (growth at 22°C, large numbers of progeny) some allelic combinations exhibit a pronounced sex skew, in one case leading to sex-specific lethality (Table 3.3). This kind of phenotype has been implicated as diagnostic of defects in dosage compensation (Lucchesi 1998). Homozygotes for EMS alleles or a deficiency for the region (*Df(3L)9-56*) die early, predominantly during the L1 phase. Certain P transheterozygotes produce escapers which eclose as adults that die within 6 days at room temperature (*PΔ2/8-1, 7-1, 1-166-37 or 72*). Other combinations (*PΔ2/ PΔ8, 8-1 or 7-1/ PΔ8)* produce sterile adults that live out a normal lifespan at room temperature. All transheterozygote escapers display a classic Minute phenotype typical of defects that lead to global reductions in protein synthesis (Figure 3.3). In addition, *Df(3L)9-56/TM3* and *Df(3L)K2/TM3* individuals exhibit a very subtle Minute phenotype suggesting that *lethal 2* is weakly haploinsufficient (data not shown, D.Sinclair personal communication).

As can be seen from Figure 3.4b, *PΔ8* and *7-1* appear to have other P elements in the background, while *PΔ2* and *8-1* do not. In addition, *7-1* was initially identified as a separate isolate of the same event that produced *8-1*. The background chromosomes were replaced (see Chapter Two for details) and the Southern analysis repeated (Figure 3.4b).

lethal 2 allele	Mutagen	Type	Reference
PΔ2	Natural P element	Hypomorph	Schulze et.al.
PΔ8	Natural P element	Hypomorph	Schulze et.al.
7-1	Natural P element	Hypomorph	Schulze et.al.
8-1	Natural P element	Hypomorph	Schulze et.al.
1-166-37	EMS	Hypomorph	Marchant & Holm
72	EMS	null	Schulze et.al.

Table 3.2(a): List and sources of *lethal 2* alleles.

lethal 2	*1-166-37*	*7-1*	*8-1*	*PΔ2*	*PΔ8*	*72*
1-166-37		0/600	0/501	0/386	0/309	0/290
7-1			1/659	1/297	86/1026	0/524
8-1				10/764	87/848	0/774
PΔ2					134/863	1/680
PΔ8						0/828

Table 3.2(b): Complementation data for pair wise combinations of *lethal 2* alleles.

GENOTYPE	RV	SR	n	n^h
PΔ2 / PΔ8	0.82	0.92	2080	567
7-1 / PΔ8	0.24	0.77	1323	105
7-1 / PΔ2	0.12	2.0	1551	62
8-1 / PΔ8	0.40	1.6	954	148
8-1 / PΔ2	0.29	0.97	661	63
1-166-37 / PΔ8	0.14	All males	1296	61

Table 3.3: Sex ratio tests for *lethal 2* transheterozygotes
RV: relative viability (=observed frequency/expected frequency); SR: sex ratio (#males:#females); n = # progeny; n^h = # transheterozygotes or homozygotes. *Note*: dramatic decrease in viability correlates with sex ratio shift.

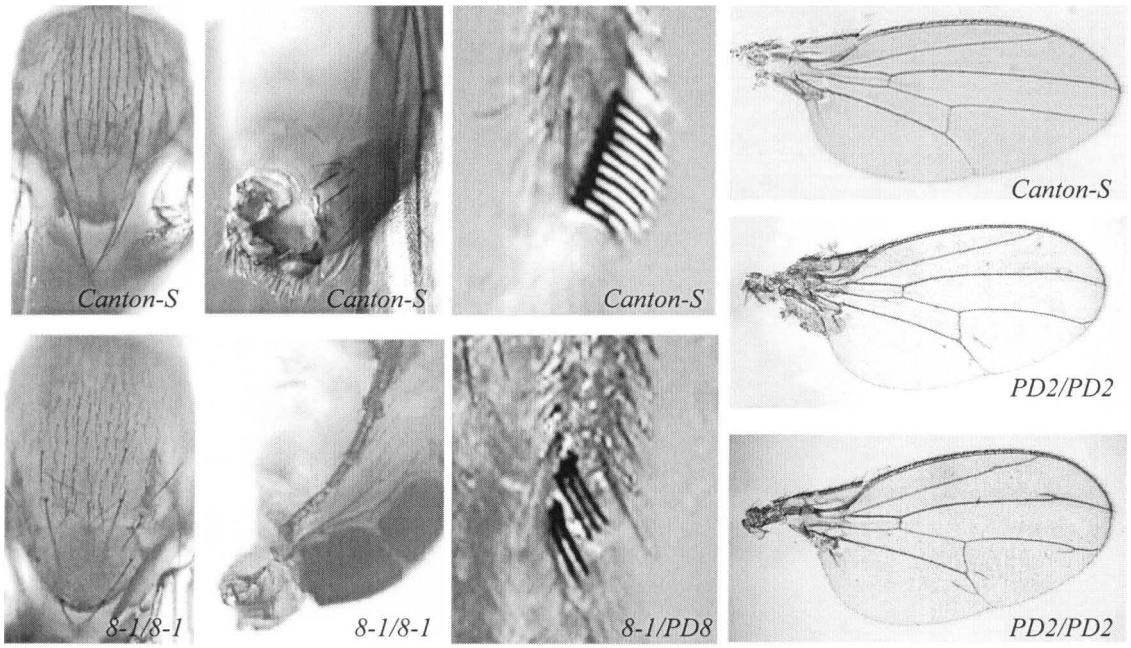
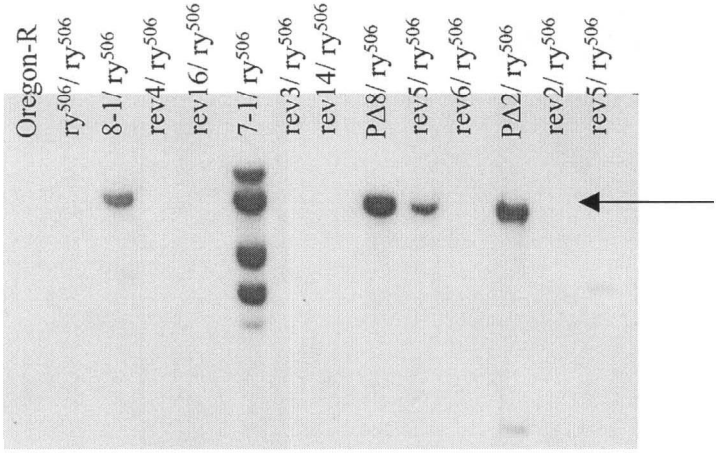
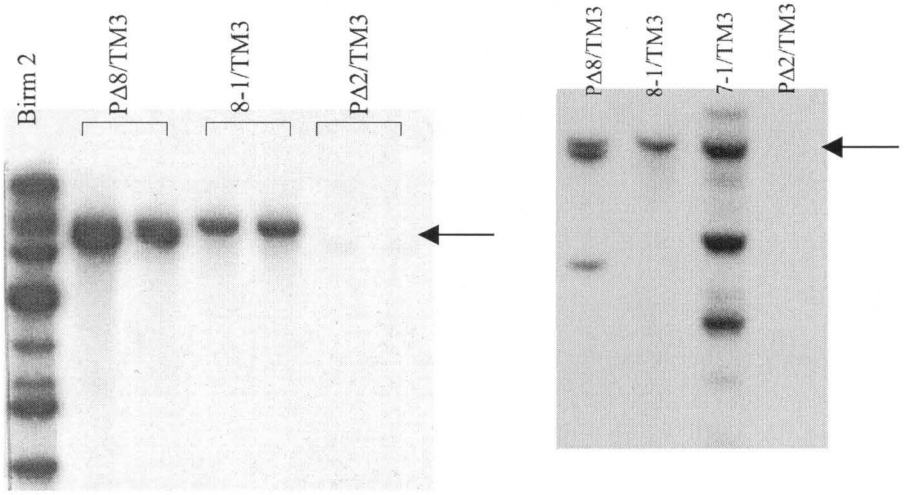


Figure 3.3: *lethal 2* phenotypes
Wild-type phenotypes (top) and *lethal* 2 transheterozygote *Minute* phenotypes (bottom): from left to right: fine bristles, mis-rotated genitalia, missing/reduced sex combs and wing vein defects. This syndrome of growth defects is associated with mutations that lead to global reductions in the levels of protein synthesis.



(a)



(b)

Figure 3.4 (a) and (b): *lethal 2* P mutant Southerns

(a): *lethal 2* P allele DNA (BEFORE removal of background chromosomes) cut with EcoRI and probed with the first third of the P element. Arrows point to the band corresponding to *RpL15* coding sequences.

(b): *lethal 2* P allele DNA (AFTER removal of background chromosomes) cut with EcoRI and probed with the first third of the P element (left) and internal P element sequences (right). Arrows point to the band corresponding to *RpL15* coding sequences.

During the course of the background chromosome replacement procedure, *PΔ2* appears to have undergone a spontaneous internal deletion event, which may have caused it to behave as the weakest allele in the series. Meanwhile, *PΔ8* and *7-1* remain unchanged, suggesting that other P elements reside on the third chromosome. Therefore, it is likely that these stocks came into our lab in different stages of the "cleaning up" process described earlier.

A number of lethal excisions in *lethal 2* were also generated, by taking advantage of the fact that certain combinations of P alleles will produce survivors. The crossing scheme to ensure independent events is outlined in Figure 3.5, and the data are shown in Table 3.4. These alleles are all lethal excisions; however, a subset produced sterile heteroallelic escapers in the F3 generation, showing a Minute phenotype. The stocks of these excisions produce no (homozygous) escapers, and are viable over mutations in flanking genes (*lethal 1* and *lethal 3*). The main purpose for generating these alleles was to aid the molecular analysis.

3.2.2: Molecular analysis of lethal 2

Initial attempts to isolate flanking DNA from *lethal 2* P alleles were unsuccessful (as they were for *lethal 3*). Fortunately, there were two promising candidate ESTs available from the BDGP: *Dbp80* which encodes an RNA helicase, and *RpL15*, which encodes a ribosomal protein.

***Dbp80* is a large heterochromatic gene which maps to 3L heterochromatin**

Dbp80 (*DEAD box protein 80*) was initially described in work published by Ari Eisen and John Lucchesi who study dosage compensation in flies (Eisen et al.1998). It was mapped by polytene *in situ* to 3L heterochromatin, and encodes an RNA helicase of the DExH box family, linked to a highly conserved protein in yeast (Dbp5) that is an essential component of the mRNA export pathway (Snay-Hodge et al. 1998) (Figure 3.6, Table 3.5). A BLASTP search against *Drosophila* predicted protein database yields 29

other DExH box helicase genes (Figure 3.7), many of which have been genetically characterized (http://flybase.bio.indiana.edu:82/), but *Dbp80* is the only one that possesses a unique six residue insertion that places it in the same family as Dbp5 (Figure 3.8).

Dbp80 was isolated from a yeast-two-hybrid screen using *Maleless* as bait (Eisen et al. 1998). *Maleless* encodes another RNA helicase that has a defined role in the RNA-protein complex which assembles at numerous sites on the male X chromosome during the process of dosage compensation (Lucchesi 1998). Other members of this complex include proteins with chromodomains and histone acetyltransferase activity, both components of chromatin remodeling complexes. The presence of an RNA helicase in the same complex foreshadowed the discovery of the role non-coding RNAs play in this pathway, and lately this has become an important feature of chromatin remodeling in general (Volpe et al. 2002). Sex-specific lethality is a diagnostic phenotype for defects in this process, and since dosage compensation takes place in males, the lethality is restricted to this sex. However, certain transheterozygote combinations of *lethal 2* alleles cause female lethality (Table 3.3), which suggested this might have represented a novel part of the dosage compensation mechanism.

Dbp80 gene organization

The *Dbp80* coding region initially mapped to three small genomic scaffolds which are listed as unlocalized (armU). In the latest heterochromatic release (Hoskins et al. 2002) these three scaffolds have been joined together to yield the map in Figure 3.9. The *Dbp80* cDNA is 1508 bp long, but the exons are spread over a very large genomic distance, which exceeds 140kb. The exons are split by a number of enormous introns, which apparently cannot be fully sequenced, probably due to repetitive DNA. The final 3' exons of this gene are still absent from the genome sequence database.

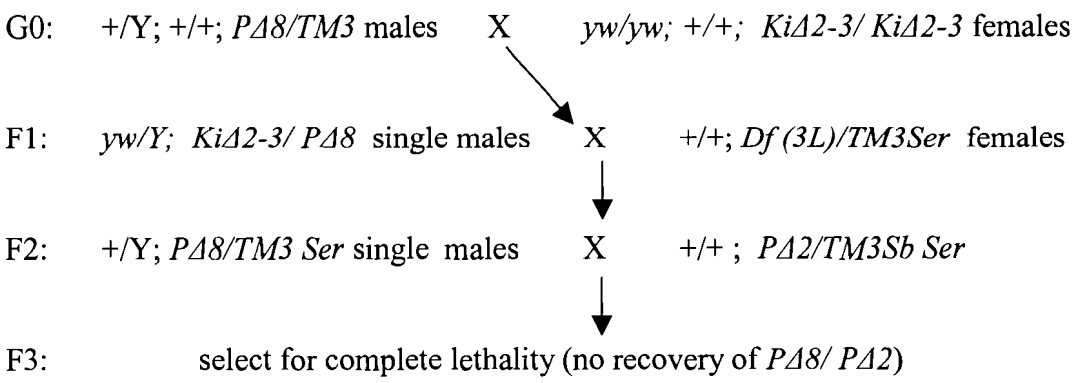


Figure 3.5: *lethal 2* excision experiment

Crossing scheme to ensure independent excision events from *lethal 2* P alleles. The same scheme was carried out starting with *8-1*, and scoring for failure to recover *8-1/PΔ8* transheterozygotes.

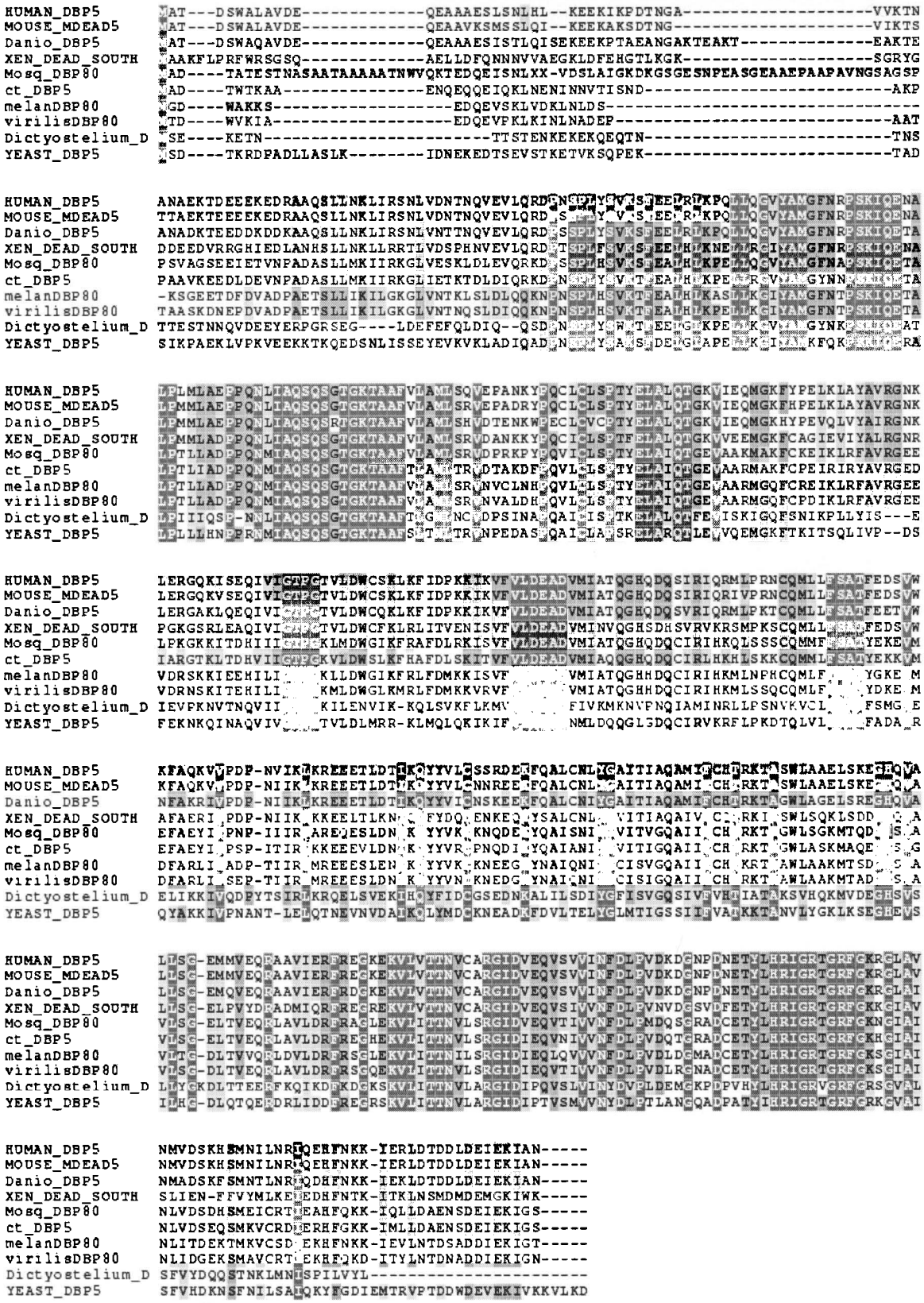
Parent Chromosome	# F1 males	#F2 progeny males	# sterile	#revertants[1]	#confirmed excisions[2]	#Independent events[3]
PΔ8/TM3	13	123	3	111	9	6
8-1/TM3	11	115	3	93	18	10

Table 3.4: *lethal 2* excision data

[1]Defined as "revertant" if F2 males produced *any* survivors in the F3 generation. Note that detailed analysis was not performed on every "revertant" cross, (since the objective was to find lethals) and a number of these were undoubtedly also excisions, which produced a mild *Minute* phenotype.
[2]Confirmed by Southern analysis, or, when mutant transheterozygotes were produced, testing for fertility.
[3]This number is minimal; Southern analysis shows that the same F1 male can produce separate excision events.

Figure 3.6: CLUSTALW alignment of DBP80 homologues across different taxa

The most highly conserved domains are indicated by white type on a gray background.


HUMAN_DBP5
MOUSE_MDEAD5
Danio_DBP5
XEN_DEAD_SOUTH
Mosq_DBP80
ct_DBP5
melanDBP80
virilisDBP80
Dictyostelium_D
YEAST_DBP5

Organism	%similarity	%identity
Anopheles gambiae	81%	68%
Chironomus tentans	81%	67%
Mus musculus	77%	59%
Homo sapiens	77%	58%
Danio rerio	76%	58%
Xenopus laevis	73%	55%
Dictyostelium discoideum	67%	47%
Schizosaccharomyces pombe	66%	46%
Saccharomyces cerevisiae	64%	44%
Neurospora crassa	62%	41%

Table 3.5: DBP80 protein homologies across different taxa

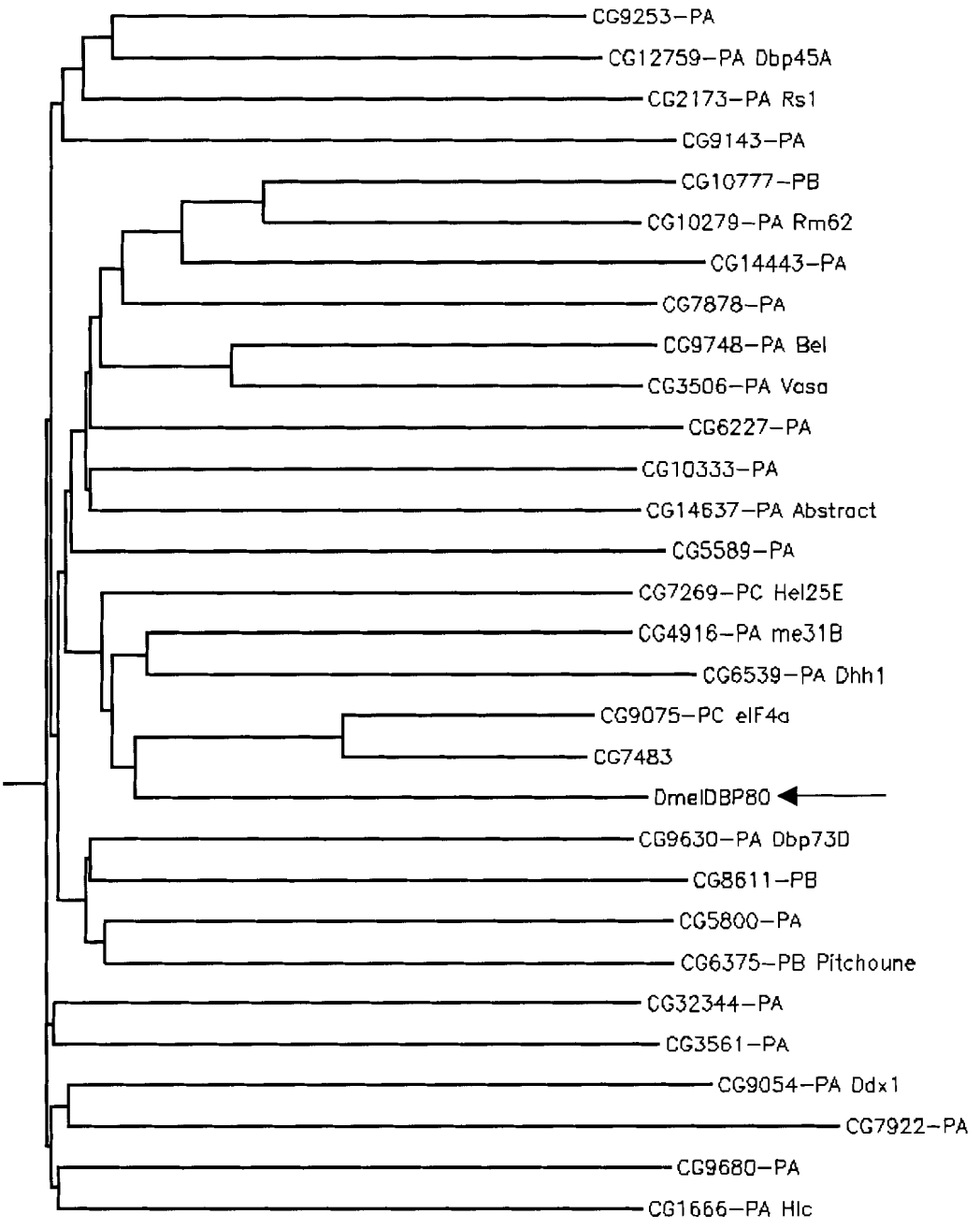
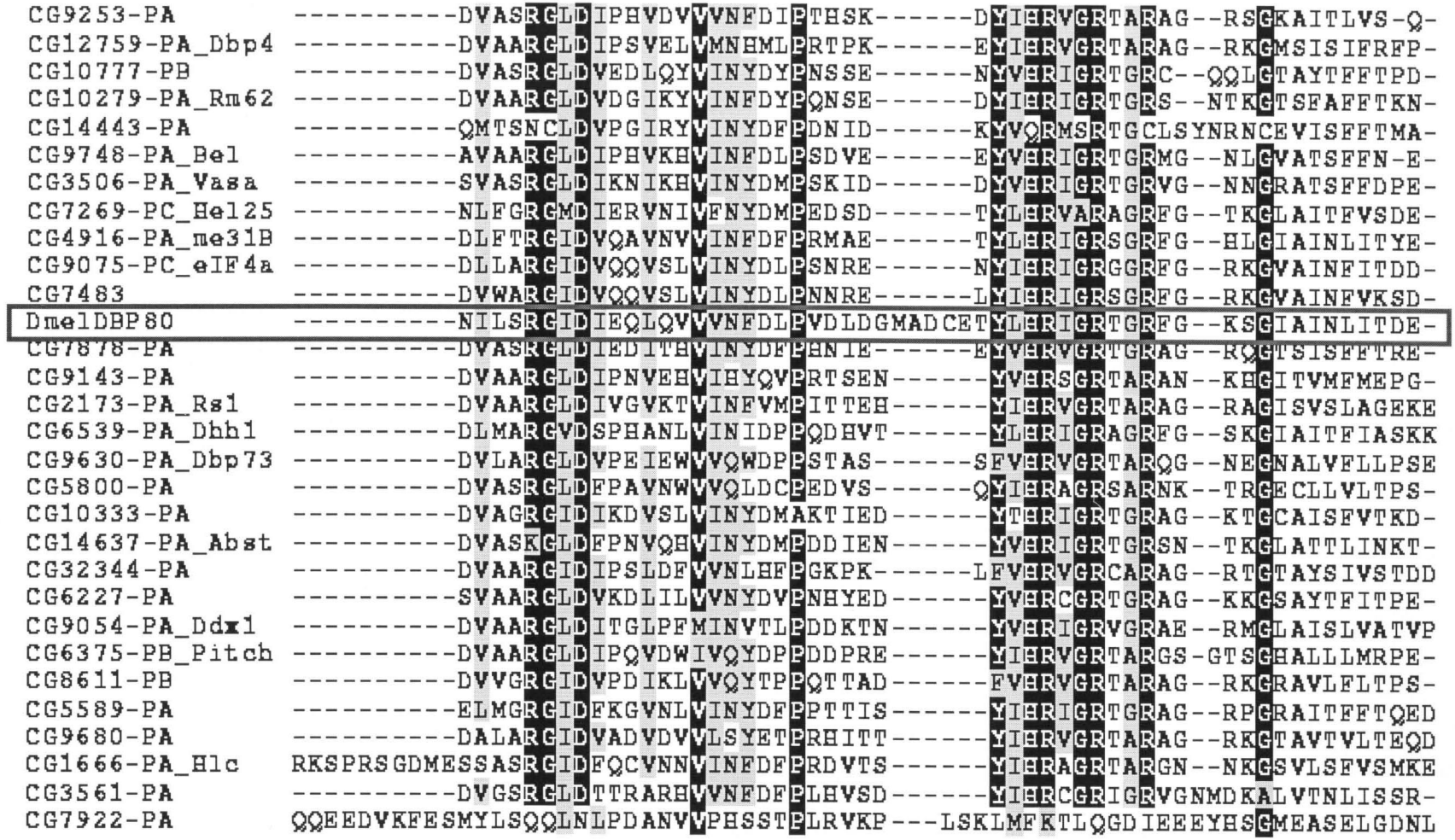


Figure 3.7: DBP80 and related proteins in the *Drosophila melanogaster* genome

Tree drawn from a CLUSTALW alignment (see Chapter 2).

Figure 3.8: DBP5 family conserved domain
Alignment of a portion of DBP80 with related proteins from the *Drosophila melanogaster* genome showing the six amino acid residue insert characteristic of the DBP5 family (in *Drosophila* this amino acid sequence spells MADCET).


CG9253-PA ----------DVASRGLDIPHVDVVVNFDIPTHSK------DYIHRVGRTARAG--RSGKAITLVS-Q-
CG12759-PA_Dbp4 ----------DVAARGLDIPSVELVMNHMLPRTPK------EYIHRVGRTARAG--RKGMSISIFRFP-
CG10777-PB ----------DVASRGLDVEDLQYVINYDYPNSSE------NYVHRIGRTGRC--QQLGTAYTFFTPD-
CG10279-PA_Rm62 ----------DVAARGLDVDGIKYVINFDYPQNSE------DYIHRIGRTGRS--NTKGTSFAFFTKN-
CG14443-PA ----------QMTSNCLDVPGIRYVINYDFPDNID------KYVQRMSRTGCLSYNRNCEVISFFTMA-
CG9748-PA_Bel ----------AVAARGLDIPHVKHVINFDLPSDVE------EYVHRIGRTGRMG--NLGVATSFFN-E-
CG3506-PA_Vasa ----------SVASRGLDIKNIKHVINYDMPSKID------DYVHRIGRTGRVG--NNGRATSFFDPE-
CG7269-PC_Hel25 ----------NLFGRGMDIERVNIVFNYDMPEDSD------TYLHRVARAGRFG--TKGLAITFVSDE-
CG4916-PA_me31B ----------DLFTRGIDVQAVNVVINFDFPRMAE------TYLHRIGRSGRFG--HLGIAINLITYE-
CG9075-PC_eIF4a ----------DLLARGIDVQQVSLVINYDLPSNRE------NYIHRIGRGGRFG--RKGVAINFITDD-
CG7483 ----------DVWARGIDVQQVSLVINYDLPNNRE------LYIHRIGRSGRFG--RKGVAINFVKSD-
DmelDBP80 ----------NILSRGIDIEQLQVVVNFDLPVDLDGMADCETYLHRIGRTGRFG--KSGIAINLITDE-
CG7878-PA ----------DVASRGLDIEDITHVINYDFPHNIE------EYVHRVGRTGRAG--RQGTSISFFTRE-
CG9143-PA ----------DVAARGLDIPNVEHVIHYQVPRTSEN------YVHRSGRTARAN--KHGITVMFMEPG-
CG2173-PA_Rs1 ----------DVAARGLDIVGVKTVINFVMPITTEH------YIHRVGRTARAG--RAGISVSLAGEKE
CG6539-PA_Dhh1 ----------DLMARGVDSPHANLVINIDPPQDHVT------YLHRIGRAGRFG--SKGIAITFIASKK
CG9630-PA_Dbp73 ----------DVLARGLDVPEIEWVVQWDPPSTAS------SFVHRVGRTARQG--NEGNALVFLLPSE
CG5800-PA ----------DVASRGLDFPAVNWVVQLDCPEDVS------QYIHRAGRSARNK--TRGECLLVLTPS-
CG10333-PA ----------DVAGRGIDIKDVSLVINYDMAKTIED------YTHRIGRTGRAG--KTGCAISFVTKD-
CG14637-PA_Abst ----------DVASKGLDFPNVQHVINYDMPDDIEN------YVHRIGRTGRSN--TKGLATTLINKT-
CG32344-PA ----------DVAARGIDIPSLDFVVNLHFPGKPK------LFVHRVGRCARAG--RTGTAYSIVSTDD
CG6227-PA ----------SVAARGLDVKDLILVVNYDVPNHYED------YVHRCGRTGRAG--KKGSAYTFITPE-
CG9054-PA_Ddx1 ----------DVAARGLDITGLPFMINVTLPDDKTN------YVHRIGRVGRAE--RMGLAISLVATVP
CG6375-PB_Pitch ----------DVAARGLDIPQVDWIVQYDPPDDPRE------YIHRVGRTARGS-GTSGHALLLMRPE-
CG8611-PB ----------DVVGRGIDVPDIKLVVQYTPPQTTAD------FVHRVGRTARAG--RKGRAVLFLTPS-
CG5589-PA ----------ELMGRGIDFKGVNLVINYDFPPTTIS------YIHRIGRTGRAG--RPGRAITFFTQED
CG9680-PA ----------DALARGIDVADVDVVLSYETPRHITT------YIHRVGRTARAG--RKGTAVTVLTEQD
CG1666-PA_Hlc RKSPRSGDMESSASRGIDFQCVNNVINFDFPRDVTS------YIHRAGRTARGN--NKGSVLSFVSMKE
CG3561-PA ----------DVGSRGLDTTRARHVVNFDFPLHVSD------YIHRCGRIGRVGNMDKALVTNLISSR-
CG7922-PA QQEEDVKFESMYLSQQLNLPDANVVPHSSTPLRVKP---LSKLMFKTLQGDIEEEYHSGMEASELGDNL

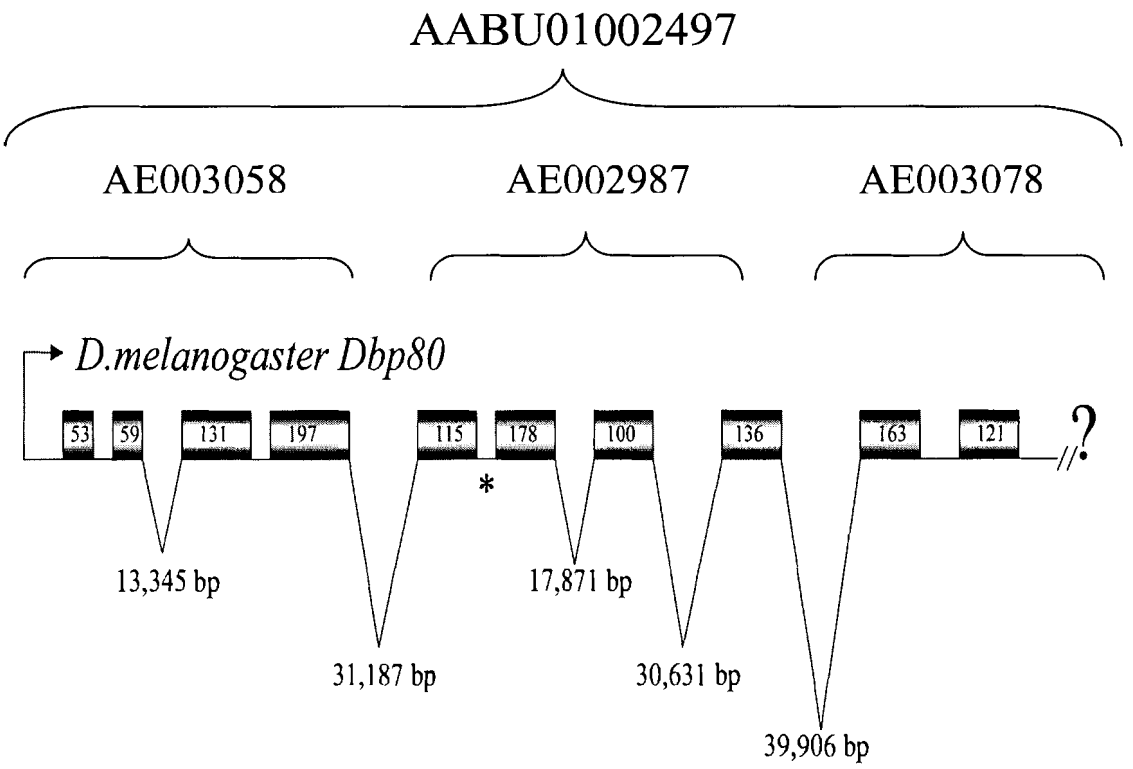


Figure 3.9: *Dbp80* gene organization
Overlay of Release 2 (AE numbers) and Release 3 (AABU number) BDGP sequenced scaffolds containing the DBP80 gene. Gray boxes indicate exons, lines connecting them represent the introns. The figure is NOT to scale. The asterisk (*) marks the position of an ancient intron, shared by the mammalian homologs and related Dipterans (see Chapter Four). The final exons of this gene are still missing from the genome sequence database, most likely because they reside in unclonable areas.

Dbp80 resides in a repetitive environment mostly consisting of degenerated transposable element sequences. These repeats are also located in the introns and result in a significant degree of sequence polymorphism, which can be visualized by Southern analysis of DNA from different strains cut with the same restriction enzyme (Figure 3.10). *Dbp80* possesses initiator and TATA box sequences at +1 and –33, is moderately highly expressed, and shows no signs of alternative splicing (Figure 3.11a,b). This is in contrast to its *C.elegans* and mammalian homologues (gene structures can be seen by searching these websites: www.wormbase.org/; www.ncbi.nlm.nih.gov/LocusLink/). There are 28 ESTs in the database, originating from libraries constructed from all major developmental stages and tissues. *Dbp80* developmental expression indicates a strong maternal component, and exhibits different levels of expression as development proceeds, suggesting it may be developmentally regulated (Figure 3.12).

Dbp80* is absent under deficiencies which remove *lethal 2

Lucchesi's group was able to establish that *Dbp80* probably mapped to 3L heterochromatin, but they could not map the gene more specifically. However, both Southern and PCR analysis shows that *Dbp80* coding sequences are removed in deficiencies which also remove *lethal 2* (see Figure 3.13a,b). Further evidence establishing the connection between *lethal 2* and *Dbp80* was, however, not forthcoming. Repeated attempts to use PCR to amplify *Dbp80* genomic DNA from a P element specific primer, and to show a band shift on a Southern blot indicating the presence of a P element in *Dbp80* were unsuccessful (Figure 3.14).

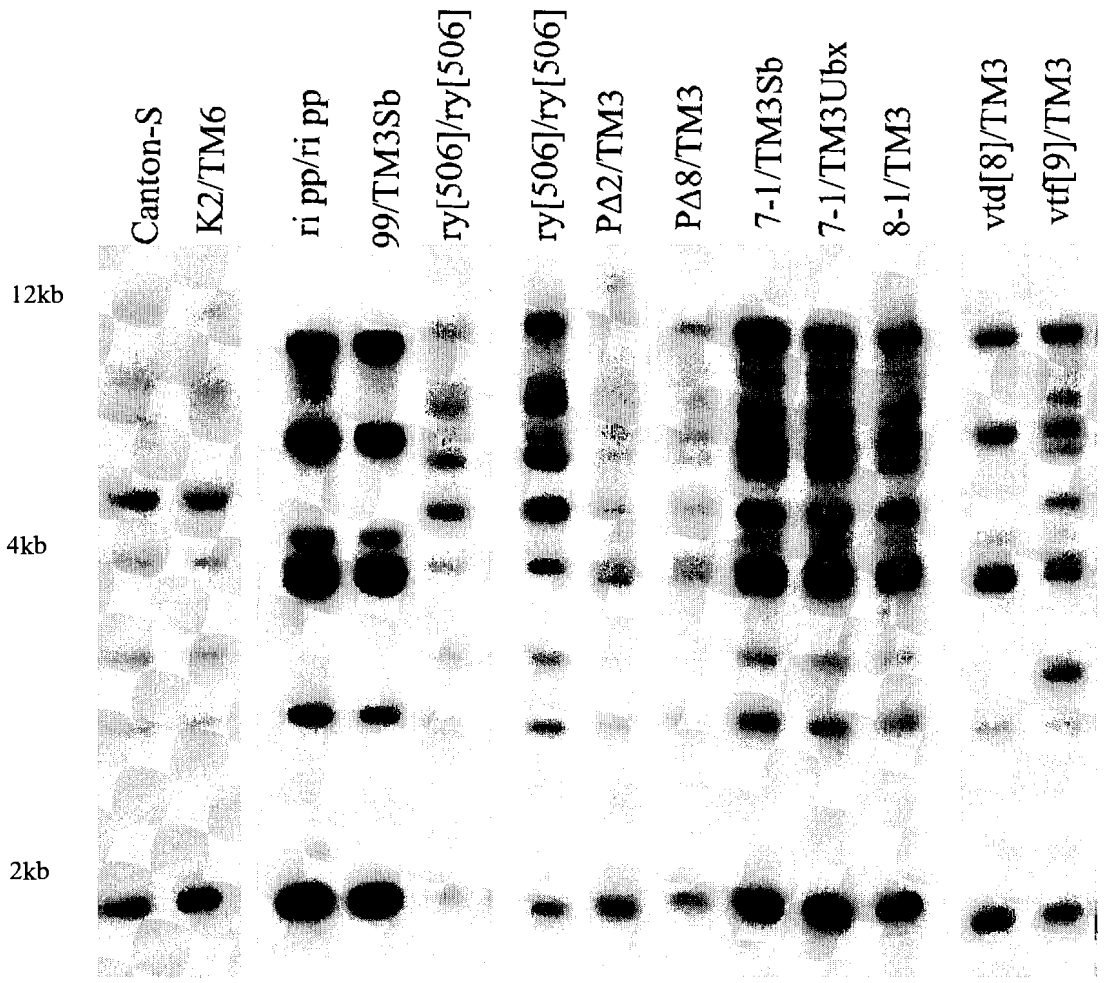
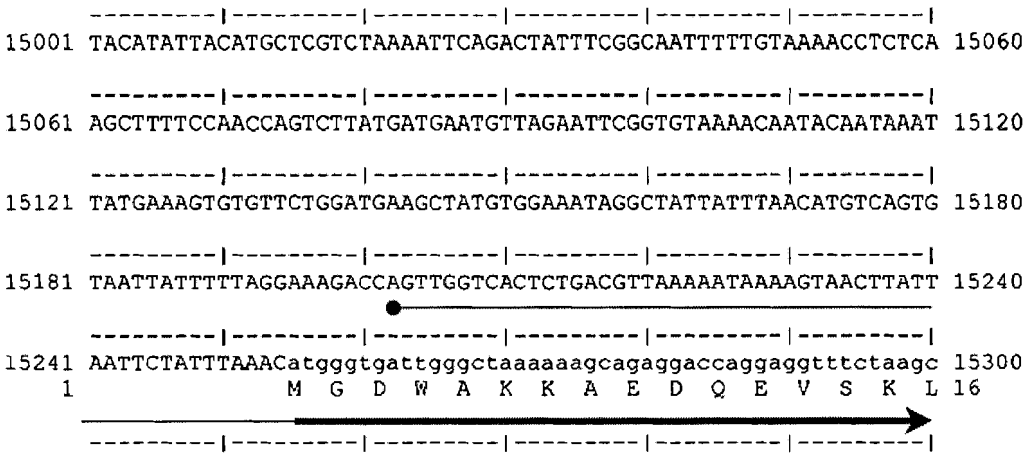
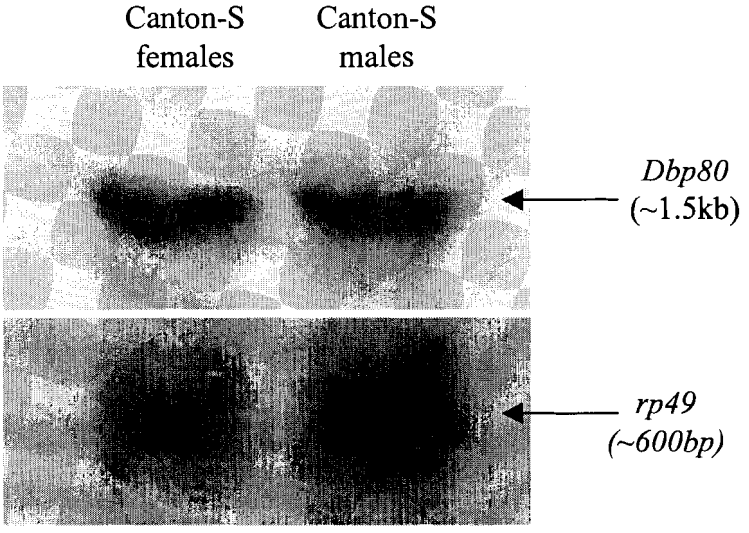


Figure 3.10: *Dbp80* polymorphism
Polymorphic banding pattern which results when genomic DNA from different genetic strains (all wildtype for *Dbp80*) is cut with EcoRI and probed with the 1.5kb Dbp80 cDNA.



(a)



(b)

Figure 3.11: *Dbp80* promoter region and Northern data
(a): Schematic of *Dbp80* 5' genomic region containing promoter consensus sequences.
(b): Total RNA from *Drosophila melanogaster* males and females probed with the *Dbp80* cDNA.

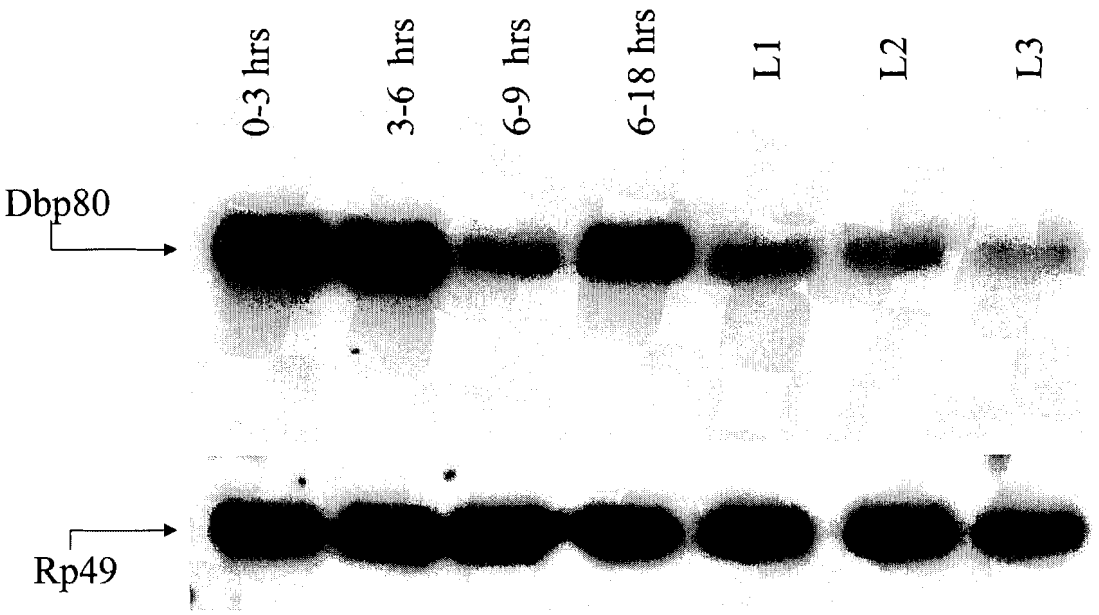


Figure 3.12: *Dbp80* developmental Northern
Poly(A+) RNA (2 ug per lane) from various *Drosophila melanogaster* stages, probed with the *Dbp80* cDNA (~1.5kb) and *rp49* (~550bp) for loading control.

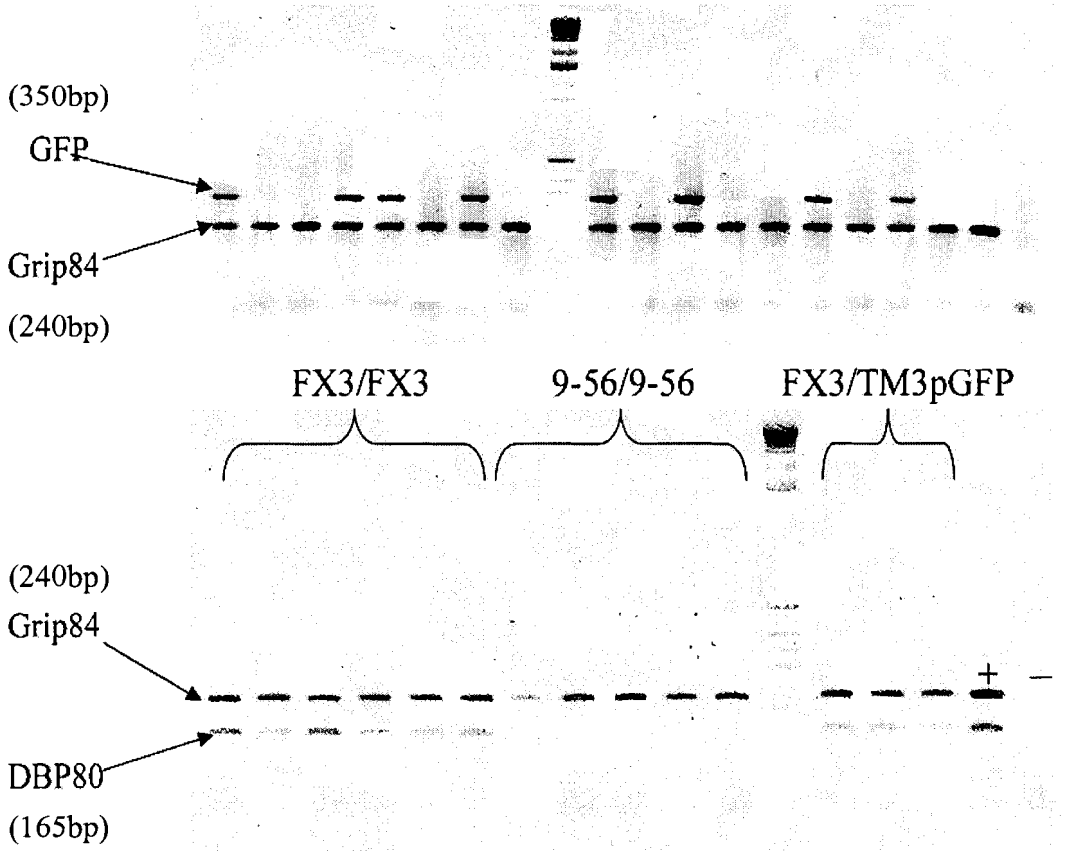


Figure 3.13(a): PCR mapping of *Dbp80*

Deficiency chromosomes are balanced with a multiply rearranged third chromosome containing sequences for GFP (Green Fluorescent Protein). TOP: DNA from individual embryos is tested with primers to the GFP sequences from the balancer, and primers to another gene in the background to show the presence of DNA (in this case, the background gene is the X-linked *Grip 84*). Lanes in which no GFP signal is evident represent genotypes homozygous for the respective deficiencies. BOTTOM: DNA from homozygous deficient embryos tested for the presence/absence of *Dbp80* coding sequences. These results show that *Dbp80* coding sequences are absent under *Df(3L)9-56*, which removes *lethal 2*, but present under *Df(3L)FX3*, which does not remove *lethal 2*. *Df(3L)FX3* does however remove *lethal 3*, which cannot therefore be *Dbp80*. (The image colour is reversed for clarity.)

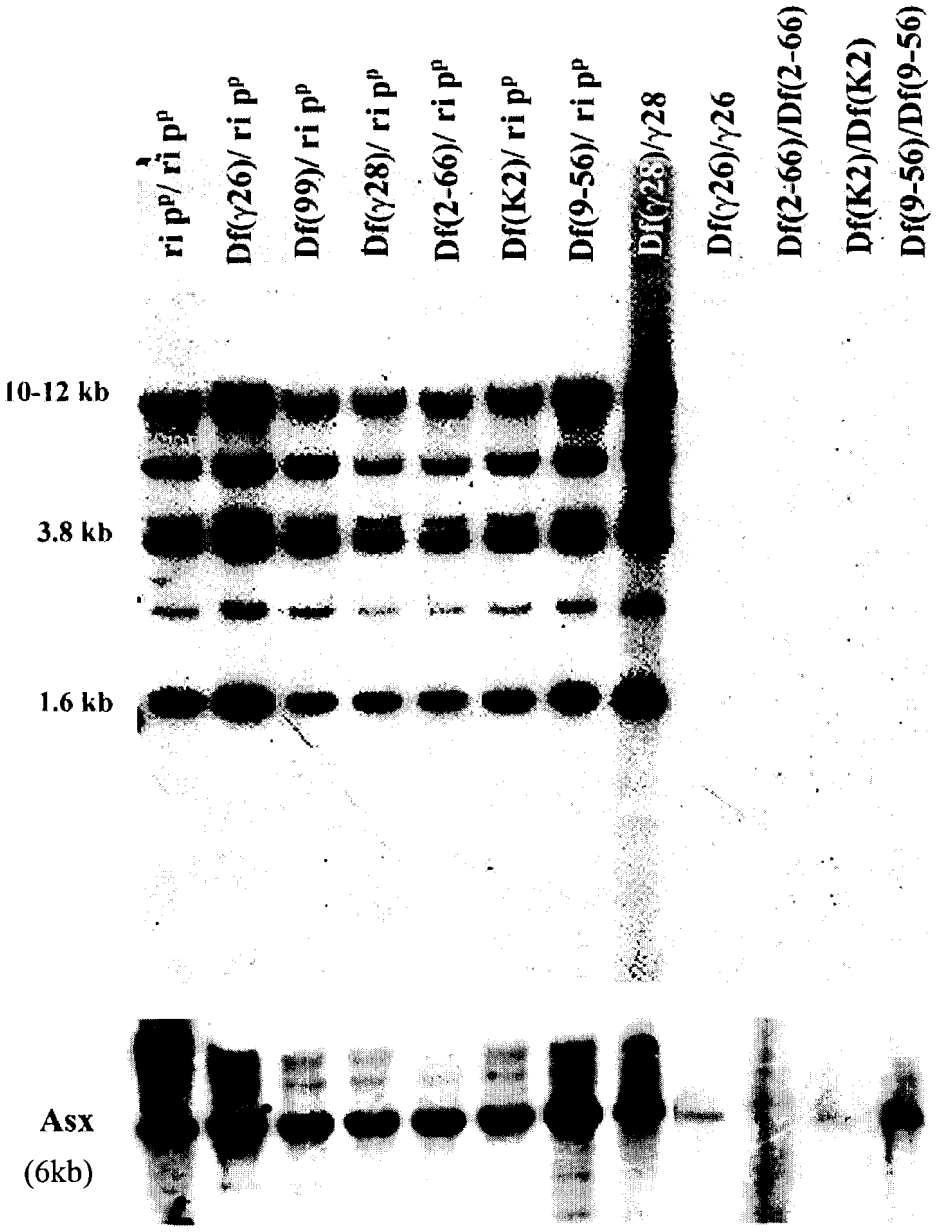


Figure 3.13(b): Southern mapping of *Dbp80*
Genomic DNA from various genotypes which include or remove *lethal 2*, probed with the *Dbp80* cDNA, and *Asx* as a loading control. *Dbp80* is absent under deficiencies that remove *lethal 2*. Note that *Dbp80* appears to be a single-copy gene (all bands representing *Dbp80* coding sequence disappear in the homozygous mutant lanes).

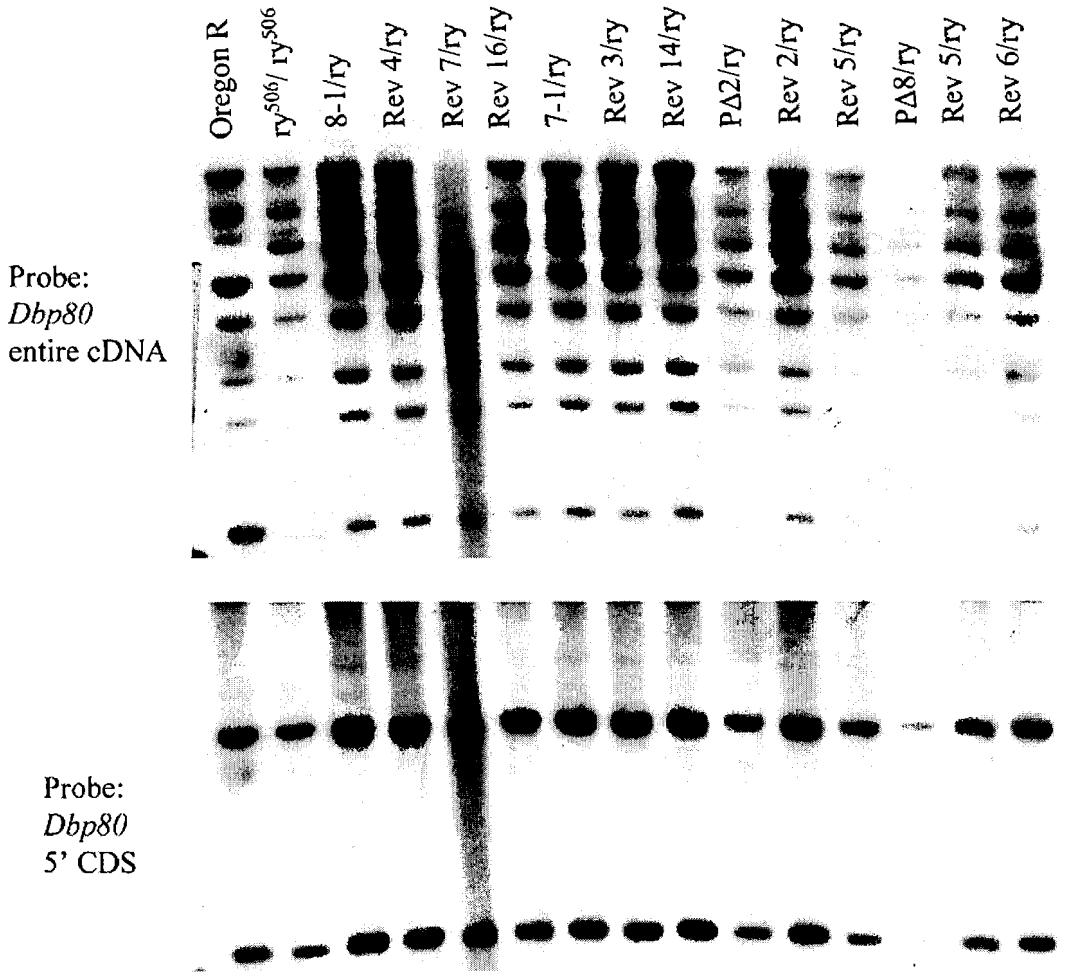


Figure 3.14: Genomic Southern of *Dbp80* showing no band shift in *lethal 2* P mutants

Eco RI-cut Genomic DNA from *lethal2* P alleles and revertants, probed with the entire *Dbp80* cDNA (TOP), and just the 5'portion (BOTTOM). No band shift indicating the presence of linked P element sequences is evident in the 5' region of this gene.

RpL15 shares a genomic scaffold with the first 4 exons of *Dbp80*

During the course of the genome sequence assembly and its revisions, *RpL15* has consistently shared a genomic scaffold with the first 4 exons of *Dbp80*, and this relationship has been preserved in the most recent heterochromatic genome project release (Figure 3.1). These two genes are approximately 10kb apart, which is comparable to the density of genes in euchromatin (Adams et al. 2000). *RpL15* encodes a ribosomal protein and belongs to a distinctive group of genes scattered throughout the genomes of all eukaryotes sequenced to date. This is in contrast to their clustered arrangement in prokaryotes, where they are organized into discrete operons, but the requirement for coordinated and constitutive regulation has not been lost. They are among the most highly conserved and ancient components of the cell's machinery (Figure 3.15, Table 3.6). In *Drosophila,* mutations in many ribosomal protein genes have been shown to express a dominant *Minute* phenotype (Lambertsson, 1998). The classical *Minute* phenotype displayed by *lethal 2* transheterozygotes, and the proximity to *Dbp80* which has been confirmed to reside in proximal 3L heterochromatin, implicated *RpL15* as an even stronger candidate for *lethal 2*, and indeed by PCR it maps to the same region (Figure 3.16). Interestingly, it does not fit the model of a typical heterochromatic gene except for its repetitive environment. It is a small gene with three exons and two introns. In total it spans less than 2 kb of genomic DNA. As has been established for nearly all ribosomal proteins genes across taxa, transcription initiates in the vicinity of a polypyrimidine tract (Mager 1988, Hariharan and Perry 1990, Bakarat et al. 2001, Yoshihama et al. 2002), and there are no other promoter consensus sequences (Figure 3.17). As for *Dbp80, RpL15* is embedded in a repetitive environment, with repeats both upstream and downstream, and within the introns (discussed further in Chapter 4, see Table 4.4). Like all other ribosomal proteins characterized to date, *RpL15* appears to be highly expressed throughout development.

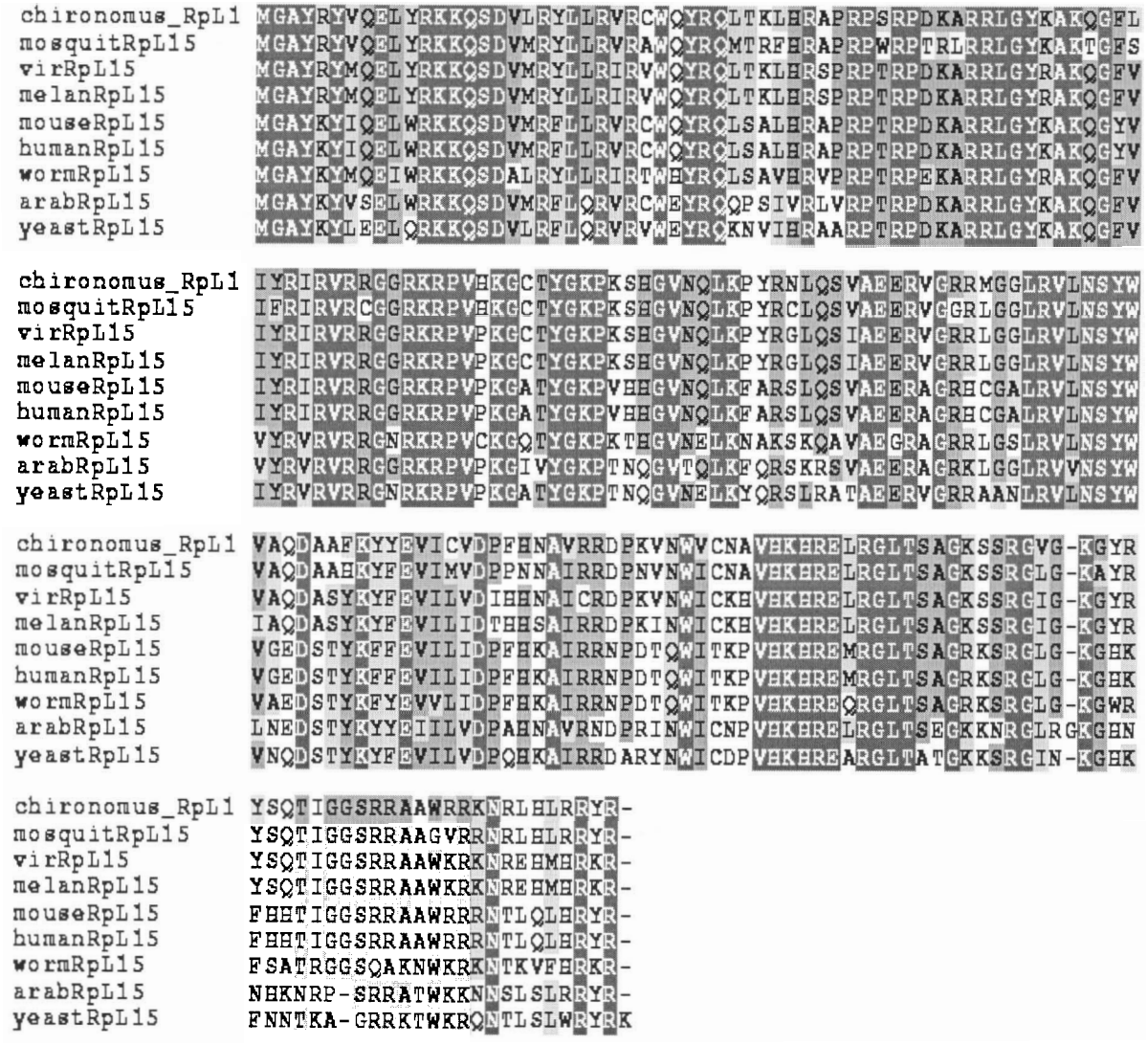


Figure 3.15: CLUSTALW alignment of RpL15 homologues across different taxa.

Organism	%similarity	%identity
Chironomus tentans	94%	84%
Anopheles gambiae	89%	79%
Mus musculus	86%	74%
Homo sapiens	86%	74%s
Ictalurus punctatus	86%	74%
Caenorhabditis elegans	83%	70%
Picea mariana	82%	66%
Neurospora crassa	80%	68%
Aspergillus niger	80%	66%
Saccharomyces cerevisiae	80%	67%
Arabidopsis thaliana	77%	63%
Leishmania infantum	72%	55%

Table 3.6: RpL15 protein homologies across different taxa.

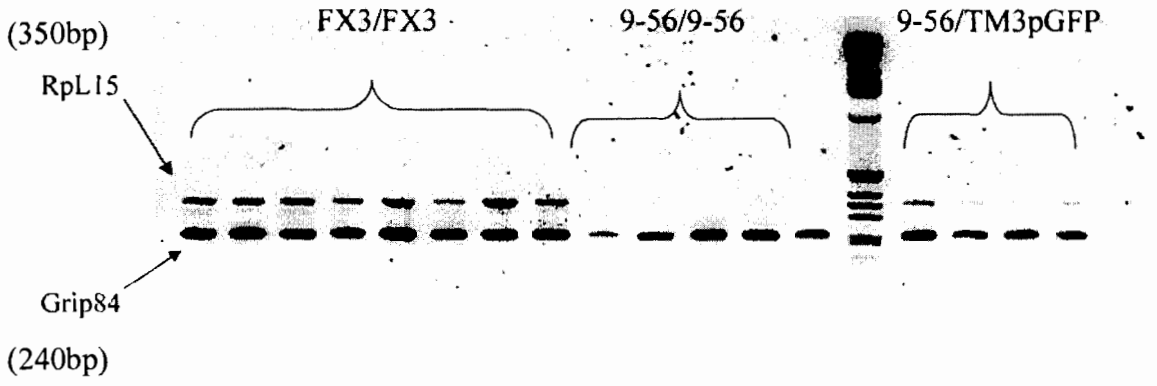
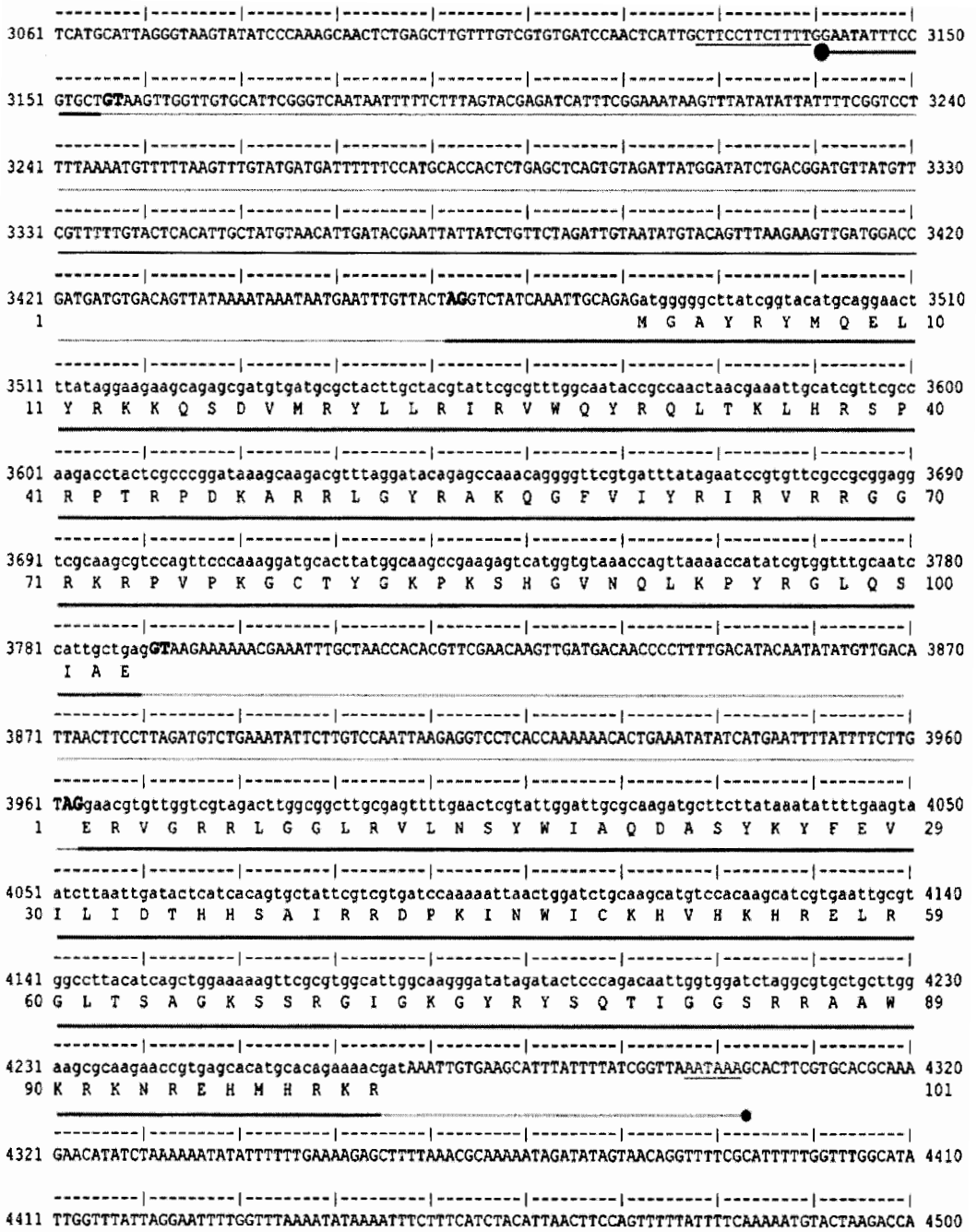


Figure 3.16: PCR mapping of *RpL15*

PCR results showing that *RpL15* coding sequences are absent under *Df(3L)9-56*, which removes *lethal 2*, but present under *Df(3L)FX3*, which does not remove *lethal 2*. *Df(3L)FX3* does however remove *lethal 3*, which cannot therefore be *RpL15*.*Grip84* is a gene on the X chromosome, used as a DNA control. The image colour has been reversed for clarity. Please see Figure 3.13(a) for a description of the PCR assay.

Figure 3.17: Schematic of RpL15 gene organization
The beginning and end of transcription is indicated by a filled gray circle. Thin gray lines represent sequences included in the unprocessed transcript; thicker gray lines represent the primary spliced product. A second processed transcript is found at much lower levels which includes the first intron (see figure 3.32). Primary splicing consensus signals are shown in bold face, and the polypyrimidine tract (CTTCCTTCTTTT) within which transcription likely begins is underlined, as is the termination signal (AATAAA).



lethal 2 mutations are lesions in *RpL15*

In order to establish whether *lethal 2* corresponded to *RpL15* (which is smaller and more tractable than *Dbp80*), the EMS and P mutants were cloned and sequenced. All six mutations in *lethal 2* correspond to lesions in *RpL15,* and these are diagrammed in Figure 3.18. The sequence of the EMS mutant allele *72* is shown in Figure 3.19a. It possesses a stop codon in the second exon of *RpL15. 72* homozygotes die as L1 larvae, as do homozygotes for *Df(3L)9-56,* so it would appear that *72* behaves as a null. The sequence of the EMS mutant allele *1-166-37* is shown in Figure 3.19b. *1-166-37* homozygotes also die as L1 larvae, so this allele also resembles a null, however the point mutation is in a splicing consensus, (GTAAA instead of the canonical GTRAG where R = purine: Mount et al. 1992).

Attempts to show that the P alleles of *lethal 2* tag *RpL15* using Southern analysis were successful, and the results were confirmed by PCR (Figure 3.20). In addition, the lethal excision alleles generated from *PΔ8* and *8-1* display a variety of molecular lesions in *RpL15* (Figure 3.21a), two of which also appear to have affected *Dbp80* (Figure 3.21b). All lethal excisions in *RpL15* complement mutations in both *lethal 3* and *lethal 1*. Sequencing from the P element indicates that all four P alleles appear to share the same insertion site 18 bp upstream of the *RpL15* pyrimidine tract (Figure 3.18), and sequencing ~500bp on either side of the P elements indicates that they are all simple insertions, with no rearrangements or deletions. However, PCR results indicate there were at least two different events, since the P element in *PΔ8* is inserted in the same orientation as the gene, while the elements in *7-1* and *8-1* are inserted in the opposite orientation. The sequence in which the P element has inserted does not conform to a published consensus (O'Hare and Rubin 1983) but this consensus is considered weak in any case, and derived from euchromatic insertions.

Southern analysis and sequencing show that *PΔ2* has undergone a spontaneous deletion event which probably happened during the removal of the background chromosomes (Figure 3.20). The P insertion in this allele also appears to be simple, but the substantial

reduction in the element's size may have caused PΔ2 to become the weakest allele in the series, with approximately 5% of the stock progeny eclosing as sterile homozygous adults. The P element in *PΔ2* has lost most of its internal sequences but retained its 31 base pair inverted repeats, and 144 bp 5' and 366 bp 3' sequences, yielding an insertion of approximately 600bp (Figure 3.22). Interestingly, there appears to have been an expansion of highly repetitive sequence between the two remaining P element fragments. There is no obvious periodicity to this AT rich expansion, but it bears no resemblance to any part of the P element sequence or surrounding genomic flank.

Dbp80 corresponds to neither *lethal 1* nor *lethal 3*

An attempt was made to position *Dbp80* relative to *RpL15*, in particular to determine whether or not *Dbp80* might correspond to either *lethal 1* or *lethal 3*. For this experiment the same PCR technique used to position both genes relative to *lethal 2* was employed with overlapping deficiencies (see Chapter Two). As Figure 3.23 shows, *Dbp80* signal is present under a deficiency which removes *lethal 3* (*Df(3L)FX3*) but absent under a deficiency which does not remove *lethal 1* (*Df(3L)K2*). Therefore *Dbp80* can be neither *lethal 3* nor *lethal 1*, which suggests that either it is not an essential gene, or that is has been missed by all screens which have been carried out in this region to date (Marchant and Holm 1988b, Schulze et al. 2001). This is further supported by the fact that those excisions from *RpL15* that also affect *Dbp80* (Figure 3.21b) are completely viable over *lethal 3* mutant alleles (see above).

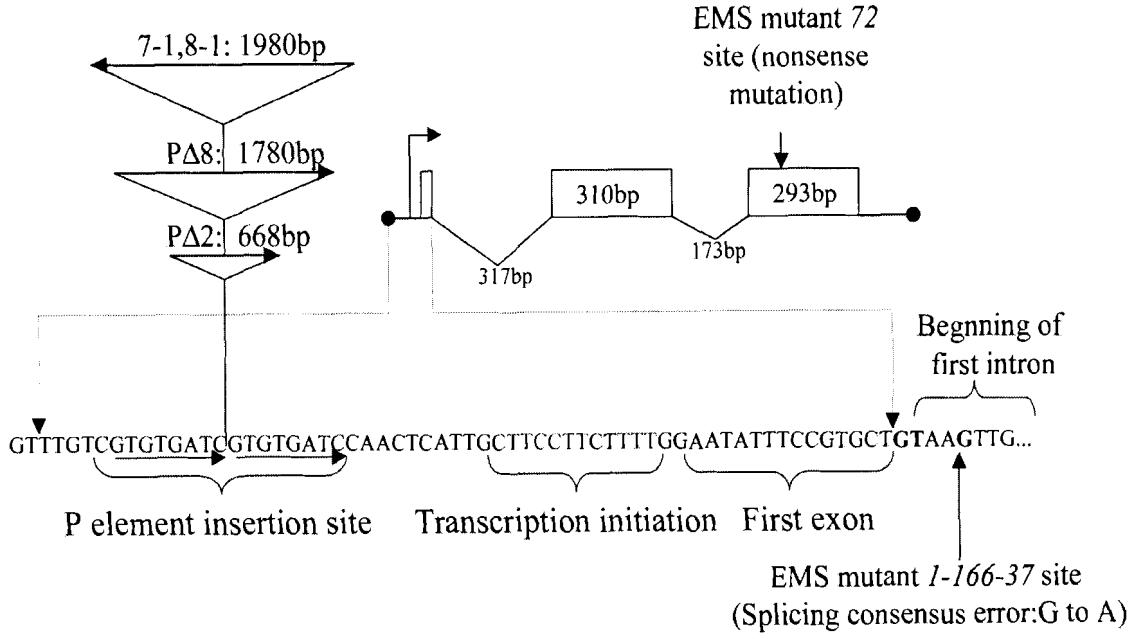
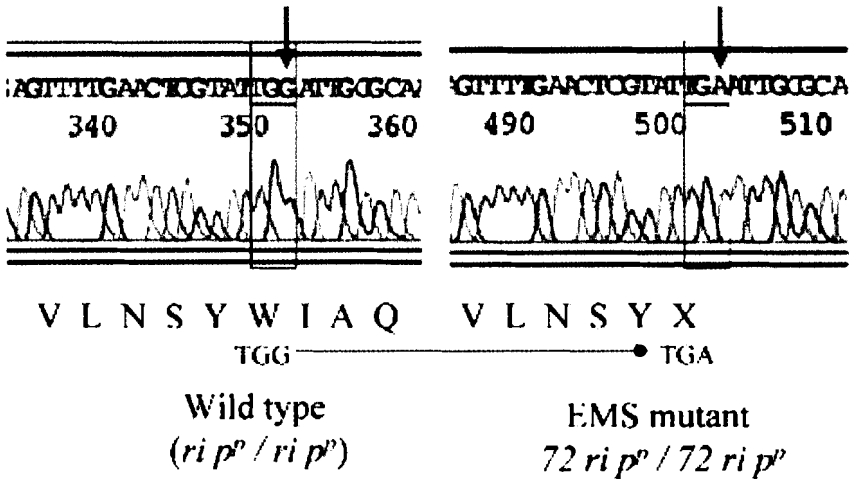
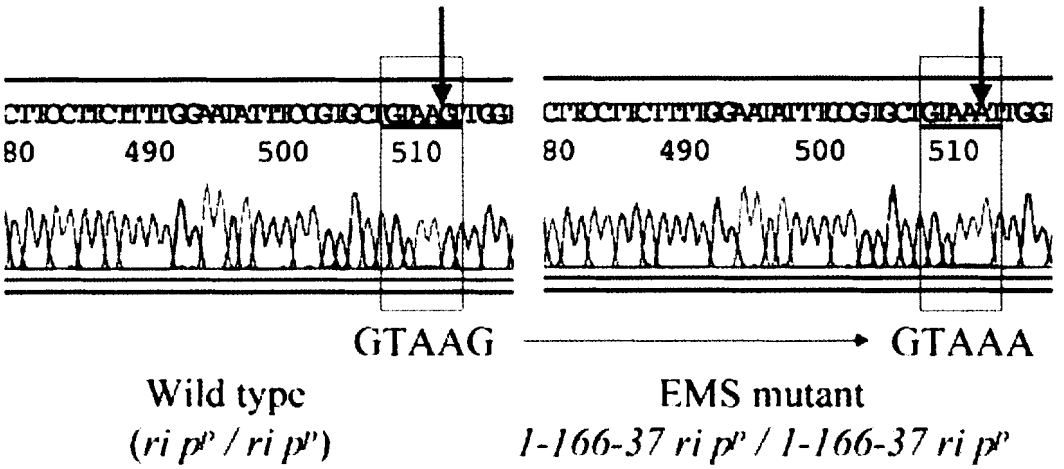


Figure 3.18: *RpL15* gene organization and location of mutations
Diagramme of *RpL15* gene structure and the location of the various *lethal 2* mutations. Note that the 5' region of the gene containing five out of 6 lesions is expanded in this figure.



(a)



(b)

Figure 3.19: Sequence of *lethal 2* EMS mutations
(a): Sequence of *lethal 2* EMS mutant *72*: a nonsense mutation in the third exon.
(b): Sequence of *lethal 2* EMS mutant *1-166-37*: a mutation in the 5'splicing consensus of the first intron.

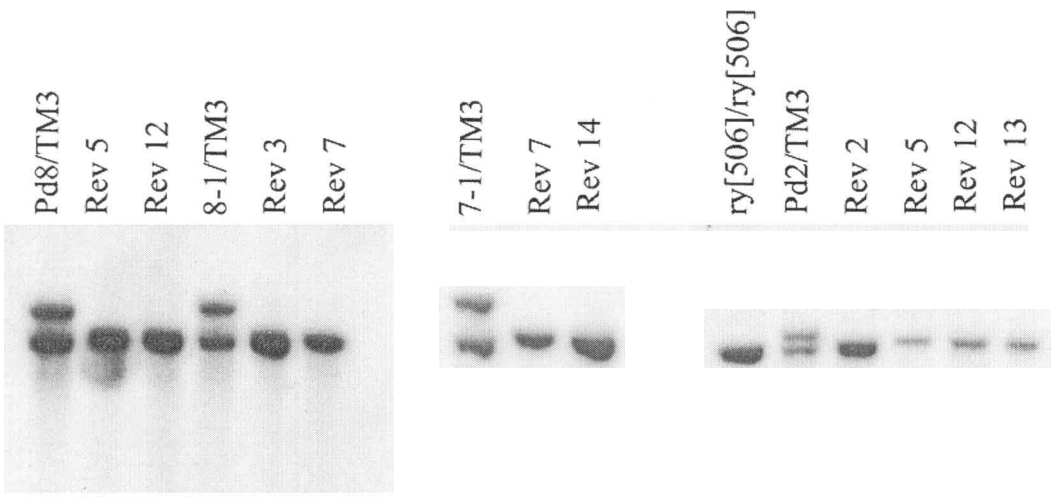


Figure 3.20: *lethal 2* P mutant Southern probed with *RpL15*
Eco RI-cut genomic DNA from *lethal 2* P mutants and revertants, probed with the *RpL15* cDNA. All genotypes are balanced over TM3 and are therefore heterozygous for the P mutant allele (except ry^{506}, which is homozygous and wild type for *lethal 2*, representing the background chromosome on which the P mutant alleles were induced). Note that the lanes containing P mutant heterozygous DNA show two bands: the lower molecular weight band corresponds to the *RpL15* gene region on the balancer (which is wild type for *lethal 2*), and the higher molecular weight band corresponds to the *RpL15* gene region including a P element tag. The revertant lanes show no band shift associated with P element insertions, as expected. Finally, *PΔ2* appears to be tagged by a smaller P element than the other alleles.

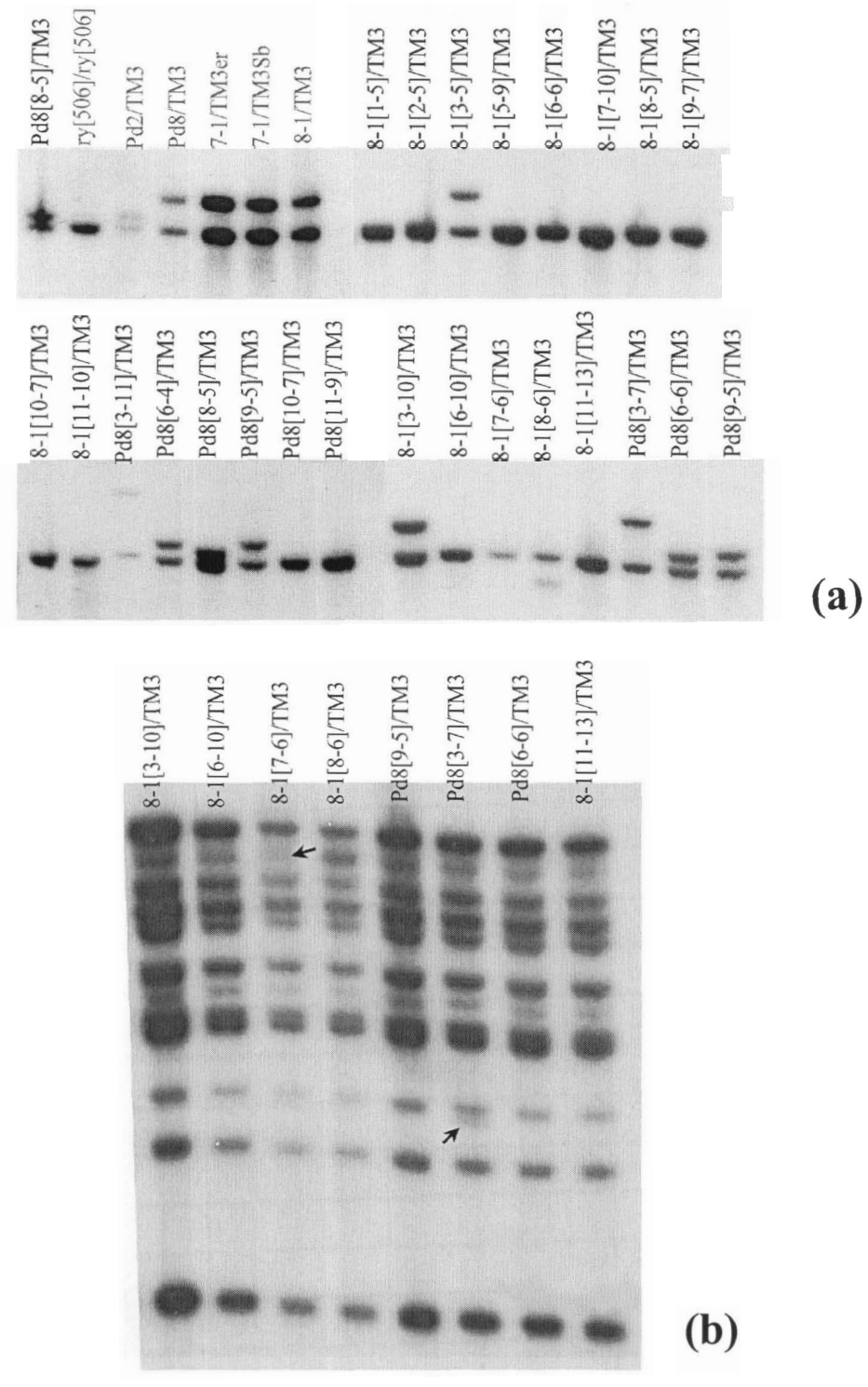


Figure 3.21: *lethal 2* excision Southerns
(a): Eco RI-cut genomic DNA from various *lethal 2* P mutant and excision derivatives, probed with the *RpL15* cDNA. (b): Eco RI-cut genomic DNA from various *lethal 2* P mutant excision derivatives, probed with the *Dbp80* cDNA. Arrows point to those bands which are altered in the excisions.

Figure 3.22: Sequence of the insertion site of the *lethal 2* P allele *PΔ2*
Flanking genomic sequence is in lower case; arrows mark the flanking direct repeats and the P element inverted repeats; P element sequences are in bold, upper case letters; the intervening AT-rich sequence expansion is depicted in italics. The polypyrimidine tract which serves as a transcription initiator is underlined and in bold face, as is the START methionine.

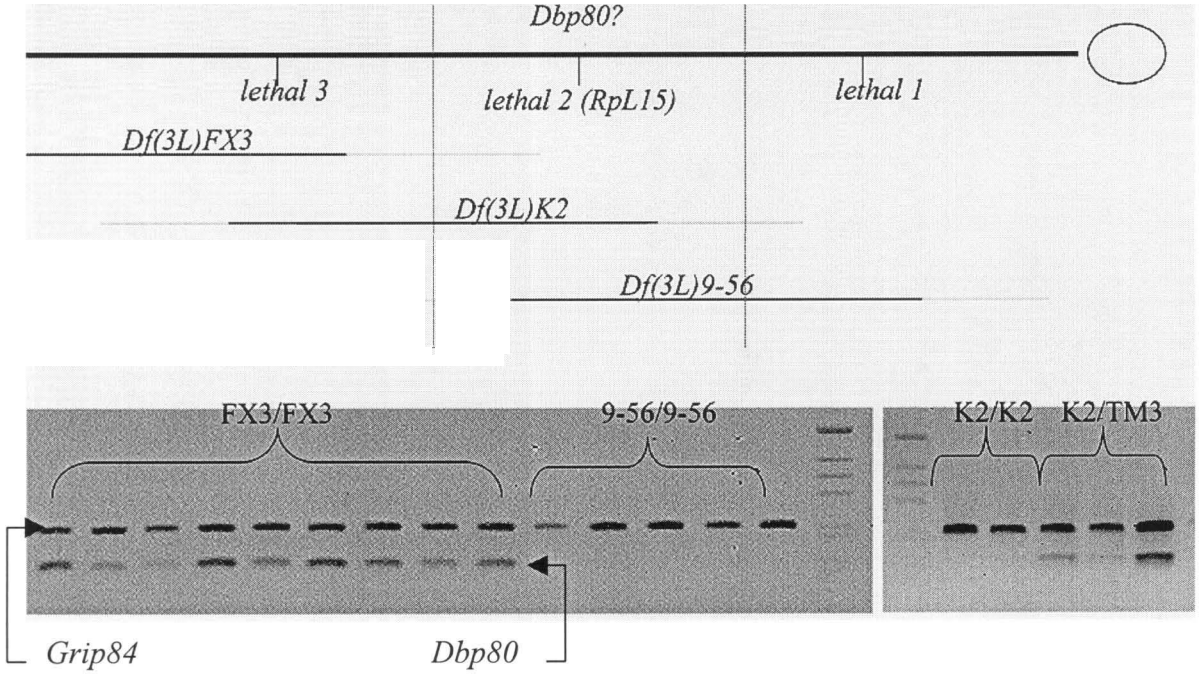


Figure 3.23: PCR mapping of *Dbp80* relative to *lethal 3* and *lethal 1*
PCR results showing that *Dbp80* coding sequences are present under deficiencies which remove *lethal 3* (FX3) or absent under deficiencies which do not remove *lethal 1* (K2), establishing that neither of these genes can be *Dbp80.* (See Figure 3.13a for a description of the PCR protocol). The lower molecular weight band is *Dbp80* and the higher molecular weight band represents *Grip84,* a gene on the X chromosome used as a DNA control. (The image colour of the PCR gel has been reversed for clarity.)

3.2.3: Transgenic analysis of lethal 2

Transgenic analysis was carried out for two purposes. Firstly, rescue of *lethal 2* mutants by germline transformation of an *RpL15* transgene would constitute further evidence establishing the connection between *RpL15* and *lethal2*. Secondly, since *RpL15* is a small gene, and genomic constructs are relatively easy to generate, a series of transgenic tools can be made and used to observe the effects of moving a heterochromatic gene into different chromatin environments. The details of transgene construction and crossing schemes are outlined in Chapter Two. In brief, two different transgenic vectors were used, one with an on board heat-shock sensitive promoter (pCaSpeR), the other with upstream activating sequences (pUAST) permitting it to be induced by a driver transgene located elsewhere in the genome. This was to address the possibility that direct expression by heat shock could cause overexpression of the transgene, which might be toxic to the system.

Four constructs were made: two genomic (one 3.4kb HindIII insert containing the gene in pCaSpeR, one 2.2kb BglII-HindIII subclone of the former in pUAST) and two cDNA (identical inserts, one in pCaSpeR, one in pUAST). The HindIII genomic fragment was isolated from a *Drosophila melanogaster* phage library. This fragment was cut with BglII to produce the smaller 2.2kb genomic construct (Figure 3.24). For the cDNA, a full length EST from the RE library, designated RE01373, was selected. All constructs were sequenced several times to ensure fidelity (Figure 3.25a,b). Note that in the genomic constructs there are some single nucleotide differences with respect to the BDGP sequence – but none of these changes fall within any coding region. In addition, I have cloned part of this region from another strain (*ri* p^p), and have determined that at least one of these changes (in the second intron) is a natural polymorphism (data not shown). Transgene lines, evidence of transgene integration and expression, summaries of genetics and heat shock protocols are outlined in Figure 3.26, Chapter Two, and Table 2 in appendix. Note that for the 3.4kb HIII construct, heat shock induction is not possible, since I was unable to obtain any inserts in the correct orientation. But the main objective of the experiment was to see if a heterochromatic genomic transgene could express in

euchromatin using its own regulatory sequences. As a control, the smaller BglII-HindIII genomic construct and the cDNA constructs were subjected to various driver protocols.

Four levels of germline rescue can be predicted: rescue of lethality (deficiencies in combination with each other or with lethal alleles), viability (improving the number of adult escapers from certain heteroallelic combinations of *lethal 2* mutants),fertility (rescuing the sterility associated with the aforementioned heteroallelic combinations that produce escapers), and the *Minute* phenotype. Rescue for lethality was not observed, with the exception of two flies (from two different *Deficiency/EMS* transheterozygote experiments), which appear to have died immediately upon eclosion and had to be picked from the floor of the bottle. Such flies were not found in control crosses. Rescue for viability was observed, and the results are shown in Figure 3.27a,b. In this experiment, a heteroallelic combination of P alleles that normally produces appreciable numbers of escaper offspring were made with both driver and transgene in the background. Since the P alleles are all marked with ry^{506}, the rescued (test) generation (F2 in the scheme outlined) was taken out of a *w-/w-* background, so that homozygosis of the ry^{506} markers (and therefore the P alleles) could be scored. In addition, since the transgene was carried by the male parent in the F1 generation, males in the test (F2) generation cannot inherit it, and thus serve as an internal control. Three external controls were set up in parallel, producing test generation genotypes that possessed neither transgene nor driver, transgene only, or driver only. For the F2 (test) generation, the number of ry^{506}/ry^{506} females and males was scored, and expressed as a ratio. An increase in female:male ratio indicates that viability has improved relative to controls. This was only observed for the class of F2 progeny that possessed both driver and transgene. Although this trend was observed for the three X-linked transgenes tested (two cDNA transgenes and one genomic), the data are only statistically significant for one experiment (1373UAS24-1: a cDNA transgene).

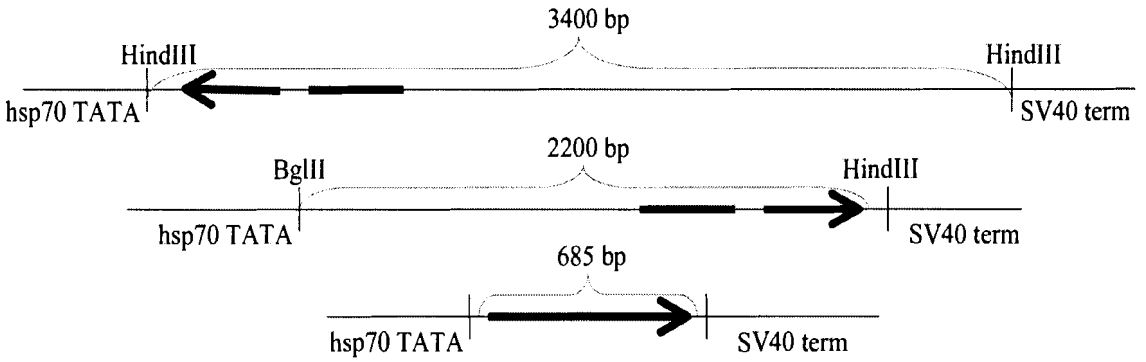
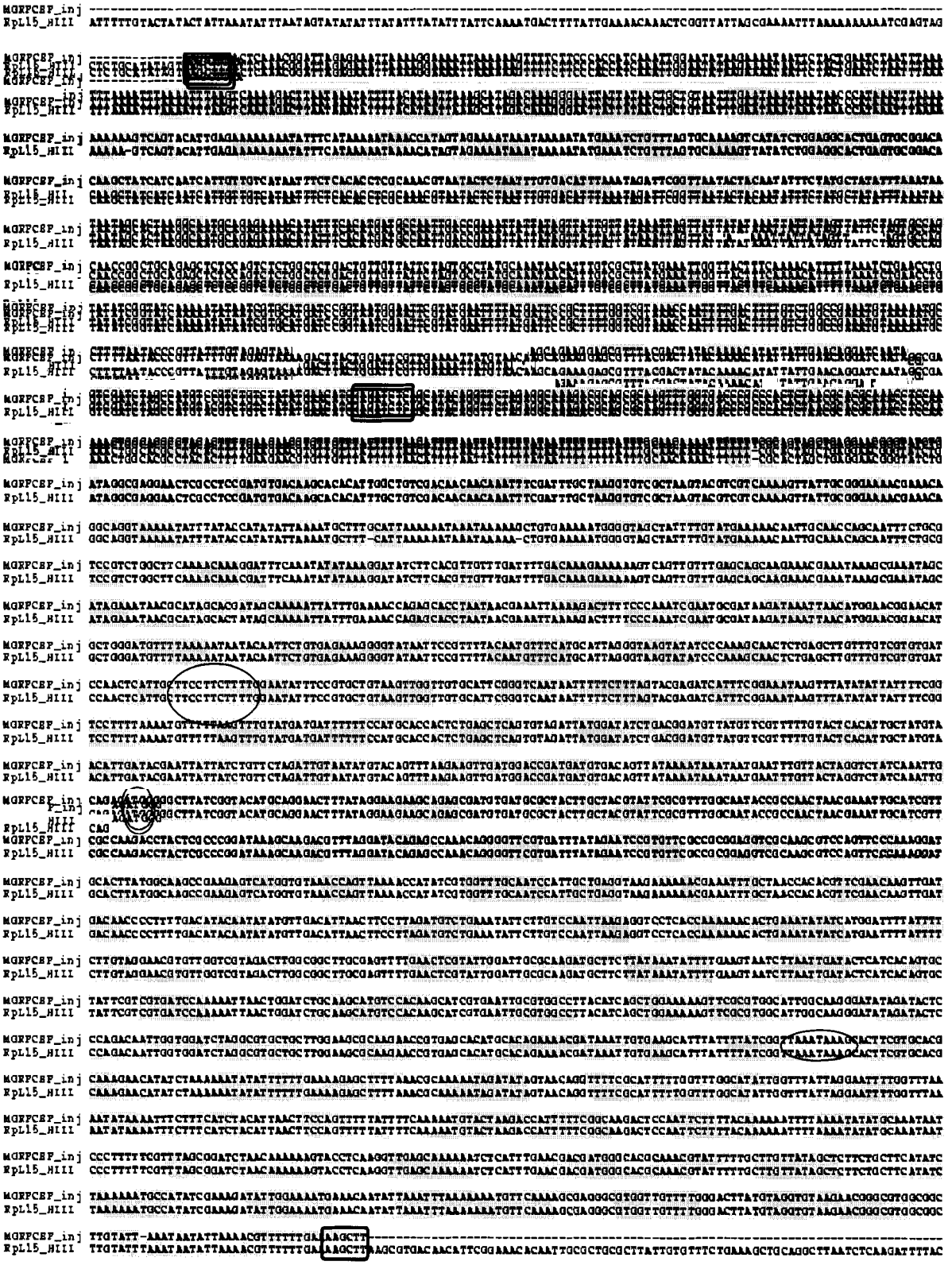


Figure 3.24: Schematic of *RpL15* transgenes
Diagrammes of *RpL15* genomic and cDNA constructs. The 3.4kb HindIII transgene and one cDNA transgene are cloned into pCaSpeR-hs; the 2.2kb BglII-HindIII transgene and the other cDNA transgene are cloned into pUAST. Only one cDNA construct is shown, since apart from the vectors in which they are cloned, they are identical.

Figure 3.25 (a): Sequence of the HindIII *RpL15* genomic transgene
Transgene sequence is aligned with BDGP sequence. Boxes indicate restriction sites (HIII-BGLII-HIII respectively) and circles mark relevant signals (pyrimidine initiator, ATG start and polyadenylation respectively). Mismatches show up as small white blocks against a grey background.



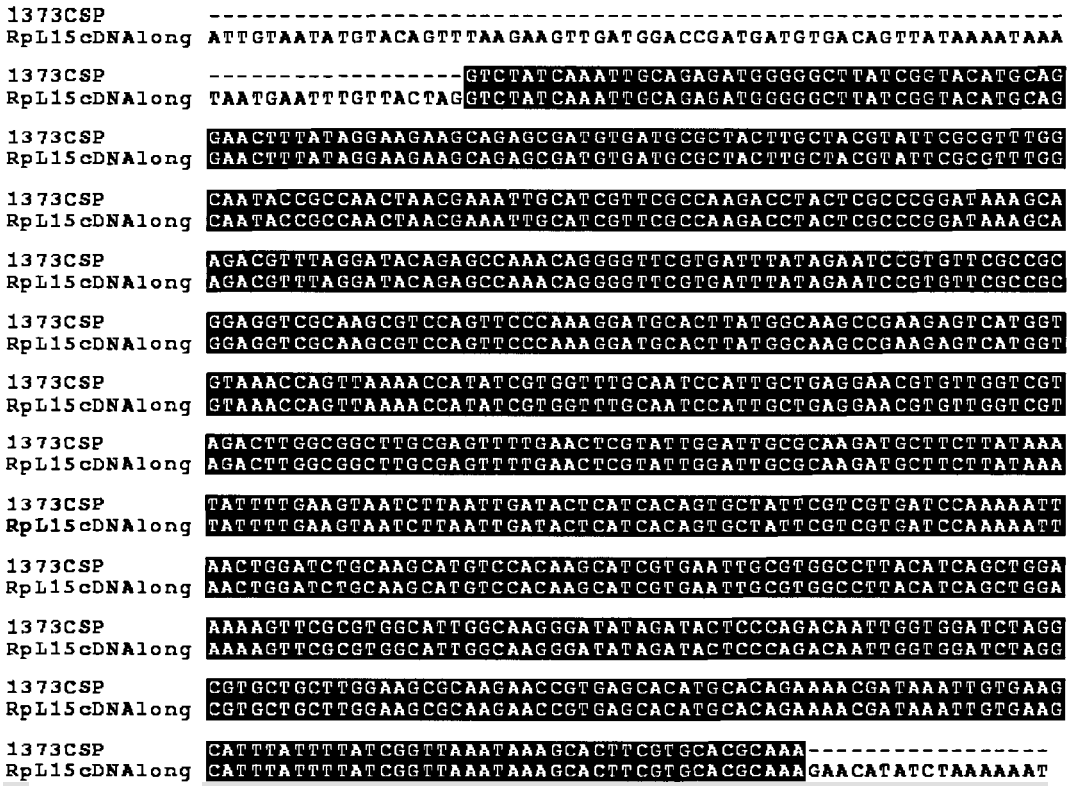


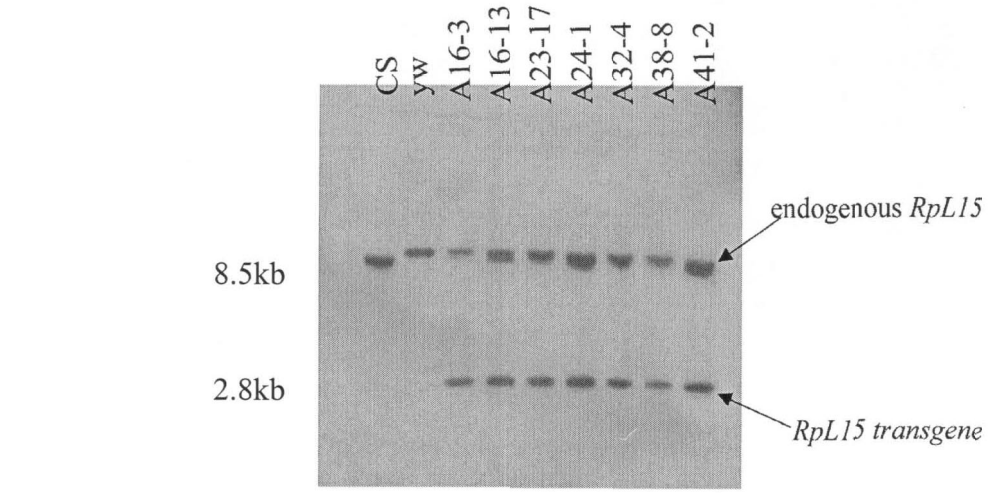
Figure 3.25 (b): Sequence of *RpL15* cDNA transgenes
1373CSP and 1373UAS: cDNA construct sequence aligned with BDGP sequence.

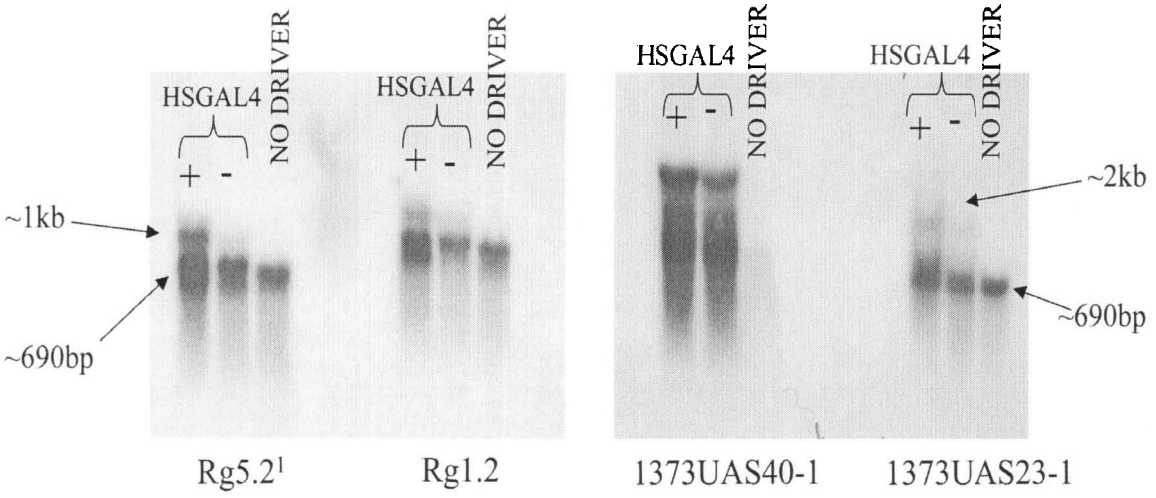
Figure 3.26: Evidence for transgene integration and expression
Samples of experiments to test for presence and expression of injected transgene constructs.

TOP: Southern analysis of transgenic lines containing the 3.4kb HIII genomic construct. Genomic DNA is cut with EcoRI and probed with the *RpL15* cDNA. "A16-3" etc. stand for the HindIII 3.4kb genomic constructs cloned into pCaSpeR-hs.

MIDDLE: Northern analysis of transgenic lines containing the 2.2kb BGL-HIII genomic transgene construct ("Rg1.2" etc.) or the cDNA ("UAS40-1" etc.), in pUAST, under heat-shock driver conditions. "+" indicates heat shock application (37°C for 40 minutes followed by a recovery period at room temperature for 2 hours), "-" indicates no heat shock application. Total RNA is probed with the *RpL15* cDNA. "HSGAL4" refers to the Heat shock driver on chromosome II.

BOTTOM: Northern analysis of transgenic lines containing the 2.2kb BGL-HIII genomic construct or the cDNA, in pUAST, under constitutive driver conditions. Total RNA is probed with the *RpL15* cDNA. "ACT" refers to the Actin driver on chromosome II, and "ARM" refers to the Armadillo driver on chromosome II.


CS
yw
A16-3
A16-13
A23-17
A24-1
A32-4
A38-8
A41-2
endogenous RpL15
8.5kb
2.8kb
RpL15 transgene


HSGAL4
NO DRIVER
+
-
~1kb
~690bp
~2kb
~690bp
Rg5.2[1]
Rg1.2
1373UAS40-1
1373UAS23-1

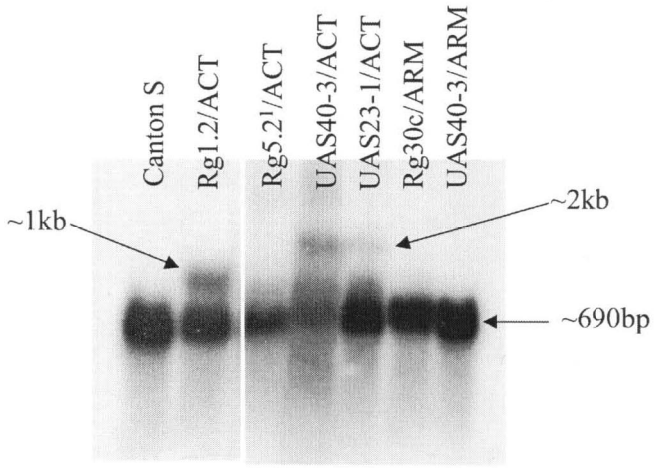

Canton S
Rg1.2/ACT
Rg5.2[1]/ACT
UAS40-3/ACT
UAS23-1/ACT
Rg30c/ARM
UAS40-3/ARM
~1kb
~2kb
~690bp

Figure 3.27(a): *lethal 2* viability rescue genetic crossing scheme
The last two generations from a multigeneration scheme are shown. In sum, F1 males carrying an X-linked transgene, a driver on the second chromosome and one P allele of *lethal 2* on the third chromosome were crossed to females bearing a second *lethal 2* P allele on the third chromosome. These females were also in a *white*$^{+}$ background so that homozygosis of ry^{506} that marks the P allele-bearing chromosome could be scored. In addition, F2 males (not shown) cannot receive the transgene and so serve as an internal control. The number of ry^{506}/ry^{506} females and males are scored and expressed as a ratio. An increase in this (female to male) ratio indicates rescue. Three independent experiments were carried out for three different transgenes; each with three independent (external) controls (F2 males with no driver, no transgene, F2 males with transgene only, F2 males with driver only). UAS24-1 and UAS26-1 are both X-linked *RpL15* cDNA transgenes, and *Rg1.2* is an X-linked genomic transgene.

F1: *w/Y; Act-5C GAL4/CyO; +/+* ♂♂ X *UAS [w+]/w; +/+; PΔ8 ry^{506} /TM3* ♀♀

F2: *UAS[w+]/Y; Act5-C GAL4/+; PΔ8 ry^{506}/+*♂♂ X *+/+; +/+; PΔ2 ry^{506} /TM3* EXPERIMENT

w/Y; + /+; PΔ8 ry^{506}/+♂♂	X	*+/+; +/+; PΔ2 ry^{506} /TM3*	CONTROL-1
UAS[w+]/Y; +/+; PΔ8 ry^{506}/+♂♂	X	*+/+; +/+; PΔ2 ry^{506} /TM3*	CONTROL-2
w/Y; Act5-C GAL4/+; PΔ8 ry^{506}/+♂♂	X	*+/+; +/+; PΔ2 ry^{506} /TM3*	CONTROL-2

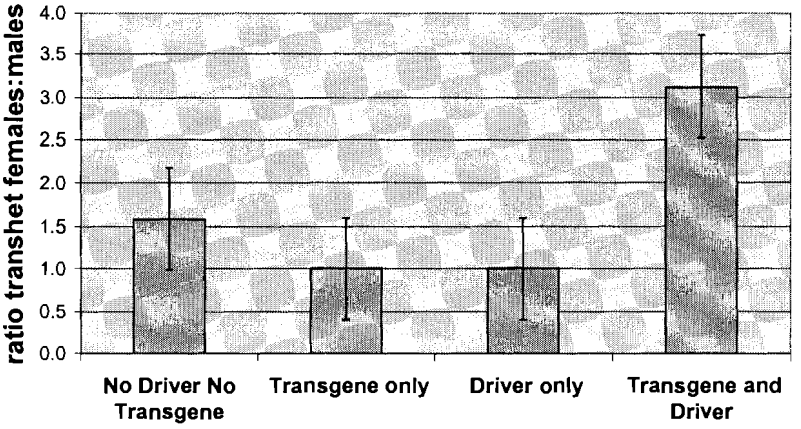
Count *ry^{506}/ ry^{506}* F3 progeny, express female to male ratio.

Experiment 1: UAS24-1 (*RpL15* cDNA transgene)
Experiment 2: Rg1.2 (*RpL15* 2.2kb genomic transgene)
Experiment 3: UAS 26-1 (*RpL15* cDNA transgene)

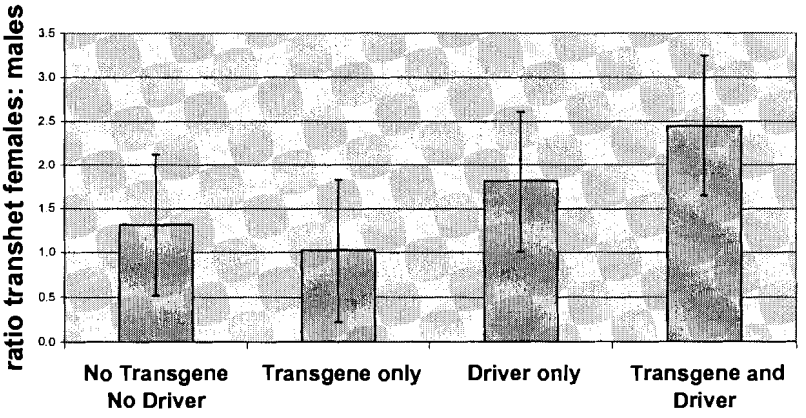
Figure 3.27(b): Statistical analysis of *lethal 2* viability rescue experiments

Data from three separate *lethal 2* viability rescue experiments, each with three independent controls. See legend to previous figure for detailed explanation. Error bars were generated from standard deviations calculated from the three controls for each experiment, and set at 95% confidence level.

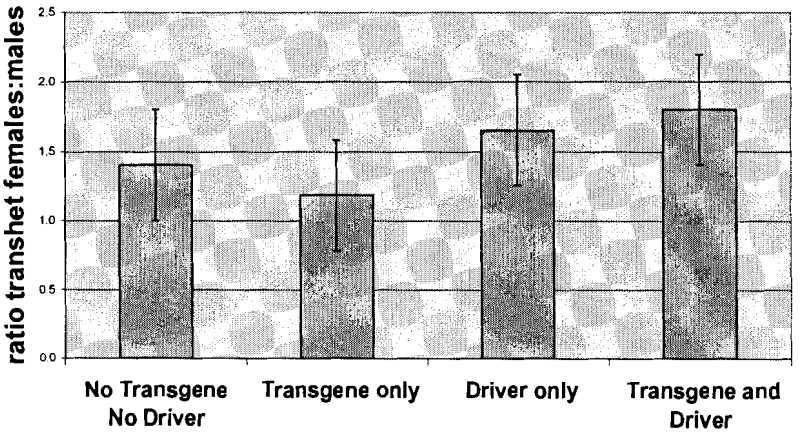
***lethal 2* transhet viability rescue with UAS24-1**
(Experiment #1)



***lethal 2* transhet viability rescue with Rg1.2**
(Experiment #2)



***lethal 2* transhet viability rescue using UAS26-1**
(Experiment #3)



3.2.4: Gene expression analysis

mRNA levels for both *Dbp80* and *RpL15* are suppressed when HP1 dose is reduced

In June 2000, Joel Eissenberg and colleagues published a paper in which they demonstrated that the two best studied heterochromatic genes – *light* and *rolled* – depend upon the product of the *Suppressor of variegation 2-5 (Su(var)2-5)* gene for their proper expression (Lu et al. 2000). *Su(var)2-5* encodes Heterochromatin Protein 1, a major protein component of heterochromatin (see Chapter One). Eissenberg and coworkers demonstrated that the expression of *light* and *rolled* was compromised in a genetic background in which the genetic dose of HP1 was reduced. This was a paradoxical result, for it implied that these heterochromatic genes were dependent upon a product which normally silences gene expression. Using Eissenberg's genetic stocks of HP1 mutants, Northern analysis was carried out on *Dbp80* and *RpL15*. The protocols and crossing schemes are described in Chapter Two, but in brief, total RNA was extracted from three larval stage L3 genotypes: mutants transheterozygous for two different *Su(var)2-5* alleles (*Su(var)2-5*^{149}/ *Su(var)2-5*^{04} : no functional HP1 dose), their heterozygous sibs (*Su(var)2-5*$^{149\ or\ 04}$*/CyO:* one functional HP1 dose) and wildtype (*yw/yw*: 2 functional HP1 doses – note that all *Su(var)* mutants were in a *yw/yw* background). The RNA was probed with *RpL15* or *Dbp80*, and the blots were exposed to a phosphor screen. After phosphoimaging data were collected, *Dbp80* or *RpL15* expression levels were measured relative to loading controls in the same lane, and the results are shown in Figure 3.28. The experiment was repeated three times for each gene. The X axis shown in Figure 3.28 represents the genotypes from which the RNA was extracted, and the Y axis expresses the level of *Dbp80* or *RpL15* expression relative to the loading control *rp49*. These values represent averages calculated from the three experimental repeats, and are normalized to the wild type (two functional HP1 doses) control. The results suggest that both *Dbp80* and *RpL15* transcriptional levels do appear to be negatively influenced by decreased HP1 dose.

The result for *RpL15* is supported by a genetic assay in which a wing phenotype associated with the weak *Notch* allele N^{55e11} is enhanced in a *lethal 2* genetic background.

Defects in ribosomal proteins have been shown to enhance wing morphogenesis mutations (Sinclair et al. 1984, Hart et al. 1993). This is partly because the primordial wing tissue experiences proliferative growth during development, and partly also because many wing morphogenesis mutants encode components of signal transduction pathways (like *Notch)* which are exquisitely dose sensitive and therefore susceptible to even subtle changes in levels of protein synthesis. Figure 3.29 depicts a series of wings in which certain mutant alleles of *lethal 2* enhance the weak N^{55e11} phenotype. Note that this effect is further enhanced when placed in a background in which HP1 dose has been reduced by one copy.

Another genetic assay that shows a similar effect involves the inversion *In(3L)C90* (Figure 3.30). This inversion breaks in euchromatin in region 62D on the left arm, and in heterochromatin in or near *lethal 1.* It is semi–lethal over deficiencies and mutant alleles of *lethal 1*, and exhibits a wing phenotype in combination with alleles of *lethal 2.* There are no phenotypes in combination with deficiencies or mutant alleles distal to *lethal 2.* The wing phenotype primarily affects the posterior margin, and is more strongly penetrant in females than in males, and this penetrance depends on the strength of the *lethal 2* allele assayed. When a single dose of HP1 is removed, this phenotype is enhanced.

RpL15 transcription is down regulated in _lethal 2_ P mutants.

Figure 3.31a shows quantitative data taken from a Northern analyses of *RpL15* expression in a *PΔ2/ PΔ2* versus a wildtype background. *RpL15* expression appears to be reduced in a *PΔ2/ PΔ2* (relative to *rp49* expression). Figure 3.31b shows the same experiment, but this time assaying for *Dbp80* expression: there is no significant change in DBP80 expression in a *lethal 2* mutant background relative to wildtype controls. Error bars represent standard deviations (95% confidence level).

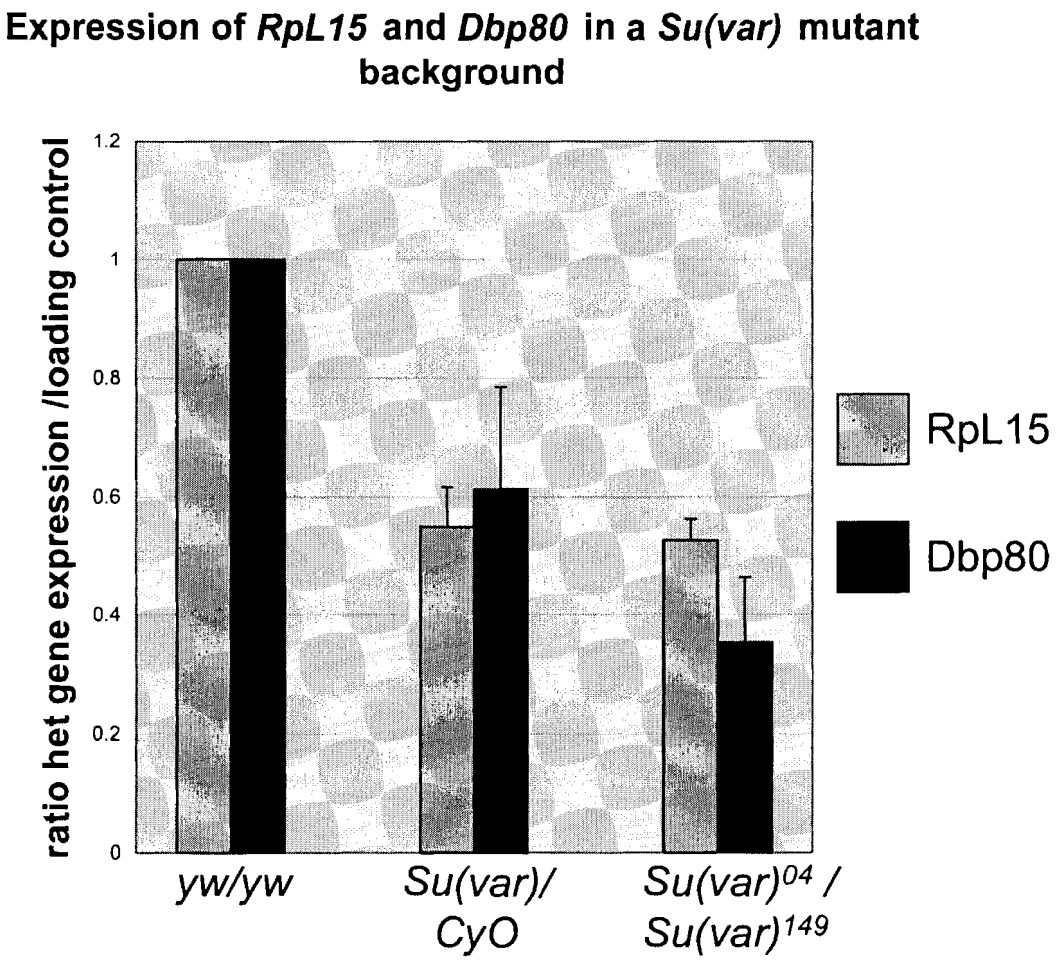


Figure 3.28: Expression of *Dbp80* and *RpL15* in a *Su(var)* mutant background

Quantitation of *Dbp80* and *RpL15* expression in a *Su(var)2-5* mutant background. The genotypes from which RNA was extracted are listed on the X axis. The Y axis represents expression levels, which were measured relative to loading controls, averaged from three separate experiments, and normalized to wildtype. Error bars represent standard error (absolute value of variance of each reading from the average).

Figure 3.29: N^{55e11} interaction assay

Photographs of wings from flies of various genotypes, showing that mutations in *lethal* 2 enhance the weak wing phenotype of the Notch allele N^{55e11}. This effect is further enhanced in a *Su(var)2-5*^{01} *(HP1)* mutant background, and the effect is specifically related to *lethal 2.*

(1): wild-type.

(2-4): expressivity of N^{55e11}.

(5-6): Balancer chromosomes. Note that N^{55e11} is interacting with *Ser* as it should.

(7-9, 19): N^{55e11} in a background containing deficiencies or mutant alleles that do not affect *lethal 2.*

(10-18) N^{55e11} in a background containing deficiencies or mutant alleles that *do* affect *lethal 2.* Severity of interaction increases with severity of *lethal 2* allele.

(20-23): control crosses: effect of N^{55e11} in a background containing a mutant *Su(var)2-5*^{01} allele (23), containing a mutant *Su(var)2-5*^{01} allele AND a lesion in a heterochromatic gene that is not *lethal 2* (20-21: 1-16-0=*SNAP25).*

Note the enhancement of the N^{55e11} phenotype in a mutant background containing a strong allele of *lethal 2* (17-18) which is further enhanced in a *Su(var)* mutant background (24-25).

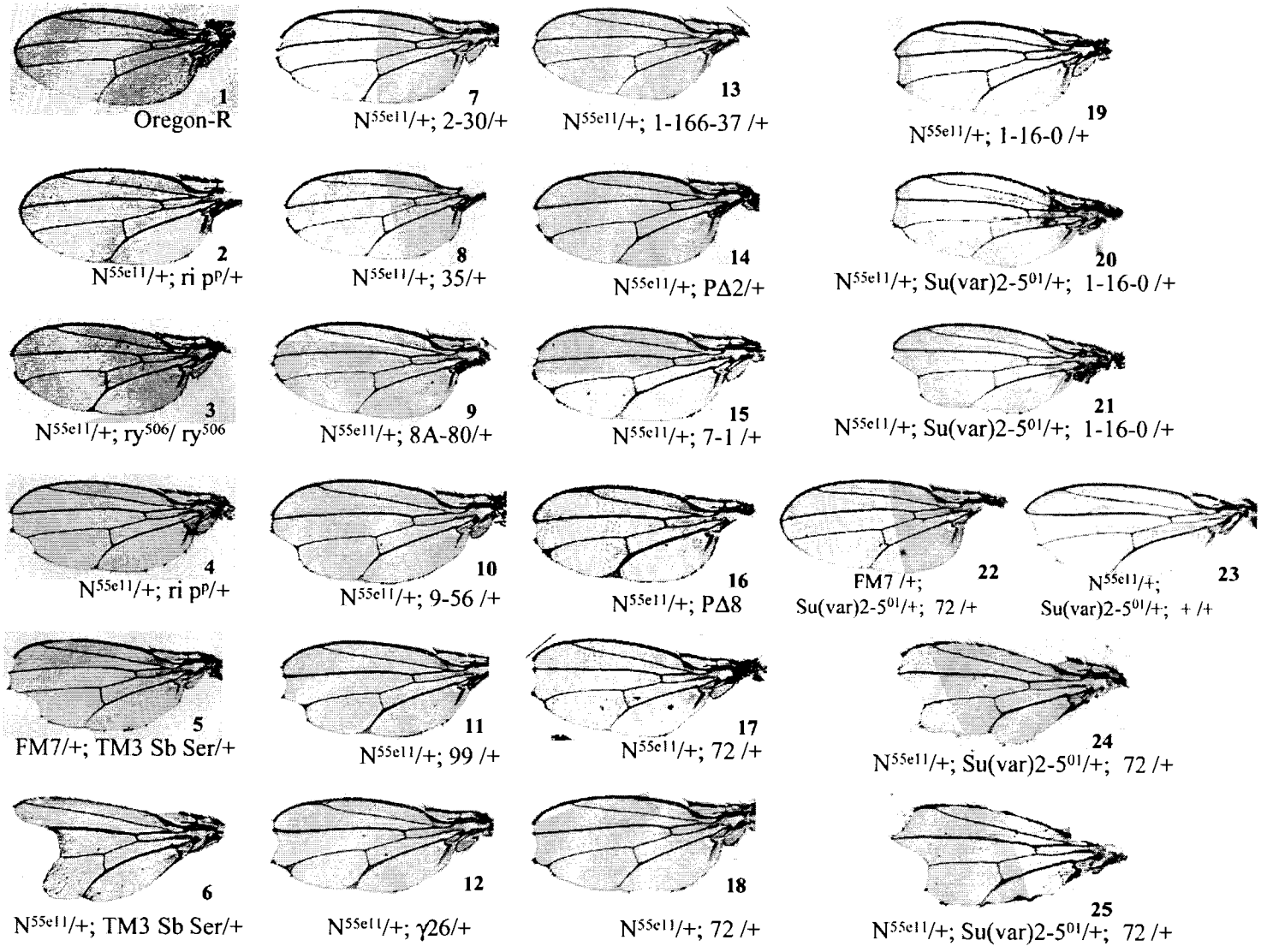

1
Oregon-R
7
N55e11/+; 2-30/+
13
N55e11/+; 1-166-37 /+
19
N55e11/+; 1-16-0 /+
2
N55e11/+; ri p/+
8
N55e11/+; 35/+
14
N55e11/+; PΔ2/+
20
N55e11/+; Su(var)2-501/+; 1-16-0 /+
3
N55e11/+; ry506/ ry506
9
N55e11/+; 8A-80/+
15
N55e11/+; 7-1 /+
21
N55e11/+; Su(var)2-501/+; 1-16-0 /+
4
N55e11/+; ri p/+
10
N55e11/+; 9-56 /+
16
N55e11/+; PΔ8
22
FM7 /+;
Su(var)2-501/+; 72 /+
23
N55e11/+;
Su(var)2-501/+; + /+
5
FM7/+; TM3 Sb Ser/+
11
N55e11/+; 99 /+
17
N55e11/+; 72 /+
24
N55e11/+; Su(var)2-501/+; 72 /+
6
N55e11/+; TM3 Sb Ser/+
12
N55e11/+; γ26/+
18
N55e11/+; 72 /+
25
N55e11/+; Su(var)2-501/+; 72 /+

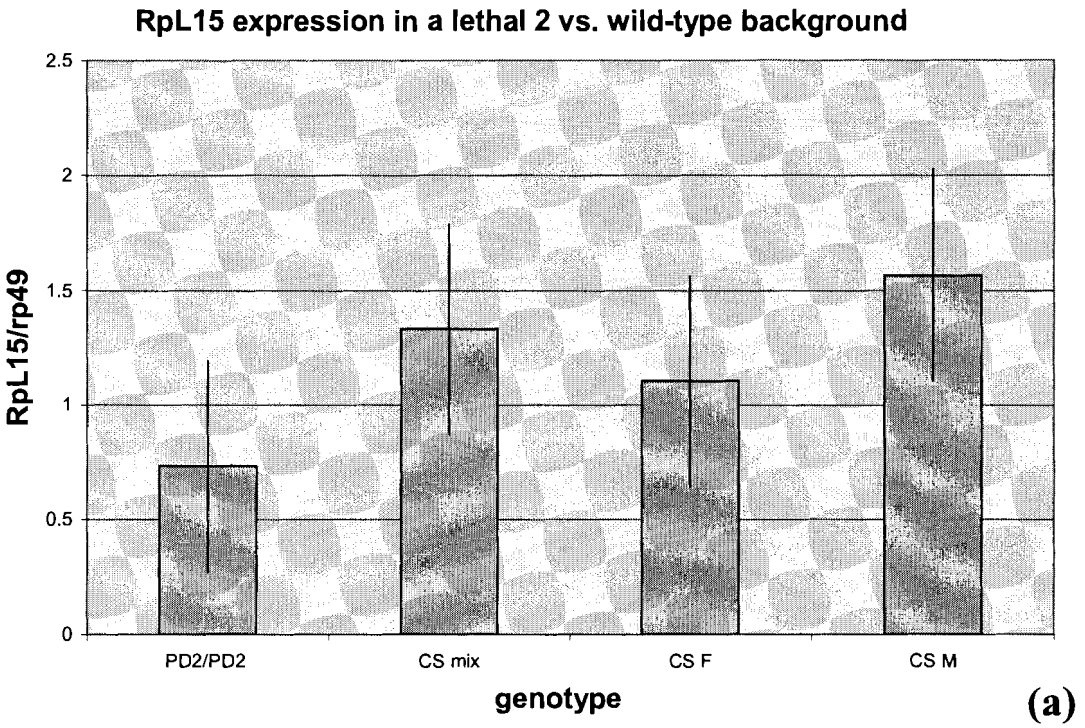
Figure 3.30: *In (3L)C90* interaction assay

Photographs of wings taken from flies carrying the *In(3L)C90* inversion and either of the *lethal 2* EMS alleles *72* or *1-166-37*. One breakpoint of *In(3L)C90* occurs in or near *lethal 1*, (resulting in a semi-lethal phenotype with alleles of this gene). The other breakpoint is in 62D, in the euchromatin of the left arm, effectively relocating *lethal 2* into euchromatin. Mutant alleles of *lethal 2* display the wing phenotype depicted in this figure, which does not occur in mutants or deficiencies distal to *lethal 2,* within the limits of the number of flies scored per cross (at least 150 flies). This phenotype is enhanced in a *Su(var)2-5*^{01} mutant background. M=male, F=Female, SV=*Su(var)2-5*^{01}, 1-16-0 is a mutant allele of *SNAP25*, a more distal 3L heterochromatic gene used as a non-*lethal 2* control.

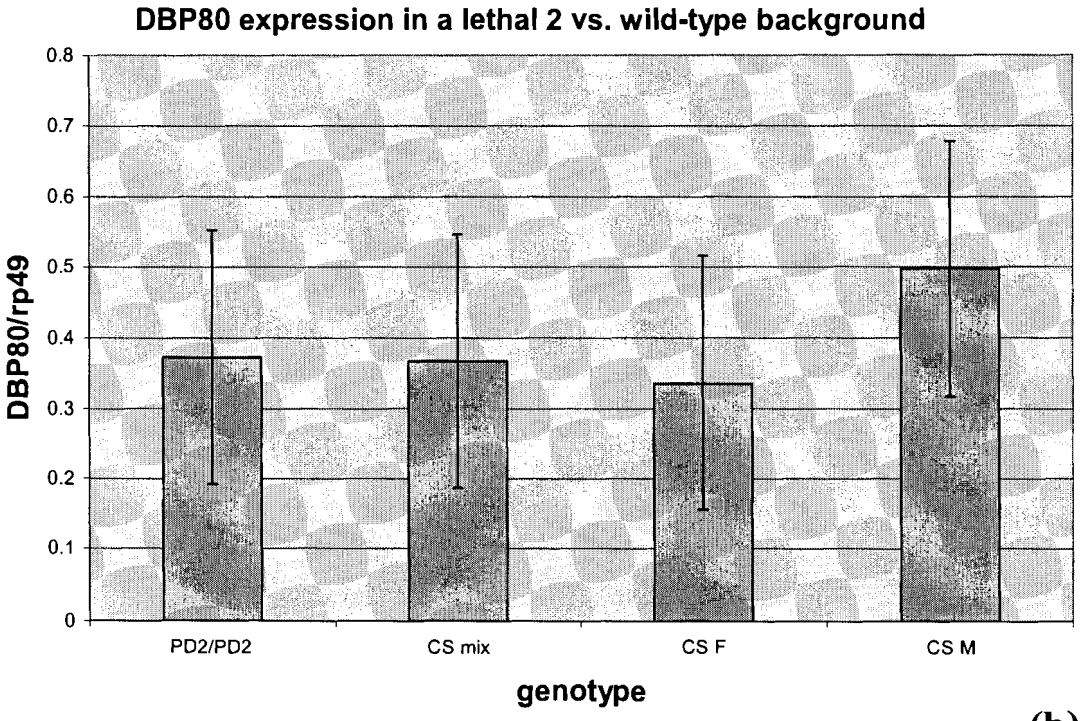
Genotype	Penetrance Male	Penetrance Female	Phenotype
1-166-37/In(3L)C90	0%	13%	 F
72/ In(3L)C90	13%	60%	 F
Su(var)2-5^{01}/+; In(3L)C90 /+	.01%	15%	 F
Su(var)2-5^{01}/+ 1-166-37/In(3L)C90	0%	81%	 F
Su(var)2-5^{01}/+ *72/ In(3L)C90*	100%	100%	 M  F
Su(var)2-5^{01}/+; 1-16-0/ In(3L)C90	6%	23%	 F

Figure 3.31: Statistical analysis of (a) *RpL15* and (b) *Dbp80* expression in a *PΔ2/PΔ2* background

Standard deviation (95% confidence) was calculated from the three wild-type controls. CS=Canton-S (wild-type strain); M=male; F=female, mix=both sexes.


RpL15 expression in a lethal 2 vs. wild-type background
RpL15/rp49
2.5
2
1.5
1
0.5
0
PD2/PD2
CS mix
CS F
CS M
genotype

(a)


DBP80 expression in a lethal 2 vs. wild-type background
DBP80/rp49
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
PD2/PD2
CS mix
CS F
CS M
genotype

(b)

DISCUSSION

The Minute syndrome and ribosomal protein genes

One of the most notable biological characteristics of *lethal 2* mutants is the classical *Minute* phenotype exhibited by certain transheterozygote combinations of alleles. In *Drosophila*, mutations in some ribosomal proteins have been shown to cause a dominant *Minute* phenotype, consisting of thin, weak bristles, rough eyes, wing vein and sex comb defects, delayed development and recessive lethality (Sinclair et al 1981). When first observed, the *Minute* phenotype was something of an enigma. It reveals itself approximately 68 times for every 5000 recessive lethals induced (Huang and Baker 1976). It is very often associated with haploinsufficiency, (associated with a cytologically defined deficiency; while acting as a recessive in the triplo-configuration M/+/+), and two different *Minutes* in the same fly are never additive, (i.e., epistatic: Schultz 1929). These observations suggest that *Minutes* encode specific products of similar function.

There are ~55 genetically identifiable *Minute* loci in the *Drosophila* genome (Lambertsson 1998) and Flybase (http://flybase.bio.indiana.edu:82/) currently lists 86 genes with homology to ribosomal proteins. Of these, at least 13 have been experimentally correlated with genetic *Minutes*, and a great many more map close to *Minute* loci (Lambertsson 1998). A few other genes have also been observed to mutate to a *Minute* or *Minute-like* phenotype. For example the rRNA genes (which are also components of the ribosome – Ritossa et al 1966), the *Suppressor of forked* gene, which is involved in transcript processing/stability, (Dudick et al. 1974), *eIF4a,* a DEAD-box helicase involved in translation initiation (Dorn et al. 1993), and *pitchoune,* another DEAD box helicase, possibly involved in cell growth and proliferation (Zaffran et al 1998). However, it is currently accepted that the majority of *Minutes* probably encode ribosomal proteins, and that the number of *Minutes* could formally equal the number of ribosomal protein genes, but many might not be seen due to their non-additive properties (i.e., a deficiency might remove more than one *Minute*).

Eukaryotic ribosomes are constructed from 4 rRNAs and about 80 ribosomal proteins. The ratio of protein to RNA increases from prokaryotes to eukaryotes (Koc et al. 2000), but the active site is still catalytic RNA, and the proteins are there primarily to maintain essential RNA structure (Aitchison and Rout 2000). Some ribosomal proteins appear to have more important roles in the ribosome than others, and it has already been demonstrated that a few of them have more than one function (Yacoub et al 1996 a,b, Wool, 1996). Genetically, different *Minutes* have phenotypes of varying severity, implying that although ribosomal protein genes are constitutively expressed, different genes may express at levels closer or farther from some threshold below which the phenotype will manifest itself. For example, there are RPS3 mutations that manifest the phenotype dominantly due to a 15% reduction in transcript abundance (Saebøe-Larsson et al.1998) while other less dose sensitive *Minute* loci will only display the phenotype when hemizygous (Saebøe-Larsson 1996) or even homozygous (Saebøe-Larsson 1998 and references therein).

Global (organism-wide) phenotypic manifestations due to defects in ribosomal protein function appear to be unique to *Drosophila* (although one report identifies a *Minute-like* phenotype in an *Arabidopsis* ribosomal protein gene mutation – Weijers et al. 2001). This may in part be due to the fact that with two exceptions (Brown et al. 1988, Yokokura et al. 1993), *Drosophila* appears only to have single copy ribosomal protein genes, while yeast, plants and humans have multiple copies, including pseudogenes (Zhang et al. 2002). Defects in human ribosomal protein genes have been implicated in a number of inherited disorders (for example Diamond Blackfan Anaemia (RpS19); Turner syndrome (RpS4), and Noonan syndrome (RpL6) – Zhang et al. 2002 and references therein), implying that mammalian ribosomes may exist in tissue/stage specific isoforms. However, no general *Minute* syndrome as such appears in organisms with multiple copy ribosomal protein genes. It is not clear why *Drosophila* has retained only single copies, but this may be explained in part by a growing body of evidence which suggests that the evolution of the *Drosophila* genome is marked by considerable DNA loss (Petrov 2002).

lethal 2 alleles are lesions in RpL15

All four *lethal 2* P alleles have tagged the ribosomal protein gene *RpL15* in exactly the same place, 18bp upstream of the pryrimidine tract within which transcription initiates (Figure 3.18). However, genetic (complementation data) and molecular evidence (P element orientation) suggests that at least two of these alleles may represent separate events. Nevertheless, when transheterozygous, the P alleles exhibit the classic Minute phenotype, most likely due to a reduction in *RpL15* transcription (Figure 3.31). Therefore the P alleles are behaving as hypomorphs. In combination with a null, the result is lethality.

PΔ2 appears to have undergone a natural rearrangement event during the course of removing the background P elements, which has left a markedly reduced P element fragment in the insertion site (Figure 3.22). This P element has lost almost all of its coding sequence, but retained the 31 base pair inverted repeats: i.e., an internal deletion has taken place, possibly accompanied by what appears to be an expansion of a sequence comprised solely of A's and T's. The *Drosophila* genome is between 4-12% middle repetitive sequences (Pimpinelli et al. 1995, Kaminker et al. 2002), the vast majority of which reside in heterochromatin. In this chromatin environment, they tend not only to be quiescent, but in fact their sequence integrity is degenerating, leading some to hypothesize that one function of heterochromatin is to serve as a "graveyard" for transposable elements which might otherwise wreak havoc (Dimitri 1999). Degeneration of transposable element sequences might include internal rearrangements and deletions, similar to the kind seen in *PΔ2.* It is also interesting to note the apparent expansion of low complexity sequence between the two remaining halves of the P element in this mutant. Expansion of highly repetitive sequence in heterochromatin has been suggested as a mechanism to reduce the density of origins of replication which appear in the coding regions of newly captured transposable elements (Csink and Henikoff 1998). One of the defining characteristics of heterochromatin is that it replicates later in the cell cycle than the rest of the genome (Ahmad and Henikoff 2001) – it is not yet clear why this is the case. Premature firing of replication origins in this region might disrupt later replication

timing, which in turn may be detrimental to the cell. It is possible that the changes observed in *PΔ2* comprise the earliest steps in this process.

The EMS *lethal 2* mutant *72* has a premature stop codon in the second exon of *RpL15*. *1-166-37* is defined by a mutation in a splicing consensus. In this regard, it is of interest to note that of the 263 *RpL15* EST's which have been isolated, 7 (~3%) possess an additional 303 nucleotides, corresponding to the preservation of the first intron. Since the *1-166-37* allele is lethal, it raises the interesting scenario that some level of ribosmal protein gene regulation may involve RNA processing, although to date there is no published research that has examined this possibility. The ribosome plays a fundamental role in controlling cell growth and development, and its constituent parts must be tightly and coordinately regulated. In prokaryotes, ribosomal proteins bind to their own mRNAs when conditions require they be down regulated (Mager 1988), and there is evidence that yeast ribosomal proteins can bind their own transcripts as well (Vilardell and Warner 1994). In eukaryotes, the polypyrimidine tract within which ribosomal protein gene transcription begins has been shown to be crucial for transcriptional (Hariharan and Perry 1990), and also translational regulation (Levy et al.1991). The complexity of gene structure and transcript processing increases with organismal complexity, so it might not be too surprising to observe RNA processing as another layer of ribosomal protein gene regulation. It is possible that a small percentage of this longer *RpL15* message plays a part in the regulation of this gene. If this intron cannot be spliced (as may be the case for *1-166-37)* the result may be lethality due to an excess of the longer form. A similar situation may exist for one of the *lethal 2* P alleles: when total RNA taken from *PΔ2/PΔ2* escapers is probed with a *RpL15* cDNA, a second, higher molecular weight band is always observed, which is consistent in size with the longer *RpL15* transcript (Figure 3.32).

Finally, a variety of molecular lesions in *RpL15* result from imprecise excisions of *lethal 2* P alleles. Combined with the genetic analysis, these molecular data identify *lethal 2* as *RpL15*.

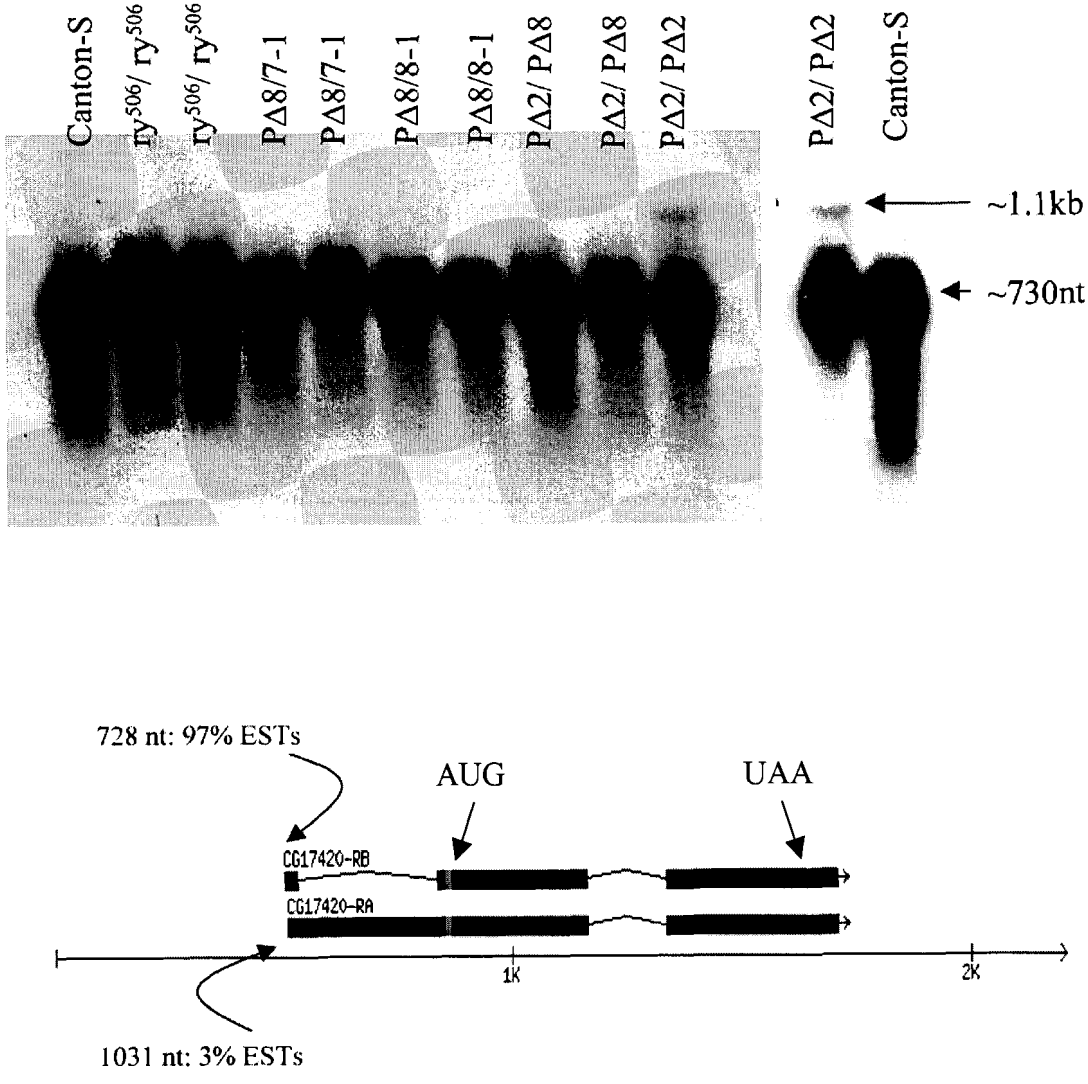


Figure 3.32: Alternative splicing in *RpL15*

Top: Total RNA taken from *lethal 2* P mutant escapers, and probed with the *RpL15* cDNA. Note the presence of a higher molecular weight band in only those genotypes associated with *PΔ 2*. Bottom: Schematic of *RpL15* alternative transcripts and their relative abundance in EST libraries. Figure courtesy of Flybase.

Difficulties with transgenic rescue

However compelling the data linking *RpL15* to *lethal 2*, germline rescue of *lethal 2* with *RpL15* transgenes would constitute the strongest evidence. Neither genomic nor cDNA constructs have rescued the lethality of *lethal 2*, in spite of all attempts to address potential artifactual problems (vector type, sequence error, positional effects etc). Therefore it remains to outline potential biological explanations for why rescue of lethality has been unsuccessful.

As has been mentioned already, Flybase lists more than 80 gene sequences with homology to ribosomal proteins. In the last 20 years, approximately 30 papers have been written about ribosomal protein genes, and in only 7 cases has an attempt at germline rescue been reported (Kongsuwan et al. 1985 (rp49), Qian et al. 1988 (rpA1), Kay et al.1988 (rp21), Voelker et al. 1989 (RpL36), Schmidt et al. 1996 (RpL19), Reynaud et al. 1997 (RpS3A) and Torok et al. 1999 (RpS21)). In 4 out of 7 cases, rescue was successful, and in 3 out 7 cases, lethality was rescued. In all successful cases, the rescue construct consisted of a genomic piece of DNA containing the gene cloned into an uninducible construct.

Based on the description of ribosomal protein genetics presented above, it is possible that *RpL15* may express to a specific level not reached by any transgenes in the absence of wildtype product. Rescue of ribosomal protein genes has been reported to be difficult for this reason (Dorer et al. 1991, Lambertsson 1998), but also because deficiencies used in rescue-crossing schemes might remove more than one gene. This potential problem was addressed by using transheterozygous combinations of specific alleles in addition to deficiencies. But as experiments using these combinations were also unsuccessful, this is not a likely explanation. While it is not possible to determine the actual size of the deficiencies that remove *lethal 2*, all combinations used in the transgenic crossing schemes demonstrate an larval stage L1 lethal phase identical to the *lethal 2* EMS allele *72* (except for some crosses which involved *Df(3L)K2,* which has an embryonic lethal phase). It is possible that to rescue the lethality associated with lesions in *RpL15* requires

a specific level of expression that may be technically impossible to reproduce. Only viability can be rescued, and only in the presence of constitutively driven cDNA transgenes.

Finally, it is theoretically possible that the *RpL15* gene has an additional and essential function in the germline. A P-induced mutation in *RpL15* was isolated as a female sterile with egg chamber defects (called "*ziti*"). The single P insertion is 52 bp away from the beginning of the coding region (therefore within the first intron), and neither it nor excisions derived therefrom display a *Minute* phenotype. The gene was cloned from DNA flanking the insertion site, and the associated transcript is reported to be ~900 bp, and it is apparently upregulated in females. None of these data have been published, and were cited in Lambertsson (1998) presumably as a personal communication. There is also a reference to conference poster that presumably outlined this work (Dej and Spradling,1997: "Heterochromatic ribosomal protein gene is specifically required during oogenesis to maintain nurse cell chromosome organization". A. Dros. Res. Conf. 38: 8B).

The data reported for *ziti* (AKA *RpL15*) are quite different from the results of the work here presented. The transcript size has been independently confirmed to be closer to 700bp (BDGP EST project), and RpL15 expression does not show a significant difference in transcript levels between males and females (Figure 3.31a). However, individuals that are heteroallelic for *lethal 2* P alleles are sterile, and can show a marked sex skew in favour of males, suggesting that oogenesis might be more sensitive to defects in ribosomal protein synthesis than spermatogenesis. This has in fact been observed with other ribosomal proteins. For example, *string of pearls,* which encodes RpS2, was isolated as a female sterile, and so-named due to the shape of the defective egg chambers (Cramton and Laski 1994), and RpS3A appears to be essential for oogenesis (Reynaud et al. 1997). In general, low fertility in heterozygous *Minute* females is often considered a part of the *Minute* syndrome (Lambertsson 1998). But some ribosomal proteins may play more important roles in this area than others; *RpL15* might be one example, which might in turn complicate a transgenic rescue scheme.

Is Dbp80 an essential gene?

The only other gene which appears to be located near *RpL15* is *Dbp80*, which encodes an RNA helicase of the DExH box family. This is a very large and diverse family of RNA helicases (more than 63 in the *Drosophila* genome-Lasko 2000) encoding products which are involved in transcript processing and export, (for example, *vasa*), and translation initiation (for example eIF4A). DBP80 belongs to a conserved subset of these proteins, which in yeast performs an essential function in mRNA export. Mutations in this single copy gene in yeast interact lethally with mutations in specific nucleoporins, and with RAN-GAP (the activating protein for the protein export factor GTPase), and exhibit a mRNA export-defective phenotype (Snay-Hodge 1998). GFP fusion proteins indicate that yeast DBP5 localizes to the cytoplasmic face of the nuclear envelope (Tseng et al.1998), consistent with its genetic interactions with nuclear pore complex components. Protein transport is unaffected in *Dbp5* mutant cells, implicating DBP5 solely in a mRNA export pathway (*ibid.*).

The mRNA export pathway involves several components including DBP5 which show a high degree of conservation. Both the human and yeast homologs of this gene interact with the same specific nucleoporins, and the human protein shows a similar nuclear-envelope delimited localization in cell culture (Schmitt et al.1999). Finally, directed mutagenesis of human DBP5 causes a mRNA export phenotype when microinjected into *Xenopus* oocytes (*ibid.*).

Dbp80 corresponds to no known lethal complementation group in proximal 3L heterochromatin in *Drosophila,* yet it is clearly located in this region. Either it has been missed in the three mutagenesis screens which have been carried out to date (Chapter Two), or it is not essential. There is some evidence supporting the latter contention: in purifying from yeast more components from the same nuclear pore complex fraction which contained DBP5, two additional RNA helicases were isolated: RNA Helicase A and UAP56. When the *Drosophila* homologues of these 2 helicases (*Maleless* and *Hel* respectively) and DBP80 were tested in Schneider cell culture (by depletion with RNAi), only UAP56 (HEL) exhibited the expected mRNA export defect – DBP80 had no effect

(Gatfield et al. 2001). HEL also associated with spliced mRNAs carrying the exon junction complex (involved in nuclear proofing of RNA processing and nonsense mediated decay of transcripts bearing premature stop codons – *ibid.*). *Hel* is an essential gene which was initially identified as an enhancer of position effect variegation (Eberl et al. 1997). HEL protein is localized to the nucleus, and associates with chromosomes, and from the experiments described above, also appears to be involved in mRNA export. In addition, recent RNAi experiments in *C.elegans* (www.wormbase.org/db/seq/rnai?name=JA%3AT07D4.4;class=RNAi) suggest that the homologue of *Dbp80* in this organism is not essential. However, it should be noted that descriptive studies in the dipteran *Chironomus tentans* showed the DBP80 homolog (CtDBP5) appears to bind mRNP particles co-transcriptionally, and accompany them to the nuclear pore and into the cytoplasm (Zhao et al. 2002).

The DBP80 protein is highly conserved, and contains a 6 amino acid residue insertion which unambiguously places it in the DBP5 class (Figure 3.8). When the DBP80 protein is used as a query in a BLASTP search against *Drosophila* predicted proteins, 29 other RNA helicases are displayed. The genes coding for twelve of these can be mutated to generate a variety of phenotypes including recessive morphology defects, sterility or lethality (http://flybase.bio.indiana.edu:82/). There is no reported experimental data on the other 17 genes. As has already been mentioned, RNA helicases comprise a very large family, but an essential role for *Dbp80* in *Drosophila* cannot be ruled out. This is in contrast to the ribosomal protein family, which in spite of its size, encodes products with specific non redundant functions. It is perhaps not surprising that one can mutate to lethality, while the other apparently does not.

The effect of reduced Su(var) dose

Su(var)2-5 encodes Heterochromatin protein 1, which is known to bind and crosslink modified histones, causing increased compaction in the chromatin fibre and transcriptional repression (see Chapter One). The expression of the well-characterized heterochromatic genes *light* and *rolled* appears to be compromised in a genetic

background in which HP1 dose has been reduced (Lu et al. 2000), which suggests that heterochromatic genes have evolved a dependence upon trans-acting factors that normally silence gene expression. This is in keeping with the paradoxical nature of heterochromatic gene expression manifested in other ways as well: for example they will variegate when transferred into euchromatin (please see Chapter One for a fuller description of position effect variegation), and yet they reside and function in a transcriptionally repressive environment. It is possible they have adapted, by evolving a dependence upon at least one of heterochromatin's primary constituents.

The results presented here suggest that both *Dbp80* and *RpL15* show some dependence on HP1 dose (Figure 3.28). The authors of the original study showing *light* and *rolled* dependence upon HP1 have extended this study by looking for genes elsewhere in the genome that are repressed by HP1 (Hwang et al. 2001). They reported three examples all mapping to euchromatin, albeit in a region with which HP1 strongly associates (region 31 on 2L). Overall their results suggest that heterochromatic genes can be activated by HP1, whereas certain euchromatic genes are repressed. This work and the previous study involving *light* and *rolled* are often cited as evidence that transcriptional activation by HP1 is a diagnostic feature of heterochromatic genes. My results do not contradict this assertion, but the sample size is too small for such a generalization to be made, and it may not be correct in any case. A very recent publication (Piacentini et al. 2003) has shown that HP1 can in fact associate and positively regulate sites of intense gene activity in euchromatin (polytene chromosome puffs), in particular, those loci which encode the heat shock proteins (87A, 87C, 95D). Formaldehyde cross-linked chromatin immunoprecipitation (X-ChIP) experiments using primers to either the promoter or coding regions of the *Hsp70* gene (which maps to 87A and C) show that after heat shock induction, HP1 protein is enriched in the coding region, and not the promoter of this gene. Moreover, this enrichment appears to depend upon the presence of RNA, and an intact chromo domain in HP1. Thus, HP1 may in fact be acting as a regulator of transcription by controlling the stability of the transcript (elongation, processing etc.) rather than by inducing or repressing gene expression *per se.* In this regard it is useful to recall that the chromo domains in proteins from the Male-specific lethal (MSL) complex

do in fact function as RNA binding modules (by interacting with the non-coding roX RNAs: Akhtar et al. 2000). The MSL complex assembles on numerous sites on the male X chromosome during the process of dosage compensation.

The mechanism by which HP1 associates with gene promoters and/or coding regions is not clear. HP1 does not itself bind DNA, but recognizes histone H3 which has been methylated on lysine 9 by another chromatin associated protein called Su(var)3-9. To date, the only functional connection between HP1 and promoter regulation comes from research in mammalian cell culture, in which HP1 and Su(var)3-9 both interact with the transcriptional repressor Retinoblastoma (Rb) (Nielson et al. 2001). Rb plays a critical role in repressing the activity of a transcription factor called E2F, which binds the promoters of many genes required for cell proliferation. It represses E2F apparently by recruiting proteins like Su(var)3-9 and HP1. Therefore it is unlikely that there are any sequence-specific elements in promoters that directly relate to HP1 involvement, and this is borne out by the contrasting nature of the two promoter types studied in the present work. *Dbp80* shows moderate levels of expression, and may be developmentally regulated (Figure 3.11,12), whereas *RpL15*, like all other ribosomal protein genes characterized to date, is very highly and constitutively expressed (data not shown – larval, pupal and adult northerns used for quantitation).

It is important to note from Figure 3.28 that dose dependence upon HP1 is not clear: the data are ambiguous with respect to the genotype which is heterozygous for a mutation in HP1. However, two genetic assays show that removing a single dose of HP1 can impair *lethal 2* function. N^{55e11} is a weak allele of the wing morphogenesis mutant *Notch*. This gene plays a central role in an ancient and conserved signal transduction pathway (Baron et al. 2002), and like most signal transducers it is exquisitely dose sensitive and therefore very susceptible to changes in the level of protein synthesis. It thus serves as an excellent sensitized background in which to assay the genetics of *RpL15*. The wing margin and vein defect in N^{55e11} is enhanced in a *lethal 2* mutant background, and further enhanced when a single copy of the gene encoding HP1 is removed (Figure 3.29).

A similar situation exists with the inversion *In(3L)C90*. This inversion has one breakpoint in euchromatin (62D and the other in heterochromatin, somewhere in or near *lethal 1*. This means that *lethal 2* is effectively translocated into euchromatin. When the EMS alleles *72* or *1-166-37* are made heterozygous with this inversion, a posterior wing margin defect results, which is enhanced when a single copy of HP1 is removed (Figure 3.30). This effect is not observed for any lesions distal to *lethal 2* – i.e., the effect appears to be polar, and may represent a variegation phenotype. Why the posterior wing margin in particular is affected is not clear - the phenotype does not resemble the wing defects associated with *lethal 2's Minute* phenotype. However, the euchromatic breakpoint of this inversion does remove ~60 genes (assayed by polytene *in situ* – Lindsley and Hardy, 1992). It is possible that one or more of these genes is sensitive to reductions in protein synthesis, and is therefore interacting with *lethal 2*, in a similar manner to N^{55e11}.

It is perhaps surprising that *RpL15* exhibits a sensitivity to both HP1 dose and position effect, since ribosomal protein genes are likely to have evolved an ability to express in almost any chromatin environment, (widely scattered, highly expressed essential housekeeping genes). In fact yeast ribosomal protein genes have been shown to carry insulators in their 5' regions containing binding recognition sites for the Rap1 activating/repressing protein (Bi and Broach 1999, Yu et al. 2003). However, there is no homologue for Rap1p in the *Drosophila* genome, and sequences matching the yeast consensus are found in neither the first intron nor upstream sequences of *RpL15*.

RpL15 transcription is reduced in lethal 2 mutants

The data in Figure 3.31 were collected from two Northern blots, probed first with the *RpL15* cDNA, then the *Dbp80* cDNA and finally with a loading control (rp49). Average values were calculated from the paired readings, and a standard deviation was determined using three wild type controls (total RNA from females, males and from a mixed population). Although there is considerable variability, *RpL15* expression does appear to be compromised in a *lethal 2* mutant background, while *Dbp80* is unaffected.

Interestingly, when *PΔ2/ PΔ2* total RNA is probed with an *RpL15* cDNA, a second, higher molecular weight band reproducibly appears (Figure 3.32). This band likely corresponds to a longer splicing variant of *RpL15* which appears in ~3% of the ESTs that have been isolated for this gene. (This percentage of longer EST's is calculated from a pool of full-length transcripts; a number of truncated/incomplete EST's are also present. Note also that the proportion of long to short EST's will vary as more EST's are added to the database.)

CONCLUSIONS

Analysis of four P alleles and two EMS alleles of the proximal 3L heterochromatic gene *lethal 2* has identified these as lesions in the ribosomal protein gene *RpL15*, thus identifying *lethal2* as encoding this large subunit ribosomal protein. There must be extensive selection pressure to keep this gene small in a region where genes tend to grow very large, due to the expansion of repeats in their introns. *RpL15* has repetitive DNA in both introns, but has maintained a size comparable to other ribosomal protein genes, which are distributed throughout the genome. This is perhaps not surprising for a gene that must be transcribed at high levels all the time, a characteristic of other small genes that are required constitutively or during times of stress (Singh et al. 2000. Castillo-Davis 2002). It will be interesting to learn how other similar housekeeping/stress genes have evolved to function in heterochromatin.

Reductions in *RpL15* transcription lead to a classical *Minute* phenotype which is modest in heterozygotes, but strong in transheterozygous combinations of weak alleles. Lethal combinations of mutant *lethal 2* alleles were not rescued by a transgene construct bearing either a cDNA or genomic clone, but this is likely due to the difficulties of generating the precise amounts of product required for ensuring proper ribosomal function. It is also possible that *RpL15* has other biological activities, which cannot be separated from its essential house-keeping function. Viability was rescued, which provides more evidence confirming the molecular identity of *lethal 2* as *RpL15*.

RpL15 resides approximately 10,000 base pairs away from another gene: *Dbp80.* This gene is very large and shows a considerable degree of sequence polymorphism in non-coding regions, and it is also embedded in a repetitive environment. Transcription from both genes appears to be impaired in a genetic background in which HP1 dose has been reduced, which is consistent with the results obtained for the well-characterized heterochromatic genes *light* and *rolled.*

Dbp80 does not correspond to any of the lethal complementation groups discovered in proximal 3L heterochromatin by Marchant and Holm (1988b). It is not yet clear whether *Dbp80* is essential, but its membership in a large and functionally overlapping family (de la Cruz et al.1999), and preliminary biological analysis using RNAi in cell culture (Gatfield et al. 2001), support the notion that it is not essential. In addition, I will show in the next chapter that this gene has undergone dramatic rearrangements over a relatively short evolutionary period, a process which may be less likely to occur if this gene were essential.

CHAPTER FOUR: Cloning and characterizing *RpL15* and *Dbp80* in *Drosophila virilis*

INTRODUCTION

Comparative analysis of homologous genes between related species has proved to be a useful tool in a number of different areas, founded on the premise that conservation of sequence implies biological constraint. For instance, it has provided unique insights into the evolution of gene regulation (Colot et al.1988, Audibert and Simonelig 1998, Wittkopp et al. 2002), in addition to revealing the kinds of rearrangements that take place as chromosomes evolve. Chromosome evolution has in fact been suggested to play a role in speciation, where significant rearrangements might lead to reduced rates of recombination, mispairing of homologues and subsequent reproductive isolation. A case in point is made by a comparison of human and chimpanzee genomes, which share 98.7% of their genetic material (Fujiyama et al. 2002), but which differ in their chromosomal configurations, including a number of fusions and pericentric inversions (Wildman 2002).

In undertaking a comparative species approach, it is important to select a species pair that reflect evolutionary lineages that are sufficiently separated in time, so that any DNA identities at selectively neutral positions will have been lost due to mutation. This would imply that sequences that have remained conserved are under selection pressure, imposed by the requirement for the proteins they encode. Estimates of evolutionary distance often vary, depending on whether morphological or molecular criteria are employed, and even within molecular analysis, estimates will vary depending on whether quickly evolving (transcription factors, introns) or slowly evolving (enzymes) templates are used. In some cases, even geology can play a role, for instance, when deriving a phylogenetic tree of *Drosophilids* which include the Hawaiian species, timed to the origin of the Hawaiian islands (Russo et al.1995).

Drosophilidae is a very diverse and widely spread Dipteran family, comprising almost 3,000 species partitioned into 61 genera (Russo et al. 1995). The genus *Drosophila* contains 14 subgenera, including *Sophophora* to which *Drosophila melanogaster* belongs, and *Drosophila* (same name as genus) which includes *Drosophila virilis* (see Figure 4.1). These two subgenera are separated by 40-60 million years of evolution,

(depending on the method of phylogenetic analysis used) which is sufficient time for conserved sequences to imply biological constraint, given the rate of nucleotide substitution per site per year per *Drosophila* lineage of 1 X 10^{-8} (Russo et al.1995).

Drosophilids exhibit a very uniform karyotype (chromosomal configuration), shared by even more distantly related dipterans like mosquitoes (Figure 4.2). Almost all of the members of this group possess 6 chromosome arms, which have been given letter designations indicating homology based on chromosomal hybridization or conservation of gene loci, where possible (Muller 1940). More recently the scheme has been extended to include dipterans that were never the subject of genetic analysis (Figure 4.2, Bolshakov et al. 2002). The reason(s) for this apparently strict karyotypic constraint is not known, but it provides an excellent model for studying chromosomal rearrangements in evolution. *Drosophila virilis* possesses what is considered to be the ancestral configuration for the genus *Drosophila*, determined from chromosomal phylogenies established by cytological analysis, consisting of five acrocentric chromosome arms plus a tiny dot sixth chromosome (Patterson and Stone, 1952). Fusion events during the course of the *D. melanogaster* lineage have produced two metacentric chromosomes, a single acrocentric X chromosome and a tiny dot fourth chromosome. In males of both species, the Y chromosome appears to be entirely heterochromatic.

There are other notable differences between the two genomes. *D. virilis* possesses a genome of ~330Mb (0.34-0.38 pg per haploid genome) which is about twice the size of *D.melanogaster*. Almost half of the *D.virilis* genome is heterochromatic, compared to a third for *D.melanogaster* (Pimpinelli et al. 1976). In addition, *D.virilis* has ~36% more euchromatin than *D.melanogaster (*Moriyama et al. 1998). Finally, the polytene chromosomes of *D.virilis* exhibit a banding pattern that suggests each band contains 4.5 times the DNA found in an average *D.melanogaster* band, or about 100kb (Vieira et al.1997a). Since there is no genome project for *D.virilis,* it is not known what these parameters mean in terms of gene density/arrangement for this species.

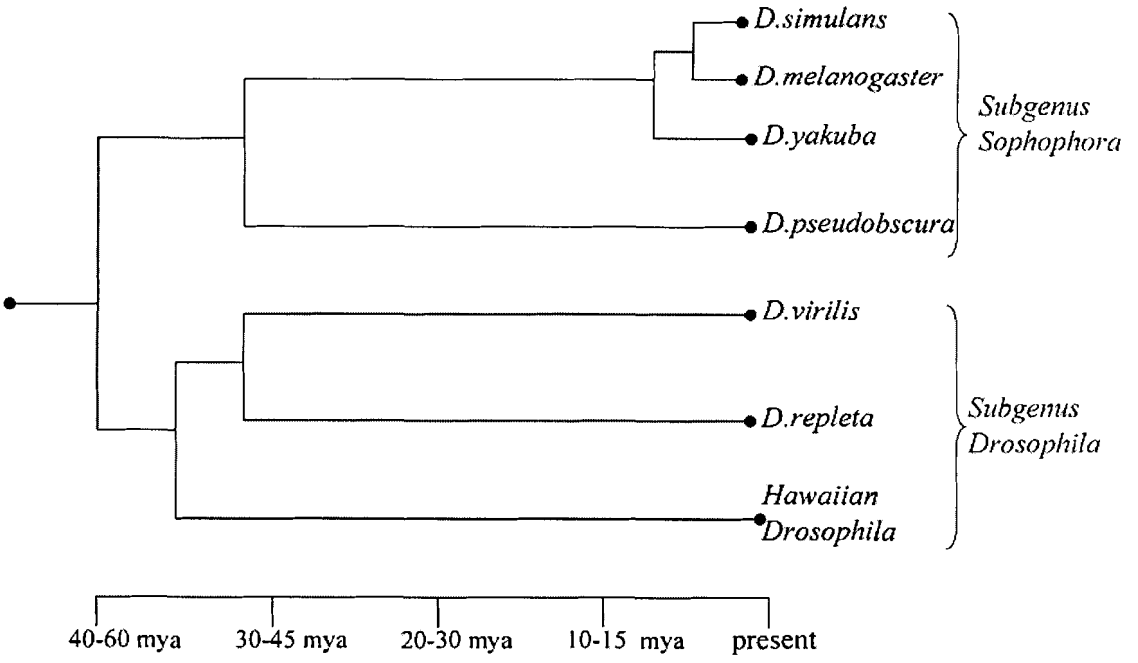
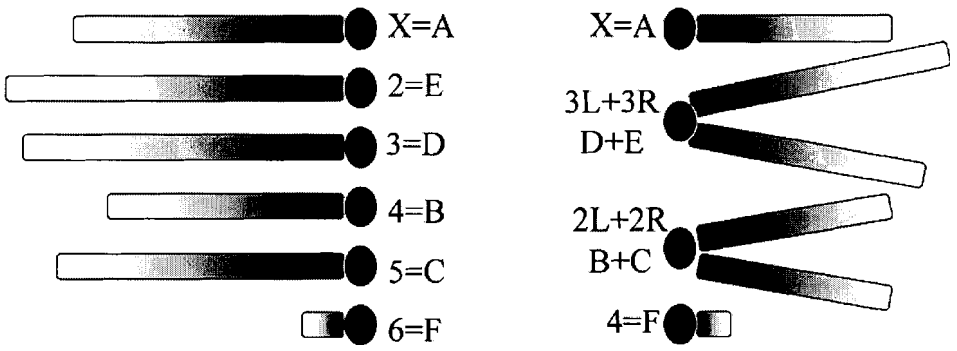
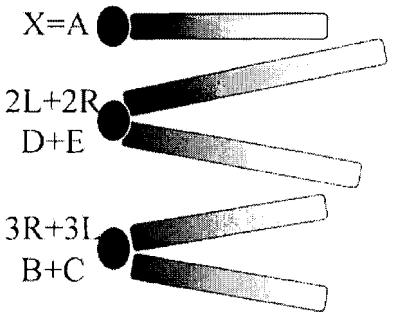


Figure 4.1: *Drosophila* phylogeny
Branch point times are depicted as ranges, to reflect estimates from the literature which are based on different methods of measuring evolutionary distance.



Drosophila virilis *Drosophila melanogaster*



Anopheles gambiae

Figure 4.2: Dipteran karyotypes
Chromosomal configurations for three dipterans: *Drosophila virilis, Drosophila melanogaster,* and *Anopheles gambiae.*

In sum, *Drosophila virilis* and *Drosophila melanogaster* are sufficiently distant evolutionarily for a comparative analysis, and they possess very similar chromosomal configurations which simplifies the karyotypic portion of that analysis. They differ in important ways, with respect to genome size and heterochromatic content, which may have implications for a direct sequence comparison.

Dbp80 and *RpL15* are closely linked genes deep within the centric heterochromatin of *Drosophila melanogaster.* A fuller understanding of how these genes came to reside and function in such a repressive context can benefit from a cross species analysis, particularly if at least one of these genes was once euchromatic. To this end the homologues of both genes were cloned from *Drosophila virilis* cDNA and genomic libraries. The homologues were then cytologically mapped to *D. virilis* chromosomes, where it was discovered that while *vRpL15* is still likely to be in heterochromatin, *vDbp80* is in fact a euchromatic gene. A preliminary genomic analysis was undertaken to identify any sequence-specific characteristics, which might explain their contrasting chromatin contexts.

RESULTS

4.1. Cloning and characterization of *vDbp80*

Two *D. virilis Dbp80* cDNA's were isolated from an embryonic cDNA plasmid library, using the *D.melanogaster* cDNA as a probe in a low stringency screen (see Chapter Two). The two *D.virilis* cDNA's differ only slightly in length at the 5' end, and the longer of the two was used to screen a *D.virilis* λEMBL3 genomic library at high stringency (see Chapter Two). DNA from two overlapping phage was digested with a combination of restriction enzymes and the fragments were subcloned into pBluescript, and subsequently sequenced. Approximately 18,000 bp of genomic DNA containing the *D.virilis Dbp80* homologue was mapped, and this data is summarized in Figure 4.3.

vDbp80 is moderately highly expressed, producing a single transcript of 1640 nt, comparable in size to the *D.melanogaster* gene (Figure 4.4a), and appears to be single-copy (Figure 4.4b: genomic DNA cut with Eco RI and probed with the cDNA produces a banding pattern consistent with one internal EcoRI site). The *vDbp80* gene encodes a protein of 465 amino acids, and is highly conserved across taxa (Table 4.1). vDBP80 possesses the same six residue insertion characteristic of members of the DBP5 family (see Figure 3.6, 3.8)

There are two principal differences between the *Dbp80* homologues. Firstly, polytene *in situ* hybridization (Figure 4.5) indicates that *vDbp80* is in fact a single copy euchromatic gene, mapping to region 35F according to the *D.virilis* chromosome maps developed by Horst Kress (1993). This places *vDbp80* approximately one third of the way down the euchromatic arm (from the centromere) of chromosome 3, which according to Muller's assignments is element D, and therefore the homologous arm of *D.melanogaster* 3L (Figure 4.2). Secondly, *vDbp80* has a dramatically different gene organization in comparison to its *D.melanogaster* homologue, and this is depicted in Figure 4.6 The *D.virilis* homologue occupies about $1/90^{th}$ the genomic territory, consisting of two large exons (564bp and 836bp) separated by a small (53bp) intron. This single intron has in fact been conserved in *D.melanogaster* both in position, phase and approximate size – in *D.melanogaster* it is 60 bp long and separates exons 5 and 6. *vDbp80* has conventional polyadenylation and transcription initiation signals, but no apparent TATA box (Figure 4.7)

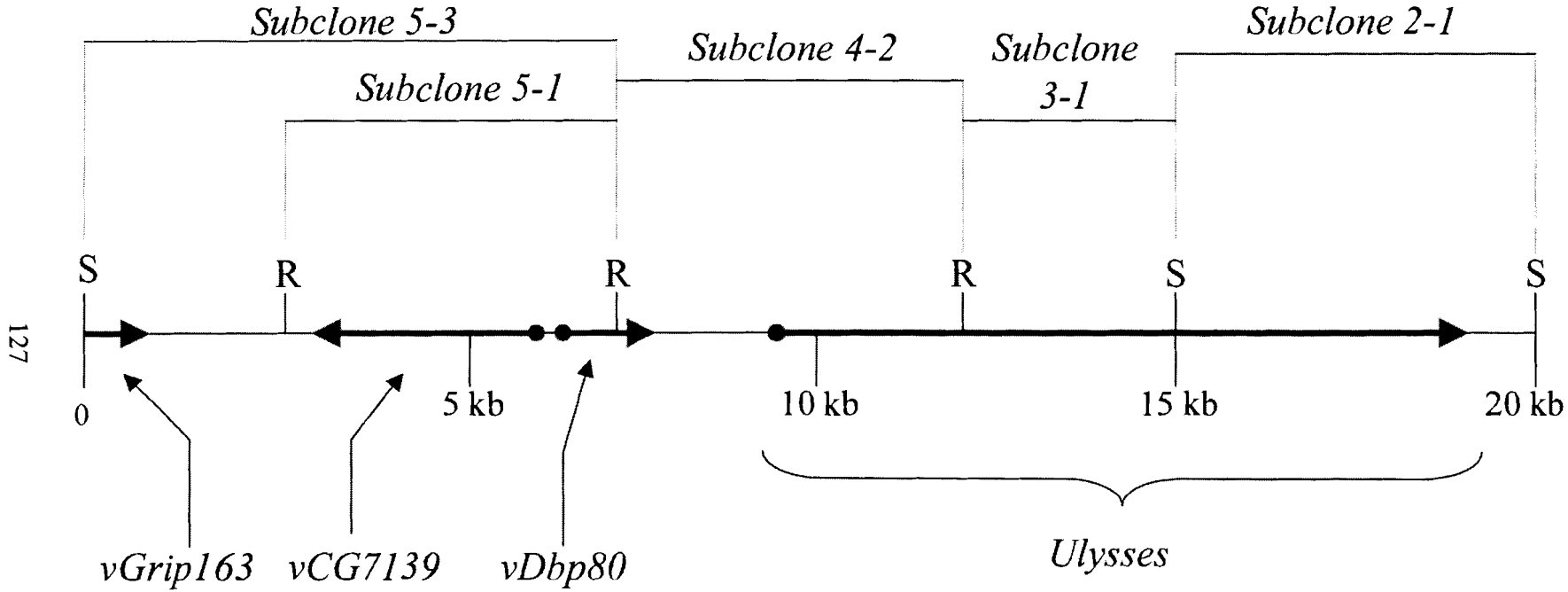
The genomic environment of *vDbp80* may provide some insight into the nature of the rearrangements that have produced these contrasting genomic organizations. A Southern analysis using genomic probes is presented in Figure 4.8. Upstream of *vDbp80* the genomic region shows two clean bands, while the downstream region appears to be repetitive. This is reflected by the sequence analysis, which reveals two genes upstream of *vDbp80* (one complete and one partial), and a large transposable element immediately downstream. The first of the upstream genes resides less than 500 bp away from *vDbp80,* and is transcribed in the opposite direction. I have called it *vCG7139,* based on the close

BLAST homology (Table 4.2). In *D. melanogaster,* this gene possesses a 5'UTR of 340 bp, and encodes alternatively spliced transcripts of 3653bp and 2810bp each, which when translated result in proteins thought to be involved in DNA repair (Eisen 1998). *vCG7139* appears to possess a very similar gene organization to its *D.melanogaster* homologue, and covers slightly less genomic territory. It is also possible it shares regulatory regions with *vDbp80*, since the genes are separated by less than a kilobase of DNA. Further upstream of *vCG7139* there appear to be at least two 3' exons from a gene with close homology to *Grip163* (Table 4.2) – which in *D.melanogaster* encodes an essential component of the centrosome (Gunawardane et al. 2000).

This chromosomal region in *D.virilis* has in fact received a lot of attention recently, with a number of other genes in the vicinity also characterized. In Figure 4.9 I have reproduced a comparative map of the region in both species, complete with cytological coordinates and the positions of other homologous genes (based on Kress 1993). Local regions of microsynteny can be observed, as well as a variety of extensive rearrangements within the same chromosomal element. This kind of organization has also been observed on the X chromosome (Vieira et al. 1997a) and between other *Drosophila* species (Ranz et al. 1997, Vieira et al.1997b, Ranz et al. 2001).

860 bp downstream of *vDbp80* there is a large retrotransposon called *Ulysses*, which measures 10.6kb, and possesses long terminal repeats of 2.1kb each (Scheinker et al.1990). It is inserted in the same orientation as *vDbp80*, and based on findings concerning the dispersed and polymorphic genomic locations of this transposable element in the *D.virilis* species group (Zelentsova et al. 1999) is very likely to be active. It has been elegantly demonstrated that P-element induced recombination can produce chromosomal rearrangements by well-characterized mechanisms (Gray et al.1996, Preston and Engels 1996). The potential role of *Ulysses* in chromosome evolution will be discussed later in this chapter.

Figure 4.3: *vDbp80* genomic map
Restriction maps and assembly of *vDbp80* genomic subclones from phage λEMBL3 isolates. R=EcoRI; S=Sal I restriction enzyme sites used in subcloning.


Subclone 5-3
Subclone 5-1
Subclone 4-2
Subclone 3-1
Subclone 2-1
S
R
R
R
S
S
0
5 kb
10 kb
15 kb
20 kb
vGrip163
vCG7139
vDbp80
Ulysses

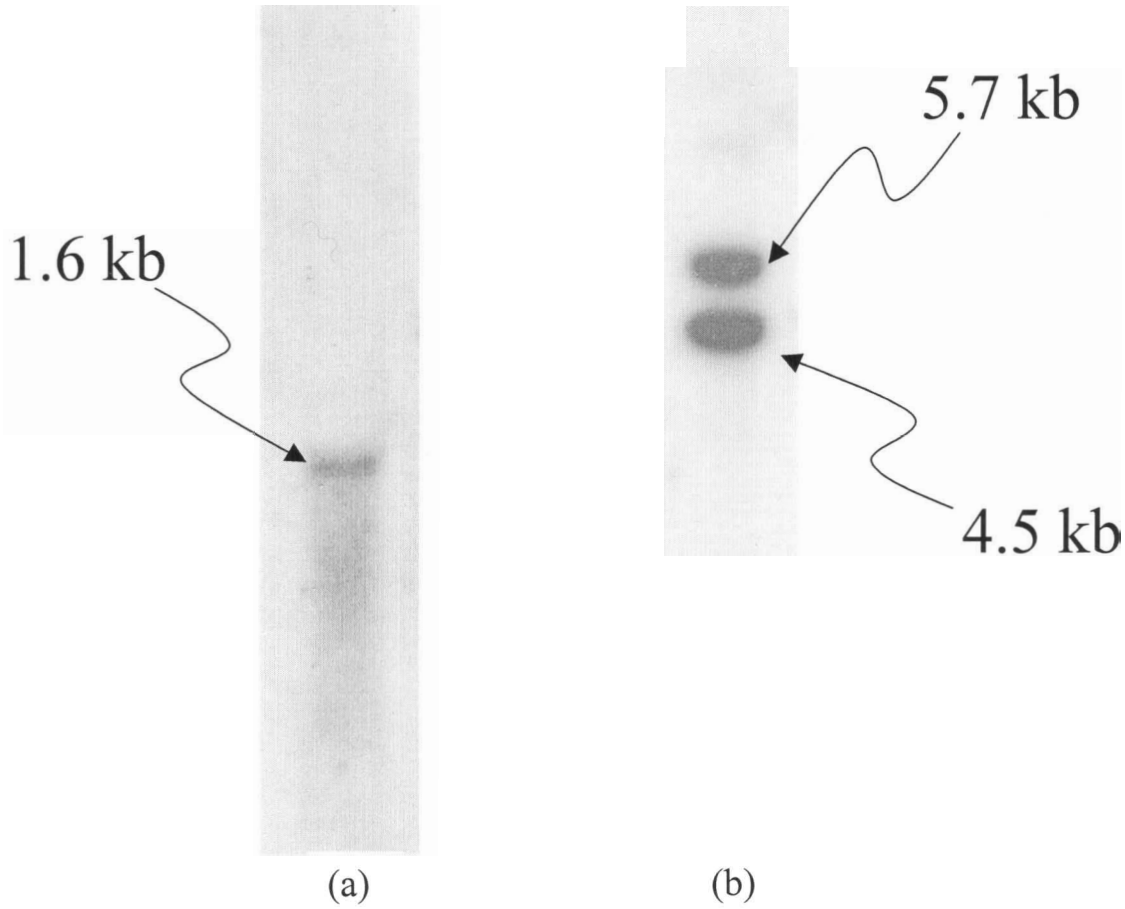


Figure 4.4 (a): Northern and (b): Southern analysis of *vDbp80*
(a): Northern analysis of *vDbp80*: total adult RNA probed with the v*Dbp80* cDNA
(b): Southern analysis: genomic DNA cut with EcoRI and probed with the v*Dbp80* cDNA. Note that the cDNA sequence for this gene has an internal EcoRI site.

Organism	%similarity	%identity
Drosophila melanogaster	90%	84%
Chironomus tentans	82%	71%
Anopheles gambiae	81%	69%
Mus musculus	79%	60%
Homo sapiens	78%	60%
Danio rerio	77%	58%
Xenopus laevis	77%	60%
Dictyostelium discoideum	70%	49%
Schizosaccharomyces pombe	64%	45%
Saccharomyces cerevisiae	68%	49%
Neurospora crassa	61%	40%

Table 4.1: vDBP80 protein homologies across different taxa

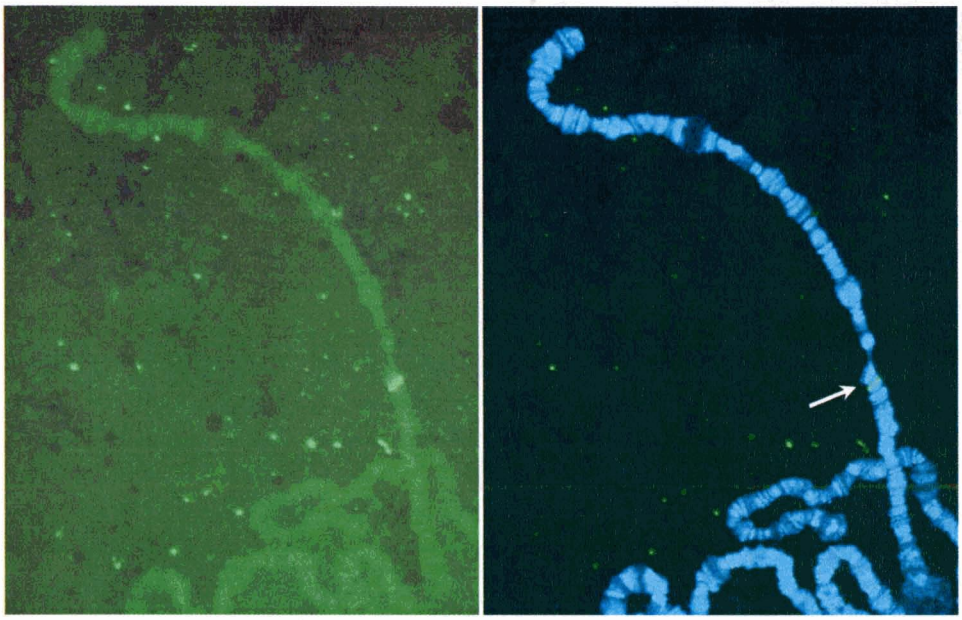


Figure 4.5: Chromosomal location of *vDbp80*
in situ hybridization of a *vDbp80* cDNA probe to *Drosophila virilis* polytene chromosomes. The signal is located on Muller's element D (chromosome 3) at approximately position 35F, according to the maps of Kress (1993).

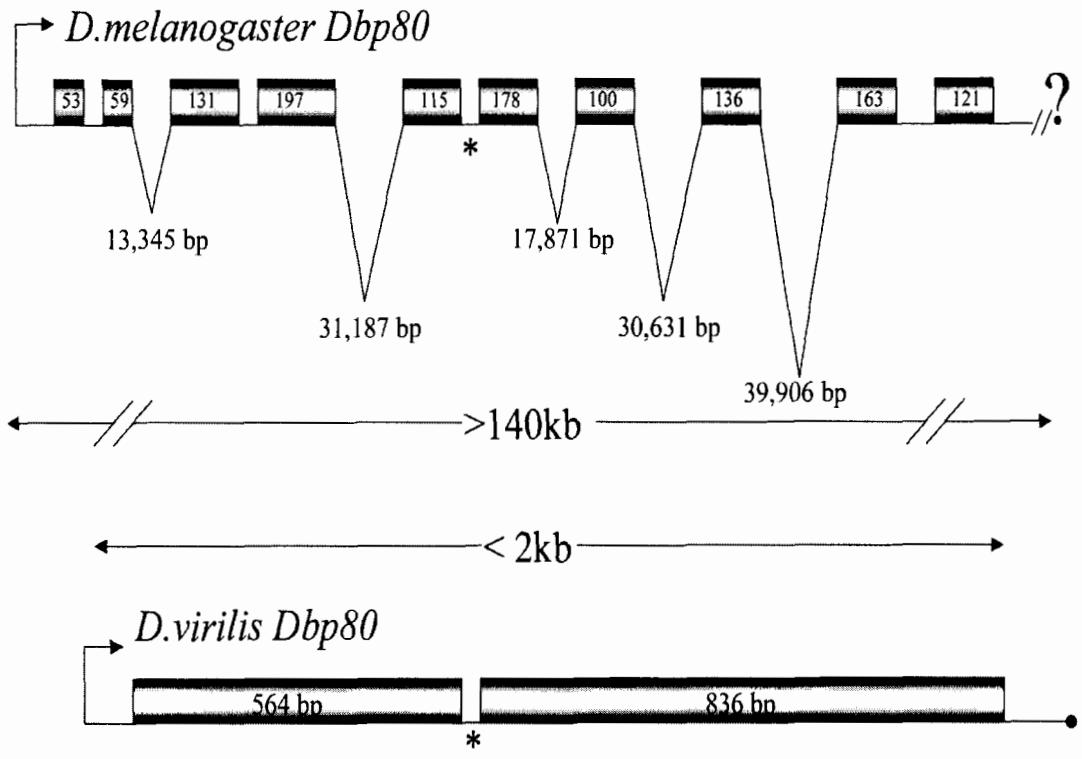
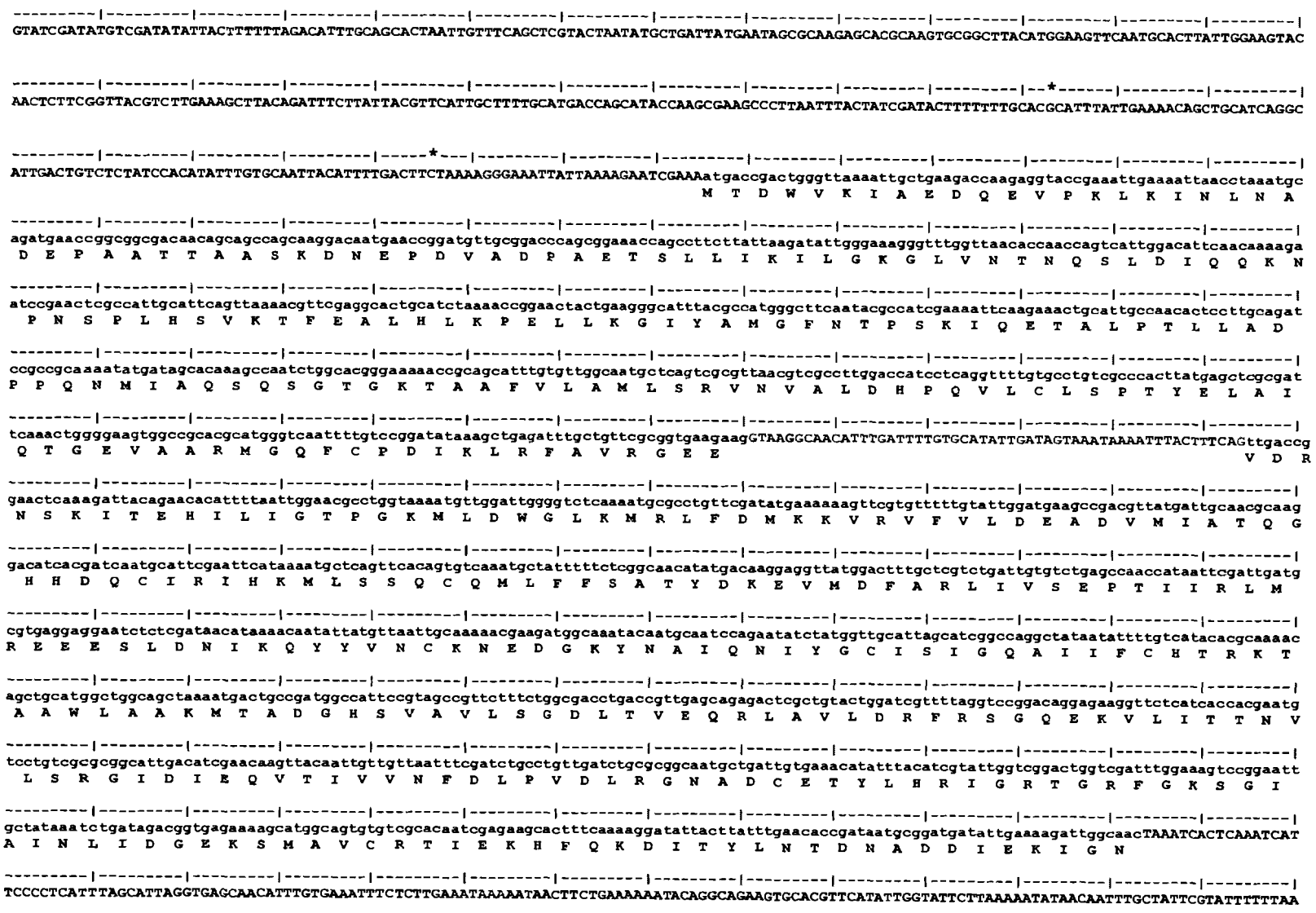


Figure 4.6: Comparative*Dbp80* gene organization
Comparison of *Dbp80* gene organization in *Drosophila melanogaster* and *Drosophila virilis*. The boxes represent exons, and the lines connecting them the introns. The diagrammes are not to scale. The asterisk marks the position of an ancient intron.

Figure 4.7: *vDbp80* gene structure
Schematic of *vDBP80* gene sequence organization showing relevant signals. (*) mark the starting positions of the two cDNA's isolated from an embryonic library.



GTATCGATATGTCGATATATTACTTTTTTAGACATTTGCAGCACTAATTGTTTCAGCTCGTACTAATATGCTGATTATGAATAGCGCAAGAGCACGCAAGTGCGGCTTACATGGAAGTTCAATGCACTTATTGGAAGTAC

AACTCTTCGGTTACGTCTTGAAAGCTTACAGATTTCTTATTACGTTCATTGCTTTTGCATGACCAGCATACCAAGCGAAGCCCTTAATTTACTATCGATACTTTTTTTGCACGCATTTATTGAAAACAGCTGCATCAGGC

ATTGACTGTCTCTATCCACATATTTGTGCAATTACATTTTGACTTCTAAAAGGGAAATTATTAAAAGAATCGAAAatgaccgactgggttaaaattgctgaagaccaagaggtaccgaaattgaaaattaacctaaatgc

M T D W V K I A E D Q E V P K L K I N L N A

agatgaaccggcggcgacaacagcagccagcaaggacaatgaaccggatgttgcggacccagcggaaaccagccttcttattaagatattgggaaagggtttggttaacaccaaccagtcattggacattcaacaaaaga

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

atccgaactcgccattgcattcagttaaaacgttcgaggcactgcatctaaaaccggaactactgaagggcatttacgccatgggcttcaatacgccatcgaaaattcaagaaactgcattgccaacactccttgcagat

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

ccgccgcaaaatatgatagcacaaagccaatctggcacgggaaaaaccgcagcatttgtgttggcaatgctcagtcgcgttaacgtcgccttggaccatcctcaggttttgtgcctgtcgcccacttatgagctcgcgat

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

tcaaactggggaagtggccgcacgcatgggtcaattttgtccggatataaagctgagatttgctgttcgcggtgaagaagGTAAGGCAACATTTGATTTTGTGCATATTGATAGTAAATAAAATTTACTTTCAGttgaccg

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

gaactcaaagattacagaacacattttaattggaacgcctggtaaaatgttggattggggtctcaaaatgcgcctgttcgatatgaaaaaagttcgtgtttttgtattggatgaagccgacgttatgattgcaacgcaag

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

gacatcacgatcaatgcattcgaattcataaaatgctcagttcacagtgtcaaatgctatttttctcggcaacatatgacaaggaggttatggactttgctcgtctgattgtgtctgagccaaccataattcgattgatg

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

cgtgaggaggaatctctcgataacataaaacaatattatgttaattgcaaaaacgaagatggcaaatacaatgcaatccagaatatctatggttgcattagcatcggccaggctataatattttgtcatacacgcaaaac

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

agctgcatggctggcagctaaaatgactgccgatggccattccgtagccgttctttctggcgacctgaccgttgagcagagactcgctgtactggatcgttttaggtccggacaggagaaggttctcatcaccacgaaatg

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

tcctgtcgcgcggcattgacatcgaacaagttacaattgttgttaatttcgatctgcctgttgatctgcgcggcaatgctgattgtgaaacatatttacatcgtattggtcggactggtcgatttggaaagtccggaatt

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

gctataaatctgatagacggtgagaaaagcatggcagtgtgtcgcacaatcgagaagcactttcaaaaggatattacttatttgaacaccgataatgcggatgatattgaaaagattggcaacTAAATCACTCAAATCAT

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

TCCCCTCATTTAGCATTAGGTGAGCAACATTTGTGAAATTTCTCTTGAAATAAAAATAACTTCTGAAAAAATACAGGCAGAAGTGCACGTTCATATTGGTATTCTTAAAAATATAACAATTTGCTATTCGTATTTTTTAA

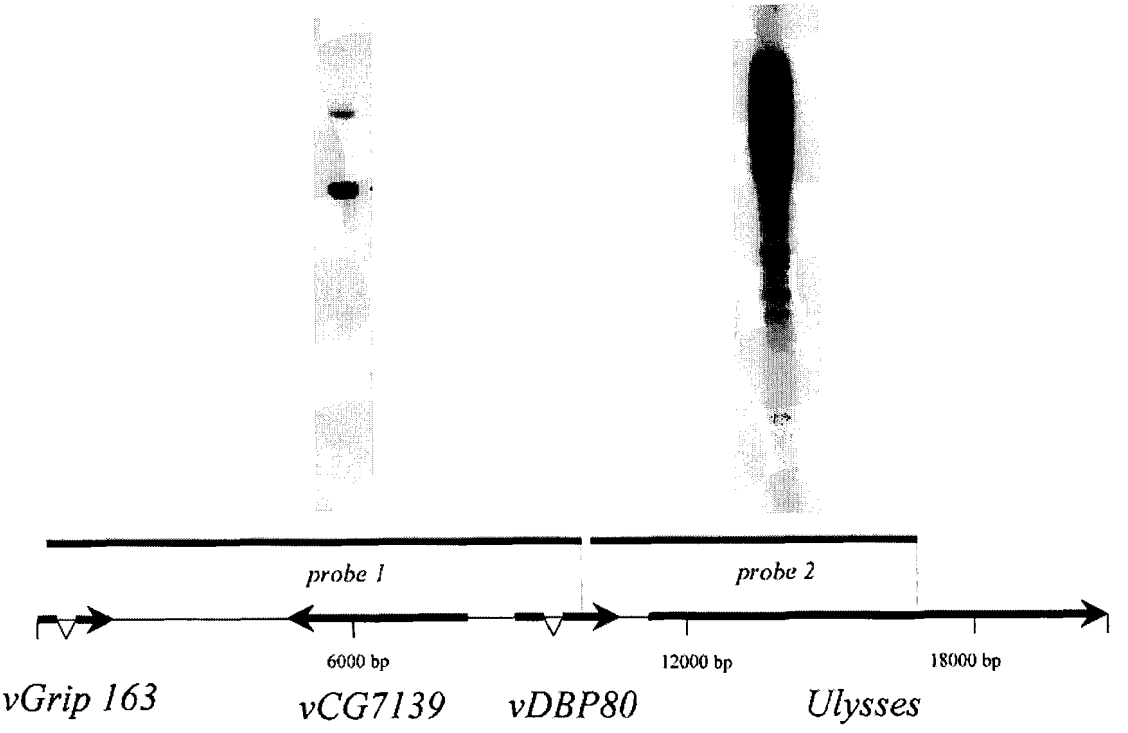


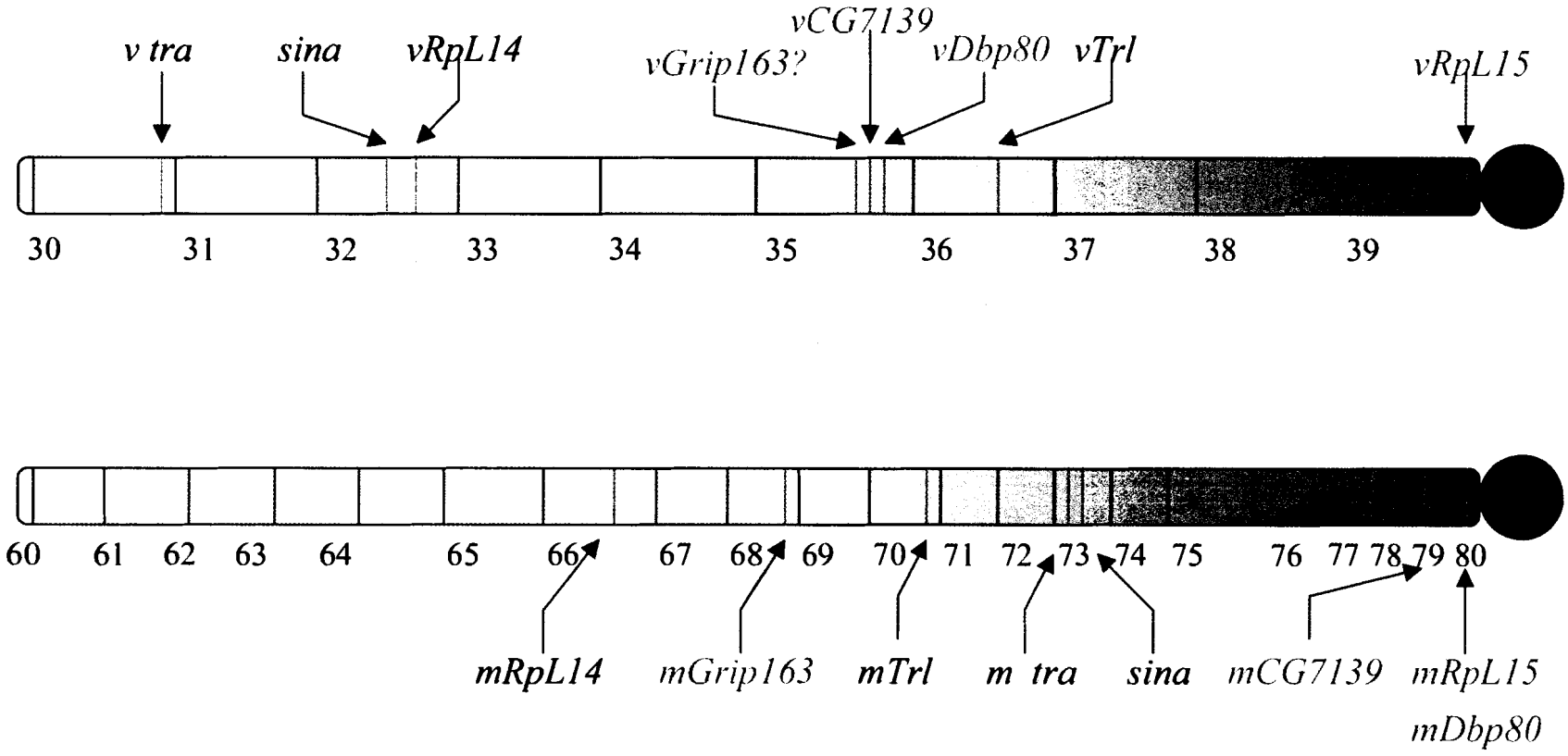
Figure 4.8: *vDbp80* upstream/downstream genomic Southerns
Genomic DNA was cut with EcoRI, transferred to nylon membrane and hybridized with genomic probes containing either the upstream or downstream regions of *vDbp80.*

PROGRAM	QUERY	SEQUENCES PRODUCING HIGH-SCORING SEGMENT PAIRS	SMALLEST SUM PROBABILITY
BLASTX against *Drosophila melanogaster* predicted proteins	vDBPgenomic05 (*Drosophila virilis* genomic contig)	CG7139-PA	2.4e-274
		CG7139-PB	9.6e-208
		CG17023 (DBP80)	7.1e-117
		CG5688 (Grip163)	2.0e-23
BLASTX against *Drosophila melanogaster* predicted proteins	vRpL15genomic05 (*Drosophila virilis* genomic contig)	CG40199 (RpL15)	5.5e-105

Table 4.2: BLASTX evidence for *D.virilis* genes upstream of *vDbp80*
BLASTX search using the *D. virilis* genomic contigs for *vDbp80* and *vRpL15* as queries against *Drosophila melanogaster* predicted proteins. Smallest sum probability reflects the likelihood that the alignment is due to chance.

Figure 4.9: Comparative cytological map of Muller's element D from *Drosophila melanogaster* and *Drosophila virilis*

This map has been adapted with permission from Horst Kress (Kress 1993). Genes in red have been mapped by the present study. Others are as follows: *transformer (tra)*:O'Neil and Belote 1992; *seven in absentia (sina)*: Neufeld et al. 1991; *RpL14*: Lyamouri et al. 2002; *trithorax-like* (*trl)*: Lintermann et al. 1998.


v tra
sina
vRpL14
vGrip163?
vCG7139
vDbp80
vTrl
vRpL15
30
31
32
33
34
35
36
37
38
39
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
mRpL14
mGrip163
mTrl
m tra
sina
mCG7139
mRpL15
mDbp80

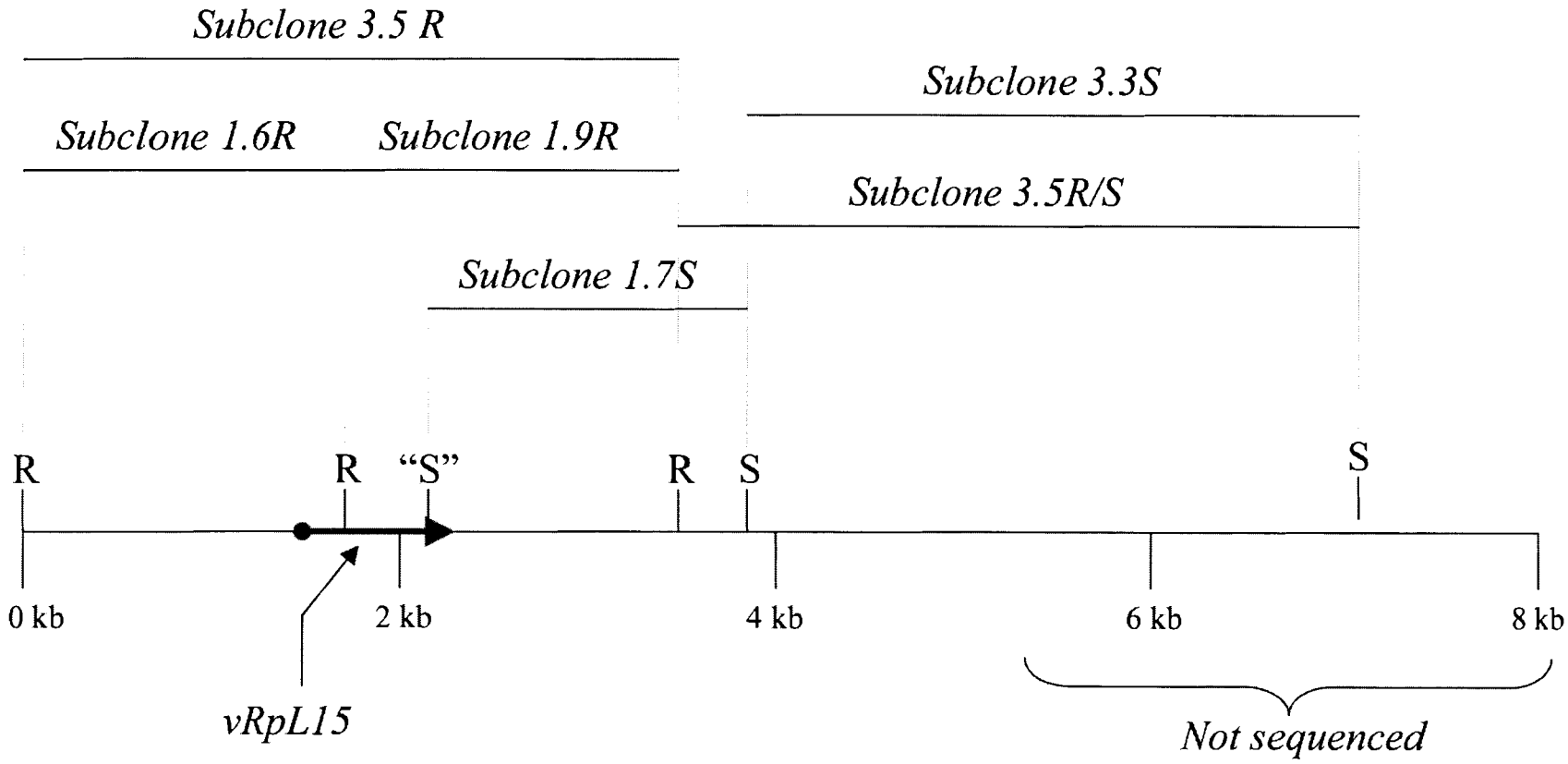
4.2. Cloning and characterization of *vRpL15*

The same cDNA and genomic libraries described in the previous section and in Chapter Two were used to isolate the *D.virilis* homolog of *RpL15*. A single transcript was obtained and used to isolate two overlapping phage. Subclones from these phage proved far more difficult to assemble and map (Figure 4.10). A total of 4,940 bp containing *vRpL15* have been sequenced. *vRpL15* is highly expressed as would be expected for a ribosomal protein gene, producing a single transcript of 700 bp (Figure 4.11a). When translated, a protein of 204 amino acids should result, exactly the same size as the *D.melanogaster* homologue. *vRpL15* is very highly conserved (Table 4.3), and based on Southern analysis is present as a single copy gene (Figure 4.11b: genomic DNA cut with Eco RI and probed with the cDNA produces a banding pattern consistent with one internal EcoRI site). However, genomic probes in a Southern analysis indicate *vRpL15* resides in a repetitive environment (Figure 4.12)

In contrast to *Dbp80*, the *D.melanogaster* and *D.virilis* homologues of *RpL15* share both chromosomal location and gene organization. Both genes consist of 3 exons separated by 2 introns, all of which are comparable in size (Figure 4.13). In both cases, transcription appears to proceed from a polypyrimidine tract located right next to the first exon, in keeping with what is understood about ribosomal protein promoter structure (see discussion and also Chapter Three). In addition, both homologues reside in a repetitive environment, which for *mRpL15* has been confirmed to be heterochromatic (see Chapter Three). Polytene *in situ* hybridization using *vRpL15* cDNA probes produces a reproducible signal in the *D.virilis* chromocentre (Figure 4.14) – diagnostic of a physical location in centric heterochromatin. It is not clear from these *in situs* whether *vRpL15* is still in Muller's element *D* (chromosome 3 in *D.virilis,* 3L in *D.melanogaster*). In the figure shown, this element is in fact missing, so either *vRpL15* is located very close to the centromere within element D, or it may be near the centromere on another chromosome. This has implications for the kinds of fusion events which ultimately must have taken place to result in the metacentric configuration in *D.melanogaster*, which will be described more fully in the discussion portion of this chapter.

Figure 4.10: *vRpL15* genomic map

Restriction maps and assembly of *vRpL15* genomic subclones from phage λEMBL3 isolates. R=EcoRI; S=Sal I restriction enzyme sites used in subcloning. "S" refers to an edge clone - the site results from the way in which the library was made (incomplete Sau3A digestion) and does not exist in the natural sequence.


Subclone 3.5 R
Subclone 3.3S
Subclone 1.6R
Subclone 1.9R
Subclone 3.5R/S
Subclone 1.7S
R
R
"S"
R
S
S
0 kb
2 kb
4 kb
6 kb
8 kb
vRpL15
Not sequenced

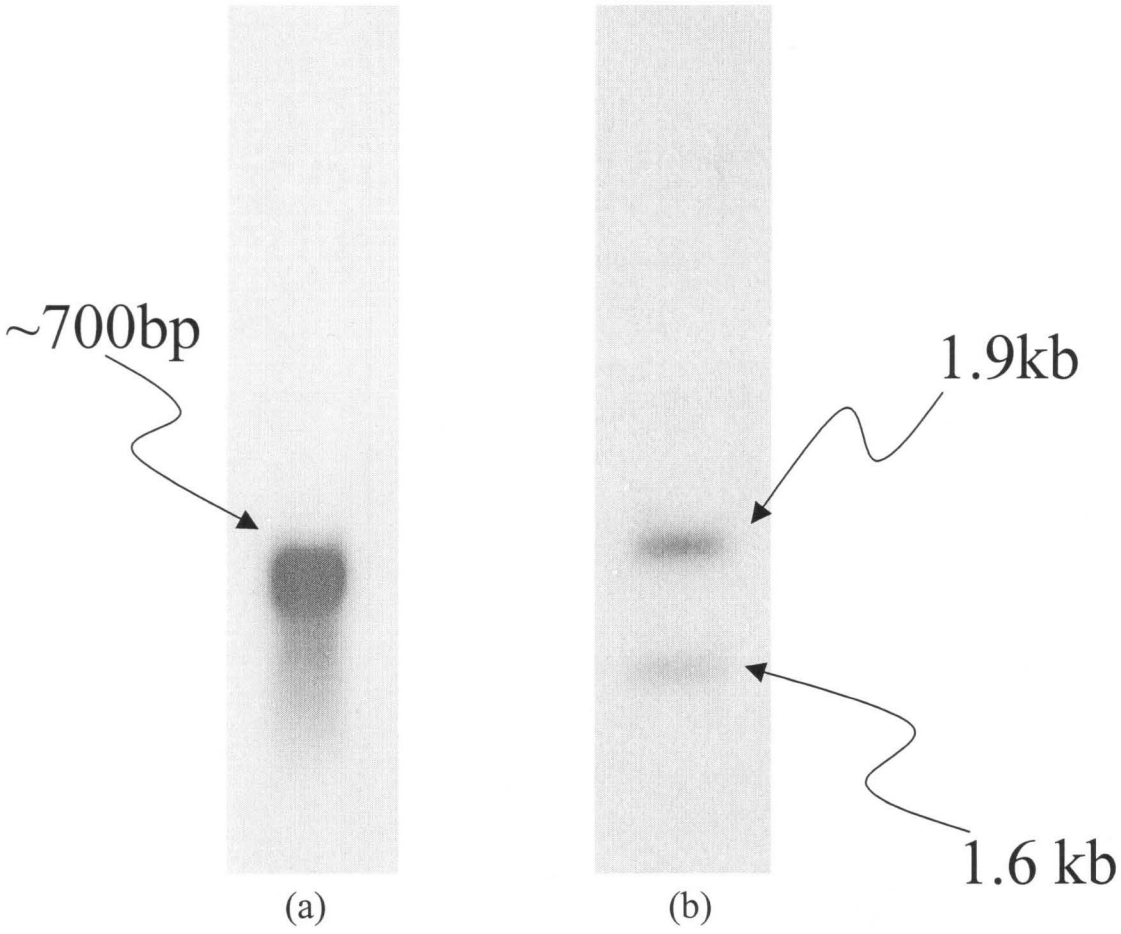


Figure 4.11 (a): Northern and (b): Southern analysis of *vRpL15*
(a): Northern analysis of *vRpL15*: total adult RNA probed with the *vRpL15* cDNA
(b): Southern analysis of genomic DNA cut with EcoRI and probed with the *vRpL15* cDNA. Note that the cDNA sequence for this gene has an internal EcoRI site.

Organism	%similarity	%identity
Drosophila melanogaster	99%	97%
Chironomus tentans	94%	85%
Anopheles gambiae	95%	88%
Mus musculus	85%	73%
Homo sapiens	85%	73%
Ictalurus punctatus	85%	73%
Caenorhabditis elegans	83%	69%
Picea mariana	81%	64%
Neurospora crassa	79%	67%
Aspergillus niger	79%	65%
Saccharomyces cerevisiae	79%	67%
Arabidopsis thaliana	79%	64%
Leishmania infantum	72%	55%

Table 4.3: vRpL15 protein homologies across different taxa

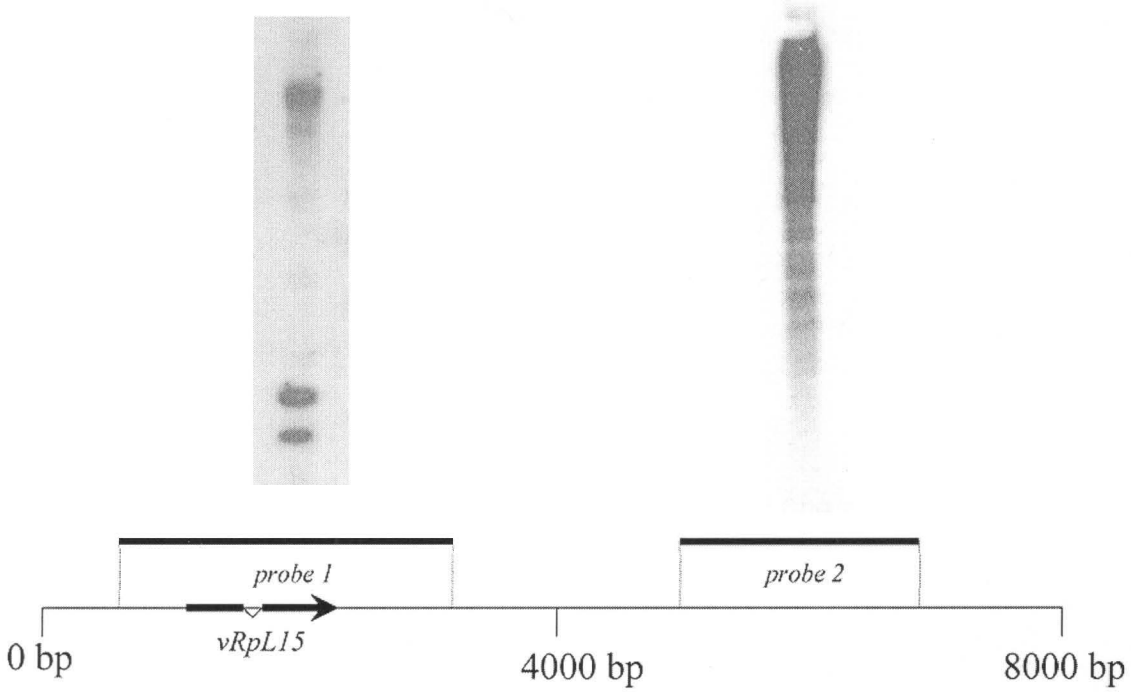


Figure 4.12: *vRpL15* upstream/downstream genomic Southerns
Southern analysis of *vRpL15* upstream and downstream regions. Genomic DNA was cut with EcoRI, transferred to nylon membrane and hybridized with genomic probes containing either the upstream or downstream regions of *vRpL15*.

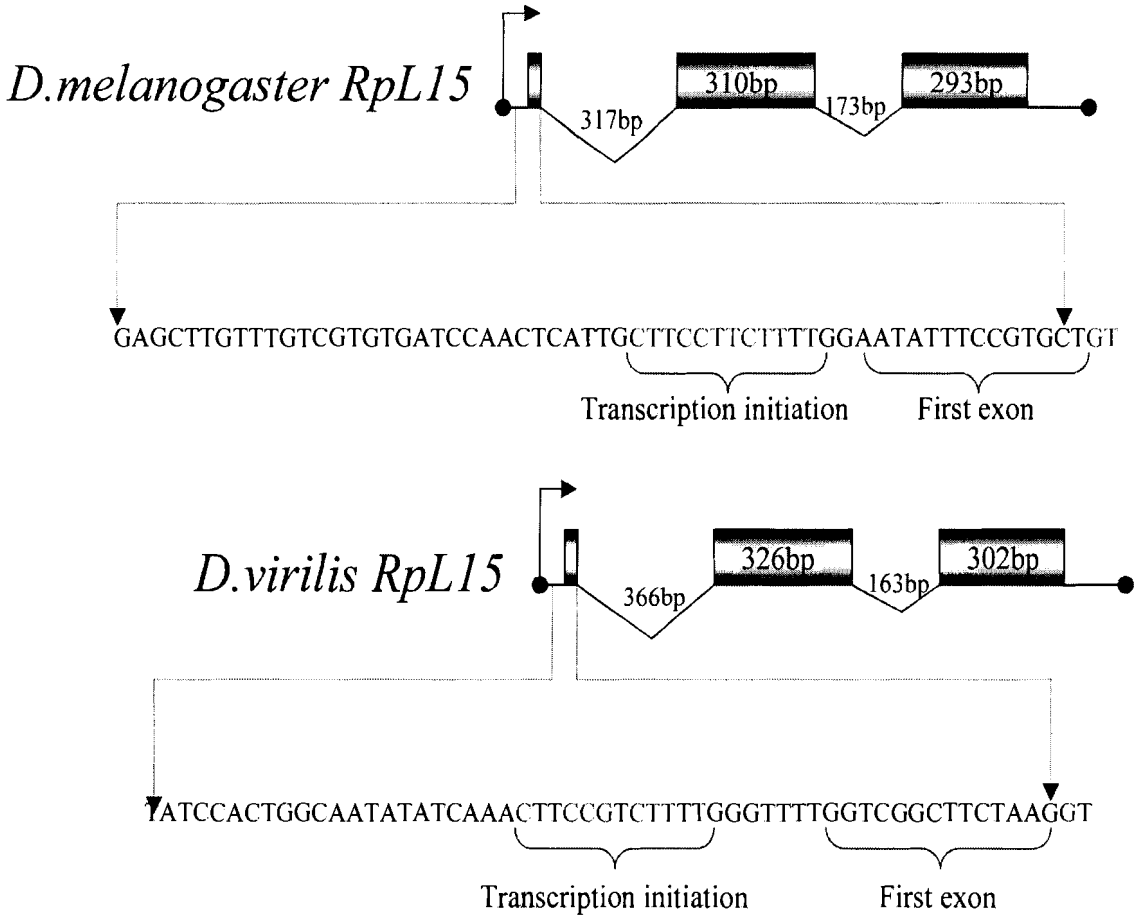


Figure 4.13 Comparative *RpL15* gene organization
Top: *Drosophila melanogaster*, and bottom: *Drosophila virilis*.

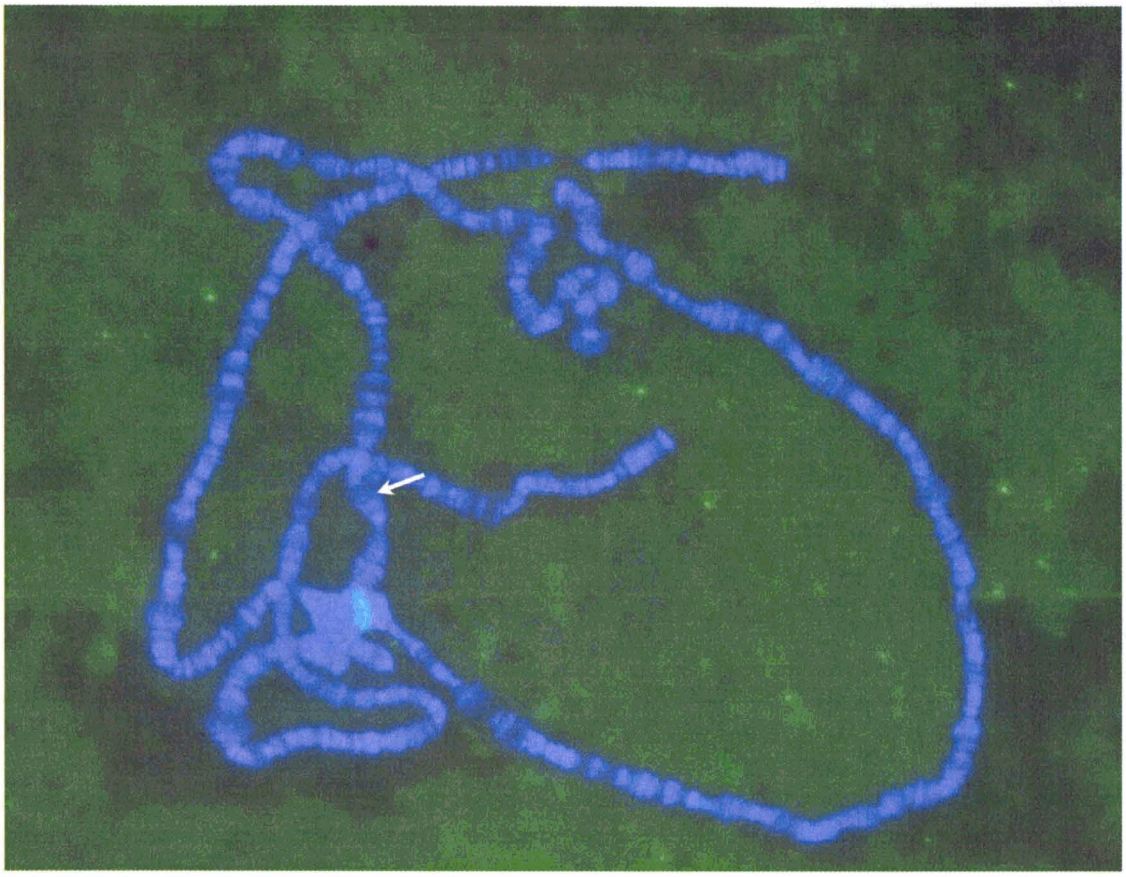


Figure 4.14: Chromosomal location of *vRpL15*
in situ hybridization of a *vRpL15* cDNA probe to *Drosophila virilis* polytene chromosomes. The signal is in the chromocentre, consistent with a heterochromatic location. Note that in this spread, the third chromosome (Muller's element D) is missing.

4.3. Sequence analysis

The *D.virilis* and *D. melanogaster* homologues of *RpL15* share both gene organization and heterochromatic environment. The same is not true for *Dbp80*, where the *D.virilis* homologue is euchromatic. This affords an opportunity to examine the same gene in two different chromatin environments. It is therefore relevant to ask whether there are any features evident from an inspection of the DNA sequence alone that might explain this difference. It is also now possible to factor in a third sequenced genome – *D.pseudoobscura*, which is separated from *D.melanogaster* by 30-45 million years (Figure 4.1). The genome sequence for this organism has recently been completed in draft form (www.hgsc.bcm.tmc.edu/projects/drosophila/update.html). Although none of these scaffolds have yet been mapped to the chromosomes, some inferences with respect to their genomic environment can be made. A BLASTN search using *mDbp80* and *mRpL15* against the *D.pseudoobscura* draft sequence makes two things clear. Firstly, both genes reside in a repetitive environment in *D.pseudoobscura* (Table 4.4). Secondly, the coding sequence for both genes in *D.pseudoobscura* exists in fragments, mapping over several scaffolds (which is generally not the case for euchromatic genes which have already been characterized in both species). Finally, *pDbp80* and *pRpL15* may well be linked, as they are in *D.melanogaster*, since exons for both genes share a scaffold (although they appear to be transcribed in opposite directions: Figure 4.15)

The repetitive environment (Table 4.4) and increased AT:GC ratio (Table 4.5) are the only relatively consistent and obvious features which characterize a heterochromatic vs. euchromatic domain. A transcription factor binding site analysis shows that the upstream regions of all these genes contain sequence homology to a range of transcription factor binding sites (Table 4.6). However, randomizing the sequences and running the same analysis produces a similar result (data not shown). Similarly, using algorithms to find conserved non-coding sequences in genomic DNA reveals significant alignments for both randomized and non-randomized sequences for these genes (non-randomized alignments shown in Figure 4.16 a,b).

Table 4.4: Repetitive sequence analysis of *Dbp80* and *RpL15* homologues in three Drosophilids

In each case, the query sequence is a specified portion of the genomic environment containing the *Dbp80* or *RpL15* homologues. Each BLASTN query was run against both the *D.melanogaster* repetitive DNA and Transposable element databases. Note that in heterochromatin, transposable element sequences show a tendency towards sequence degeneration. Therefore, alignment with intact sequence models for transposable elements is expected to be poor (i.e., smallest sum probability will approach 1.0).

DATABASE	QUERY	SEQUENCES PRODUCING HIGH-SCORING SEGMENT PAIRS	SMALLEST SUM PROBABILITY
Repeats	*mDbp80* genomic upstream (655bp)	Satellite DNA fragment 1.672-573	0.73
		mdg1	0.59
Transposons		Quasimodo	0.63
		Copia	0.97
Repeats	*pDbp80* genomic upstream (2200bp)	Alpha-gamma heatshock fragment	0.044
		Anon. Similar to 1.5kb repeat flanking Su(f)	0.22
		Auton. Replic. Seq.	0.37
		HP1-VS	0.93
Transposons		Micropia	0.0064
		Tabor	0.998
Repeats	*vDbp80* genomic entire (1991bp)	None	
Transposons		None	
Repeats	*mRpL15* genomic 3.5kb (used in rescue experiments)	Anon. Similar to 1.5kb repeat flanking Su(f)	6.7e-13
		Su(Ste)-like repeat	5.9e-12
		D.melanogaster suppressor of forked gene	1.1e-06
		HP1-VS	0.00010
Transposons		INE-1	2.8e-16
		1360 element	9.1e-07
Repeats	*vRpL15* gene region (1616bp)	Su(Ste)-like repeat	0.040
		Anon. Similar to 1.5kb repeat flanking Su(f)	0.98
Transposons		Het-A element	0.62
		Idefix	0.90
Repeats	*pRpL15* genomic incomplete (1104bp)	18S, 5.8S 2S and 28S rRNA – like sequences	0.93
Transposons		Het-A element	0.72
		TRANSIB1 element	0.98
		Hopper	0.995
		1360 element	9.1e-07
		Stalker	0.032
		Stalker4	0.032
		Rover	0.072

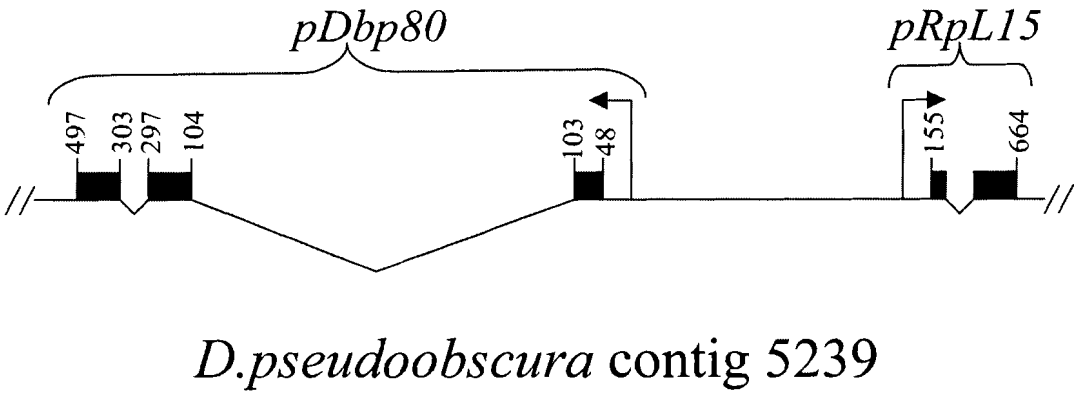


Figure 4.15: Preliminary organization of *Dbp80* and *RpL15* in *Drosophila pseudoobscura*

The *D.pseudoobscura* draft genome sequence is incomplete for both genes. Isolated exons can be found on a number of different scaffolds, which also contain a high density of repetitive sequences. One scaffold (contig 5239 shown here) contains exons from both genes, which appear to be transcribed in opposite directions, in contrast to their arrangement in *D.melanogaster*. Numbers above the exons (black boxes) refer to positions in the *D.melanogaster* cDNAs.

Gene region (genomic upstream)	Cytology	Total bp	%AT	%GC	Ratio AT:GC
mDbp80	Het (h51)	1397	65.4%	34.6%	1.9
vDbp80	Euch (35F)	918	60.6%	39.0%	1.5
mRpL15	Het (h51)	677	67.4%	32.6%	2.1
vRpL15	Het	647	68.2%	31.8%	2.1
pDbp80	Het?	2200	64.4%	35.6%	1.8
pRpL15	Het?	697	69.0%	31.0%	2.2
CG18001	Het (41C)	1761	67.6%	32.4%	2.1
RpS3A	Het (102A3)	957	71.3%	28.7%	2.5
CG12775	Het (40F)	1331	70.1%	29.9%	2.3
Yip6	Het (40E-F)	1703	66.9%	33.1%	2.0
Qm	Het 80D1-2	1140	67.4%	32.6%	2.1
RpL18A	Euch (54C3)	1151	55.7%	44.3%	1.2
RpL22	Euch (1C4)	1247	51.0%	49.0%	1.0
CG9871	Euch (59D3	1317	57.6%	42.4%	1.3
RpL23a	Euch (62A10)	243	51.4%	48.6%	1.0
RpL29	Euch (57D8)	562	62.5%	37.5%	1.7
RpL37A	Euch (25C4)	404	62.4%	37.6%	1.6
Hoip	Euch (30C)	208	65.5%	34.5%	1.9
RpL7A	Euch (6C1)	1598	64.1%	35.9%	1.8
RpPO	Euch (79A7)	517	59.0%	41.0%	1.4
RpL14	Euch (66D)	440	56.1%	43.9%	1.3
RpL36	Euch (1B13)	850	55.1%	44.9%	1.2
RpL7	Euch (31A)	412	46.4%	53.6%	0.9
RpL27A	Euch (31E)	504	64.3%	35.7%	1.8

Table 4.5: AT:CG ratios for heterochromatic and euchromatic upstream genomic regions

All genes from CG18001 on are ribosomal protein genes in *D.melanogaster*. The cytological positions are based on polytene banding patterns; the chromosomal locations for the *D.pseudoobscura* genes are not yet known.

QUERY	HITS	SEQUENCE	PROBABILITY	POSITION
mDbp80 683 bp upstream of ATG	pannier_CS	KKSYGATAAGGR	9.11e-02	4
	dFRA-hmtIIA	TGACTCA	7.62e-02	42
	dJRA-hmtIIA	TGACTCA	7.62e-02	42
	ZESTE_CS	YGAGYC	7.20e-01	43
	tailless_site	AAATTAA	7.62e-02	51
vDbp80 648 bp upstream of ATG	tailless_site	AAATTAA	7.55e-02	143, 514
	actin_5c_US	TATAAAA	7.55e-02	147
	antp/en_homeodo	STAATKG	2.69e-01	334, 599
	Ftz-artificial_	AACACATTACACGC	1.03e-02	474
	cuticle-gene-US	TGCATCA	7.55e-02	561
pDbp80 700 bp upstream of (putative) start	Ttk-eve	CCAGGACC	2.10e-02	141
	ZESTE_CS	YGAGYC	7.44e-01	280
	Bicoid-CS	BBTAATCYV	2.48e-01	406
	HB4	GATGCCAAAAAACG GC	1.12e-02	492
	Bicoid-X3	GATCATCCA	5.27e-03	535
	dl-GPIIb	CGAGAAAATCG	1.12e-02	594
mRpL15 677bp upstream of ATG	HB3	GAAAAAGAAAAA	2.94e-03	14
	tailless_site	AAATTAA	7.87e-02	130,168
	actin_5c_US	TATAAAA	7.87e-02	628
	Hb-en_(1)	CAAATAAATAA	1.08e-02	631
	engrailed_CS	HCWATHAAA	8.80e-02	658
vRpL15 647 bp upstream of ATG	tailless_site	AAATTAA	7.53e-02	10
	chorion_upstrea	RTCACGTW	7.52e-02	53
	engrailed_CS	HCWATHAAA	8.42e-02	97, 330
	Ftz-artificial_	AACACATTACACGC	1.03e-02	330, 518
	antp/en_homeodo	STAATKG	2.69e-01	386
pRpL15 607 bp upstream of ATG	engrailed_CS	HCWATHAAA	9.05e-02	177
	ZESTE_CS	YGAGYC	7.42e-01	212
	zeste-Ubx	CGAGCG	2.87e-01	240
	su(Hw)-MHC_class	YRYTGCATAYYY	1.93e-01	449
	HB4	GATGCCAAAAAACG GC	1.11e-02	515
	B-factor-hsp70	TATAAATA	2.09e-02	649

Table 4.6: Transcription Factor binding site analysis
Sequence analysis of the upstream genomic regions for *Dbp80* and *RpL15* in *D.melanogaster, D.pseudoobscura* and *D.virilis*. Note that the START methionine can only be molecularly confirmed for the *D.melanogaster* and *D.virilis* genotypes. The programme used to identify transcription factor recognitions sequences was TFSITE (see Chapter 2).

Figure 4.16(a): (following 3 pages): DiALIGN analysis of the upstream region of *RpL15* in *D.melanogaster, D.pseudoobscura* and *D.virilis*
The number and significance of hits is conserved when these sequences are randomized, but their distribution is not.

```
Alignment (DIALIGN format): M,V&P RpL15 upstream (upto and including Start ATG)
   ============================

mRpL15              1   gttgtttgat tttgacaaag aaaaaagtca gttgtttgag cagCAAGAAA
vRpL15              1   ctgt------ ---------- ---------- ---------- ----------
pRpL15              1   ---------- ---------- ---------- ---------- ---CAGAAAA

                                                                      *******


mRpL15             51   CGAAATAAAG CGAAATAGCA TAGAAATAAc gcatagcact atagcaa---
vRpL15              5   ---------- ---------- ---------- ---------- ----------
pRpL15              8   CAAAAATAAG AAACAAAAAA CAGAAAAAAa catttatttt tctcaacaaa

                        ********** ********** *********


mRpL15             98   --AAATTATT ---------- ---------- ---------- ----------
vRpL15              5   GCTAATTAAT TTTaaaggaa acactcgaat gaacttgcta gttatagaaa
pRpL15             58   GCAAATTATT TTTtat---- ---------- ---------- ----------

                        ********** ***


mRpL15            106   ---------- ---------- ---------- ---------- ----------
vRpL15             55   cgtgactact catcaacata tctatacggt acagatcgag TATCAATAAA
pRpL15             74   ---------- ---------- ---------- ---------- TATCAAAAAA

                                                                    **********


mRpL15            106   ---------- ---------- ---------- ---------- -------TGA
vRpL15            105   attggaa--- ---------- ---------- ---------- ----------
pRpL15             84   cgaatatcag agtgtacaca tattcgaatc tgagtactag gctgcagTGA

                                                                         ***


mRpL15            109   AAACCAGAGC ACCtaataac gaaattaaaa gacttttccc aaatcgaatg
vRpL15            112   ---------- --------TT ---------- ---------- ----------
pRpL15            134   GAACCCGAGC TCCcttacTA ---------- ---------- ----------

                        ********** ***     **


mRpL15            159   cgataagata aattaacatg gaacggaaca tgctgggatg ttttaaaaat
vRpL15            114   ---------- ---------- ---------- ---------- ----------
pRpL15            154   ---------- ---------- ---------- ---------- ----------


mRpL15            209   aatacaattc tgtgagaaag gggTATAATT CCGTTTtaca atgtttcatg
vRpL15            114   ---------- ---------- ---TATTGTT CCGTTTCCAG AGTTTAAGAA
pRpL15            154   ---------- ---------- ---TGTTGTT TCATTTACTT CGTTCCACAA

                                               ******* ********** **********


mRpL15            259   cattagggta agtatatcc- ---------- ---------- --CAAAGCAA
vRpL15            141   TCCTTGTTAA Tatcacagtc tggcatattc gatccgaata ttCAAAGCAA
pRpL15            181   TCAAAGTTTA Tttattatgt cgtgaacgct gtgagtcggt ------GCAA

                        ********** *                                 ********
```

```
mRpL15          286   CTCTGAgctt gtttgtcgtg tgatccaact ca------TT GCTTCCTTCT
vRpL15          191   CTCTtacaga gaatatccac tggcaatata tctaa----- -CTTCCGTCT
pRpL15          225   CTCTGAttgt attttcgctc gtcagaagcc aacttcacTT CCTTCCTTCT

                      ******                                  ** **********
                                                                **********
                                                                **********

mRpL15          330   TTTGGAATAT TTCCGTGCTG TAAGTTGGTT Gtgcattcgg gtcaATA---
vRpL15          235   TTTGGGTTTT GGTCGGCTTC TAAGGTGGGT Gggatacgg- ----ATA---
pRpL15          275   TTTAGAATTT GGTCGTCTTG TAAGCAGGGT Gtgtataata ttggatgtaa

                      ********** ********** ********** *              ***
                      ********** ********** ********** *
                      ********** ********** ********

mRpL15          377   ---------- ---------- ---------- ---------- ----------
vRpL15          277   ---------- ---------- ---------- ---------- ----------
pRpL15          325   agaatctatt gataccttg caatctttcg tttagtgtgt tatttttcaa


mRpL15          377   ---------- ---------- ---------- -ATTTTTCTT TAGTAcgaga
vRpL15          277   ---------- ---------- ---------- -GTTTTTAAA TAGAAttgcc
pRpL15          375   tttattgtgt cgatcagcaa tacccgagtt cATTTTTCTA TAtatatcgt

                                                  ********* *****


mRpL15          396   tcatttcgga a---ATAAGT TT-------- ATATATTATT TTCGGTcctt
vRpL15          296   tacgttgtaa tacgATCACT TT-------- ACATTTTgat aa--------
pRpL15          425   agacctcgtt atgcggcaaa tatttcttgc ATATTTTACT TTGGCTggac

                              ****** **         ********** ******

mRpL15          435   ----TTAAAA TGTTTTTAAG TTTgta---- ---------- ----------
vRpL15          330   ----TTTAAT TGTTTTTCTG TTTTTCAGGT ATGTGCAACT aacgggtatt
pRpL15          475   gaccTTAATA TATTTTTAAA TATTTGAAGC TAGTCTACCT tctgaata--

                    ****** ********** ********** **********


mRpL15          457   ---------- ---------- ------TGAT GATTTTTTCC ATGCACCACT
vRpL15          376   tattaaatta caattacAAT TTGAACTGAT GATATTTTCC ATGCACCACT
pRpL15          523   ---------- -------AAT TTGAACata- ---------- ----------

                                *** ********** ********** **********
                                          **** ********** **********

mRpL15          481   CTgagctca- ---------- -GTGTAGATT ATGGATAT-C TGACGGATgt
vRpL15          426   CTtatgcctt tgttgagcaa tGTGCAGATT ATGGATATgC TGACGCATta
pRpL15          535   ---------- ---------- ---------- ---------- ----------

                      **                     ********* ******** * ********
                      **


mRpL15          518   tatgttcgt- ---------- ---------- ---------- ----TTTTGT
vRpL15          476   gtagctataa attacagtat ttaactatat tatgtattat atctTAATAG
pRpL15          535   ---------- ---------- ---------- ---------- ----TATTAA

                                                                   ******
```

```
mRpL15          533   ACTCACATTG CTATGTAAca ttgatacgaa ttattatctg TTCTAGATTG
vRpL15          526   CCTCTTGTTT TACTTTTAT- ---------- ---------- GTTTAGCTAA
pRpL15          541   ACCCTTATTG ATATTTAAA- ---------- ---------- TTATAGATAA

                      ********** *********                       **********

mRpL15          583   TAATATGTAC AGTTTAAGAa g--------- ---------- ------TTGA
vRpL15          555   CTTAAaaagt tat------T TTAAAA---- ---------- ----------
pRpL15          570   CAAAATGCAC ATCTTTGGAT TTAAAAtgac tttacagtca aatgaaTTGA

                      ********** ********** ******                     ****
                                                                       ****
                                                                       ****

mRpL15          608   TGGACCGATG ATGTGACAGT TATAAAATAA ATAAtg---- ----AATTTG
vRpL15          575   --GACCAATG ATGTGACAcg ttatgtcaag --AAATAACA Tttc-ATTTG
pRpL15          620   TGGACCGATG ATGTGACTGT TATTAAATAT ATAAATAACA TataAATTTG

                      ********** ********** ********** ********** *   ******
                      ********** ********** ********** ****           ******
                      ********** ********** ********** ****            *****
                        ******** *******                               *****
                                                                       *****

mRpL15          650   TTACTAGGTC TATCAAattg cagagatg
vRpL15          620   TTACTAGGTC TATCAATCGT TGGATATG
pRpL15          670   TTACTAGGTC TATCAAGCAT TGGAAATG

                      ********** ********** ********
                      ********** ********** ********
                      ********** ******
                      ********** ******
                      ********** ******
```

Figure 4.16(b): (following 3 pages): DiALIGN analysis of the upstream region of *Dbp80* in *D.melanogaster, D.pseudoobscura* and *D.virilis*
Because these sequences are very long, only those regions containing significant alignments are shown.

Aligned sequences: M,V&P Dbp80 upstream (including ATG start)

1) mDBP80 2447
2) pDBP80 2200
3) vDBP 918

```
mDBP80          201   taGATCGATT TCAAATATGG TGCCGTGATA ACGTGTTaga cttaaatggc
pDBP80           10   ---------- ---------- ---------- ---------- ----------
vDBP             23   --GATTGTTG CGTACAATGG TGCGGTGGTG GGGTGCTctc ttctctaagn
                        ******** ********** ********** *******

mDBP80          301   cactctaagt ggatgcaatg tggacagaat aacacgaatt gatgatatat
pDBP80           10   ---------- --ATTTTTGC aattc----- ---------- ----------
vDBP            121   cccgccgaag acATTTTTGC tgtgttcgca gtttctttcg gttcctagcg
                                   ********

mDBP80          451   tttgatgatt ccttacttga CTAAAACCTA Acttatttcg ttagtccgtc
pDBP80           23   ---------- ---------- CTAAAATCTA Aaaatcatat ctctttgtag
vDBP            176   ---------- ---------- ---------- ---------- ----------
                                             ********** *

mDBP80          551   tcggACCGCA TGGAATcggt ccaaaaaaac ttcttacttT TTGCCTTGGg
pDBP80           62   ----ACCGTA TGGAATata- ---------- ---------- ----------
vDBP            176   ---------- ---------- ---------- --------TT TTGCCTTGGc
                          ****** ******                         ** *********
                                                                ** *********
                                                                ** *********

mDBP80          901   ttgactctct gccgctcttt aagcaatcga tTTTAACTTT TTTATTACAT
pDBP80           77   ---------- ---------- ---------- -TTAAACTTT TAAATTAGAA
vDBP            223   ---------- ---------- ---------- ---------- ----------
                                                       ********* **********

mDBP80          951   AATTGGatcc tactaacact aatatttaat atggttattt tacattatat
pDBP80           96   AGTGGGcgaa cctccttaaa tcg------- ---------- ----------
vDBP            223   ---------- ---------- ---------- ---------- ----------
                      ******

mDBP80         1151   atagaaccca gaaaaaattt taatttacgt gattataccc aaattgACAT
pDBP80          119   ---------- ---------- ---------- ---------- ----------
vDBP            223   ---------- ---------- ---------- ---------- ------ACAT
                                                                        ****
                                                                        ****

mDBP80         1201   ATTTACTACA TACTAATTAT TTTCAGTTAC AACGTTtctg ttaagtctcg
pDBP80          119   ---------- ---------- ---------- ---------- ----------
vDBP            227   ATATATAACT GCGCAGTTAT CTTGTTTAGC AACGTTcgca gcaacaccac
                      ********** ********** ********** ******
                      ********** ********** ********** ******

mDBP80         1822   agggtgtaca tttgactcat gaaattaaca taaaagtatc gcattgtgcg
pDBP80          204   aTATAGACTG ---------- ---------- ---------- ----------
vDBP            458   -TAAAGTTTG ---------- ---------- ---------- ----------
                       *********
                       *********
                       *********
                       *********
```

```
mDBP80          2022   atccaatgag tcttaacgta ataactggaa cactCCTAAA AGATCATTGT
pDBP80           214   ---------- ---------- ---------- ----CCACAA AAAACAacta
vDBP             467   ---------- ---------- ---------- ----CCAAAA AAAACATTGT
                                                         ****** **********
                                                         ****** **********
                                                         ****** ******
                                                         ****** ******
                                                         ****** ******

mDBP80          2072   AACATCATTA TAACC----- ---------- ---------- ----------
pDBP80           230   ttattcataa aacaaatttt aaaatagctt aaagatgtct tgatcttttt
vDBP             483   ACAATGTACA TTACA----- ---------- ---------- ----------
                       ********** *****
                       ********** *****

mDBP80          2087   ---------- ---------- ---------- -------AAA TATTTTTTTG
pDBP80           280   caatcacgca ctttcttccg tctaaacaag ctatggaAAA TATTATTATT
vDBP             498   ---------- ---------- ---------- -------GAA CAacgacatg
                                                            *** **********
                                                            *** **

mDBP80          2100   AATATATGAT ATCATATAag attata---- ----TATCGT ATTTATTTAA
pDBP80           330   AAAAAAAAAA AATTTATAct tacgaaaggt caacTATTGT TGTTACTGAA
vDBP             511   tccgattgta gtatcgatat gtccacagaa caacgacatg tccgattgt-
                       ********** ********             ****** **********
                                                       ****** **********
                                                       ****** **********
                                                       ****** **********
                                                       ****** **********

mDBP80          2142   TCATATTGAA AGTTAAGAAA tcaatgct-- ---------- ----------
pDBP80           380   ACATATGTAT AGATGAGAAA aatttgtgaa agcgggatta cctaaaaacc
vDBP             560   ---------- ---------- ---------- ---------- ----------
                       ********** **********
                       ********** **********
                       ********** **********
                       ********** **********
                       ********** **********

mDBP80          2170   ---------- ---------- ---------- ---------- ----------
pDBP80           430   caaatgtaat ttaaaactgg agtctacctt aatattttaa tgctatccag
vDBP             560   ---------- ---------- ---------- ---------- ----------

mDBP80          2170   ---------- ---------- ---------- ---------- ----------
pDBP80           480   tgttcgatta tttagtgttc taactttgca ttagttcttc aaatcgacaa
vDBP             560   ---------- ---------- ---------- ---------- ----------

mDBP80          2170   ---------- ---------- --TATATTAA ATTATATCTG GTTTACATAT
pDBP80           530   ttcgatatcg aatgAGTATC GACTTATCGA CTTATTGATA TTTacgctta
vDBP             560   ---------- ----AGTATC GATATGTCGA TATATTACTT TTTTAGACAT
                                   ****** ********** ********** **********
                                   ****** ********** ********** ***
                                   ****** ********** ********** ***
                                   ****** ********** ********** ***
                                   ****** ********** ********** ***

mDBP80          2198   Tacatgctcg tctaaaattc agactatttc ggcaattttt gtaaaacctc
pDBP80           580   aatttttca attcttttgt tgtgagctaa acattaaata tgacggattg
vDBP             596   Ttgcagc--- ---------- ---------- ---------- ----------
                       *
```

```
mDBP80          2248   tcaagctttt ccaaccagtc tt-------- ---------- ----ATGATG
pDBP80           630   ggttaaaaaa gcagaagacc aggaagtaac caagctggta tggaATGATG
vDBP             603   ---------- ---------- ---------- ---------- ----------
                                                                    ******
                                                                    ******

mDBP80          2276   AATGTTAGAA TTCGGTGTAA AACAATACAA TAAattatga aagtgtgttc
pDBP80           680   TACTTGAACA TTTTGTGTAA ATATATATAT TTAtaatttg tcgattatag
vDBP             603   ---------- ---------- ---------- ---------- ----------
                       ********** ********** ********** ***
                       ********** ********** ********** ***

mDBP80          2360   ---------- ---------- ---------- ---------- ----------
pDBP80          1030   gagtaACTAG TCGTTTCCGG TCGAAATTAA TTG------- ----------
vDBP             603   -----ACTAA TTGTTTCAGC TCGTACTAAT ATGctgatta tgaatagcgc
                            ***** ********** ********** ***

mDBP80          2360   ---------- ---------- ---------- ---------- ----------
pDBP80          1063   ---------- ---------- -TGAAACTTA ATAGAGTTCg caaagtggtg
vDBP             698   tacaactctt cggttacgtc tTGAAAGCTT ACAGATTTCt tattacgttc
                                              ********* *********


mDBP80          2365   ---------- ---------- ---------- ---------- --------TC
pDBP80          1430   cgaacaaaca ggaaaccagt tgatttgtaa tgctttgttt caaaagctTT
vDBP             863   ---------- ---------- ---------- ---------- --------TT
                                                                    **
                                                                    **
                                                                    **


mDBP80          2367   AGTGTAATTA TTTTTAGgaa agaccagttg gtcactctga cgtt------
pDBP80          1480   TGTGCAATTt ggatcgcttt aaaataccaa tcgctgacca ataccaatcg
vDBP             865   TGTGCAATTA CATTTTGact tcta------ ---------- ----------
                       ********** *******
                       *********
                       *********
```

DISCUSSION

The homologues of two linked heterochromatic genes in *Drosophila melanogaster* were cloned from a related species *Drosophila virilis*. cDNA's were first isolated from an embryonic plasmid library, and these were used to screen a λEMBL3 genomic phage library in order to obtain their genomic organization. The cDNA's were labeled and hybridized to *D.virilis* polytene chromosomes, in order to establish their physical locations.

D.virilis was selected as the comparative species for a number of reasons. Firstly, there are several proven libraries available, including many recent cDNA libraries, which have been generated by the BDGP specifically to aid in comparative research. Secondly, *D.virilis* and *D.melanogaster* belong to different subgenera (Figure 4.1), which division likely represents one of the most ancient evolutionary splits within the entire genus (40-60 my). Thus any observed sequence and/or positional conservation likely reflects important underlying biological functions. Thirdly, *D.virilis* exhibits what is considered to be the ancestral karyotype of this group, possessing 5 acrocentric "rod" chromosomes and a tiny "dot" sixth. *D.melanogaster* is considered to be a "derived" form, with specific chromosomal rearrangements leading to a karyotype consisting of two metacentric chromosome pairs (fusion of "rods"), an acrocentric X chromosome, and a tiny "dot" fourth. A full evolutionary analysis would require the examination of homologous genes through a range of intermediate species between *D.virilis* and *D.melanogaster*, however, some inferences can be made based on the data here presented.

vDbp80 is a small single copy euchromatic gene flanked by an LTR retrotransposon.

There are three principle features that emerge from a comparison between *D.virilis* and *D.melanogaster Dbp80* homologues. Firstly, *vDbp80* is a euchromatic gene. It resides on chromosome 3, which is Muller's element D and therefore homologous to *D.melanogaster* chromosome arm 3L. Secondly, *vDbp80* is flanked downstream by a full-length and very probably active LTR retrotransposon called *Ulysses*. Members of

this retrotransposon family have in fact been linked to a variety of chromosomal rearrangements, both in nature and in the laboratory, resulting from their activity during hybrid dysgenic events (Petrov et al.1995, Evgen'ev et al. 2000). Thus *Ulysses* or something like it could have played a role in the kinds of chromosomal rearrangements which might result in the diverse physical locations of this gene today. Thirdly, *Dbp80* has undergone a dramatic enlargement over time, due to the apparent insertion of new introns, and expansions of repetitive DNA therein. This kind of development is always valued for providing insight into the mechanism of intron evolution.

The first two points described above may be causally linked. The extent to which transposable elements have contributed to the kinds of chromosomal rearrangements responsible for moving *Dbp80* into heterochromatin will require a fuller analysis of this gene in a number of intermediate species. For example, the gene might be cloned and positioned in *D.texana*, which belongs to the same subgenus as *D.virilis* but has undergone the D+E (3L+3R) fusion characteristic of *D.melanogaster* (Patterson and Stone 1952). A similar analysis in *D.yakuba*, which shares the subgenus *Sophophora* with *D.melanogaster*, would also be fruitful, since it represents the deepest evolutionary split within this subgenus, (and is also a target species for genome sequencing). The objective would be to find out when and how often heterochromatic genes in *Drosophila melanogaster* became heterochromatic. For instance, preliminary evidence suggests that *Dbp80* might be heterochromatic in *D.simulans* – a sibling species. A cross-species genomic Southern suggest that *sDbp80* is also a very large gene (Figure 4.17). The same may be true for *D.pseudoobscura.*

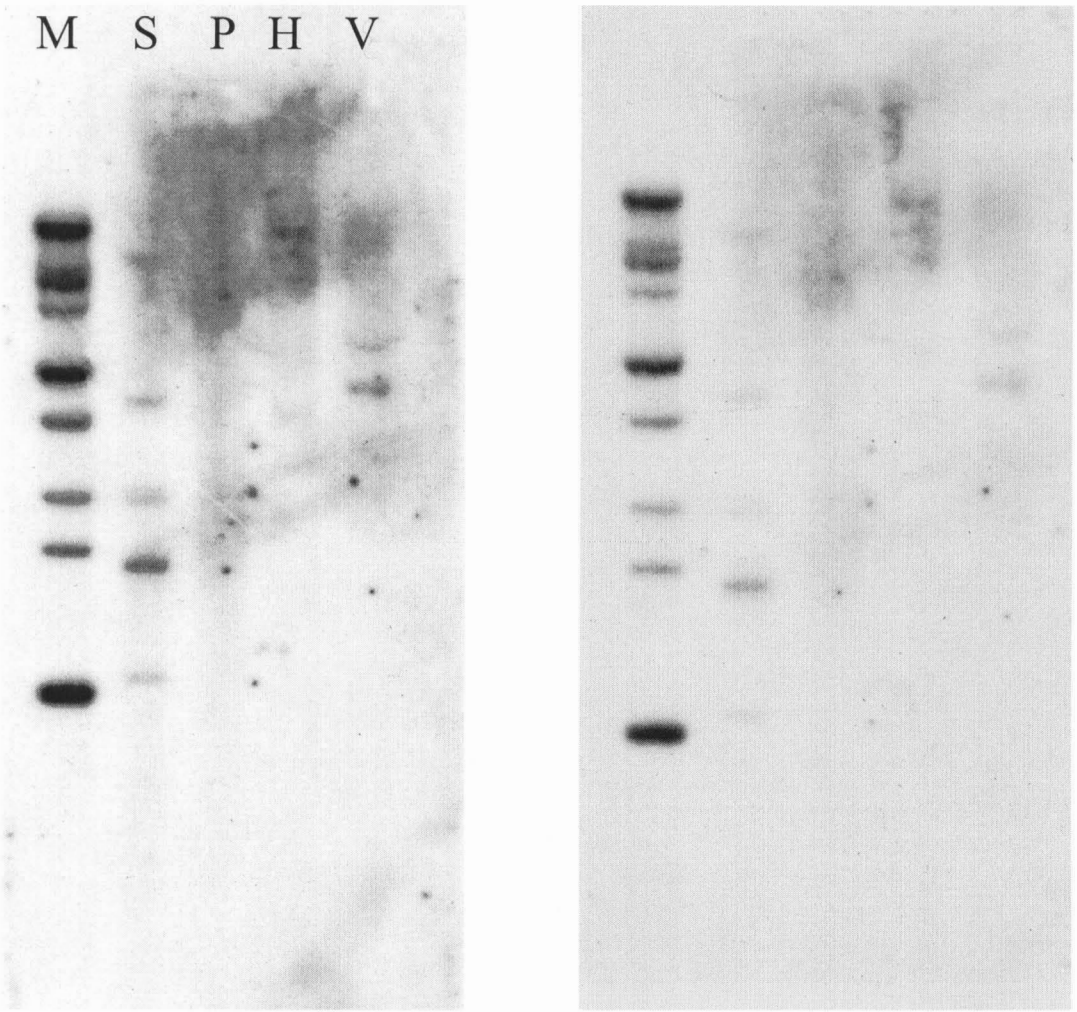


Figure 4.17: Southern analysis of *Dbp80* in 5 *Drosophila* species
Long and short exposures of a cross-species southern of genomic DNA taken from 5 species of *Drosophila.* M=*melanogaster*, S=*simulans,* P=*pseudoobscura*, H=*hydei*, V=*virilis.* Genomic DNA was cut with EcoRI and hybridized at low stringency with the *mDbp80* cDNA.

The first draft of the *D.pseudoobscura* genome has recently been released (www.hgsc.bcm.tmc.edu/projects/drosophila/update.html). An attempt to align *mDbp80* and *mRpL15* with *D.pseudoobscura* genomic contigs produces a fragmentary and incomplete map (Figure 4.16). This may not be significant, since the *D.pseudoobscura* genome project is in its earliest phases, and there has as yet been no cytological mapping. Yet it is notable that exons from *pDbp80* and *pRpL15* can be found on the same contig, suggesting that they may be linked in this organism, as they are in *D. melanogaster*. Since both genes are heterochromatic in *D.melanogaster*, and *RpL15* is also heterochromatic in the more distantly related *D.virilis* (see below), this strongly suggests both genes are heterochromatic in *D.pseudoobscura* as well. In addition, using the entire *D.pseudoobscura* contig containing exons from both genes as a query in a BLASTN search of the *Drosophila melanogaster* transposable element and repetitive DNA databases shows that this region of the *D.pseudoobscura* genome is rich in repetitive sequences (Table 4.4). Finally, since *RpL15* is highly conserved in terms of its chromatin environment, and since it is linked with *Dbp80* on Muller's element D in *D.melanogaster* (chromosome 3L), this suggests that both genes also reside on Muller's element D in *D.pseudoobscura*. Muller's element D has become fused to the X chromosome in *D. pseudoobscura*, and has long since lost its homolog, and is therefore fully dosage compensated (Bone and Kuroda 1996). Therefore both genes are not only likely to be heterochromatic, but they must be dosage compensated as well.

vDbp80 is flanked by a large transposable element of the LTR retrotransposon family called *Ulysses*. Transposable elements are probably ubiquitous in the animal kingdom (assayed by cloning or PCR: Arkhipova 2001) and have long been thought to play a critical role in genome evolution (Bowen and Jordan 2002). In *D.melanogaster*, certain types of parasitic elements have in fact been co-opted for specific vital functions, most notably the HeT-A and TART elements that comprise essential components of the telomeres (Fanti and Pimpinelli 1999). There are also examples of genes which remain associated with transposable elements throughout long evolutionary periods (McCollum et al. 2002), implying a functional dependence. It has been elegantly demonstrated that P elements can cause chromosomal rearrangements via proven recombination mechanisms

(Gray et al.1996, Preston and Engels 1996). This has been exploited in experimental designs, but may also help to elucidate the role natural transposable elements play in evolution. It is clear that transposable elements can cause all kinds of rearrangements, and if these rearrangements take place in germ cells, they can be inherited, and serve as raw material for genomic evolution.

Ulysses is (so far) unique to the species group to which *D.virilis* belongs, and exhibits interstrain polymorphism (Zelentsova et al.1999). In addition, it can be induced to transpose, both by creating a dysgenic cross (similar to the kind used in P element screens), and by germline transformation with an active element (Evgen'ev et al.1997). Interestingly, *Ulysses* itself may not be responsible for this phenomenon, but another element, *Penelope*, also unique to the *virilis* group, appears to be a trigger, causing the co-mobilization of several different transposable elements (*ibid.*). As for other parasitic elements, *Penelope* and *Ulysses* both have preferential insertion sites ("hotspots") which probably correlate less with DNA sequence than they do with the state of the chromatin (ie., weak consensus, *ibid.*). It is also significant that many of the insertion sites and "hot spots" which have been studied in this group coincide with natural breakpoints for inversions (Evgen'ev et al.2000). It has been proposed that these inversions might come about as a result of ectopic recombination between the varied locations of these elements along a chromosome arm. In fact, according to a recent study (*ibid.*), there appears to be a *Ulysses* element located in the vicinity of 35F on chromosome 3 that coincides with an inversion breakpoint. It is likely that this *Ulysses* element is the one flanking *vDbp80*. This mechanism of chromosome rearrangement provides one acceptable explanation for the very large numbers of intra-chromosomal inversions that appear in all members of this genus (Ranz et al. 2001). A specific example is shown in Figure 4.9, a comparative map of the region containing *vDbp80*. From this figure it is clear that multiple rearrangements must have taken place during the course of *Drosophila* evolution.

Perhaps the most striking difference between these two homologues is their drastically different genomic organization. With rare exceptions (an example of which forms a considerable part of this thesis – *RpL15*), genes in heterochromatin are very large (see

Table 1.1) in terms of the genomic territory occupied by their exons. This is believed to be due the expansion of repetitive DNA within the introns, which can approach sizes on the order of megabases (Kurek et al.1998). The contrasting gene structures of *mDbp80* and *vDbp80* certainly appear to follow this pattern. *vDbp80* possesses a single intron which may be considered ancient, because it is shared by the *Drosophila,* mouse and human homologues (but not the *Anopheles*: www.ensembl.org/Anopheles_gambiae/geneview?gene=ENSANGG00000012461). *vDbp80* spans less than 2 kb of genomic DNA, but when relocated during the course of evolution to heterochromatin, the genomic territory has expanded by a factor of 90.

There are two competing hypotheses to explain the presence of introns (Cho and Doolittle 1997). The "introns early" hypothesis states that introns were in the earliest ancestral genes, and were important for the "exon shuffling" that might have been crucial in evolving diverse protein functions. Support for this theory is circumstantial: rare instances of introns in bacterial and mitochondrial genes (Belfort et al. 1995, Eskes et al. 1997). The explanation for the current paucity of introns in these genomes today is that they have been lost over time. Intron loss is considered an essential part of this theory, since an important implication is that all introns today represent a subset of all that ever existed. The "introns late" hypothesis suggests that introns should be treated just like any other parasitic kind of DNA – they inserted ("invaded") eukaryotic genes long after the divergence from prokaryotes. In support of this hypothesis, there is evidence that some group II introns can act like mobile elements (Eskes et al. 1997, Palmer et al. 2000).

The single intron in *vDbp80* is a phase 1 intron (i.e., the intron splits the codon in the first position). Phase 0 introns are thought to be significantly more common than Phase 1 or Phase 2 introns (Rzhetsky and Ayala 1999). This is often cited as evidence supporting the "introns early" hypothesis, since the proto-exons which were shuffled as a result of the presence of ancient introns were unlikely to be split within codons. But most of the introns which split *mDbp80* are not phase 0, and in general, the "introns late" hypothesis provides the simplest explanation for the expansion of *Dbp80* in *D.melanogaster.*

Otherwise, duplicate genes would have to be invoked, one member of each pair subsequently and coincidentally lost in the two lineages.

There is also evidence which suggests that RNA helicase protein genes exhibit a huge range of diversity in terms of their genomic organization (while the proteins remain very highly conserved - Boudet et al. 2001). These proteins are involved in almost every step of RNA metabolism; the flexibility in gene structure may imply a similar regulatory flexibility, allowing different members of this large family to overlap in function. This idea may go some way to explain the discrepancy in *Dbp80*'s function across taxa. It is essential for mRNA export in yeast (Snay-Hodge et al. 1998) and appears to retain that function in humans (Schmitt et al. 1999), yet is dispensable in this context in *Drosophila* (Gatfield et al. 2001). In *Chironomus tentans* (midge) it appears to behave as a kind of chaperone, remaining associated with certain genes undergoing expression, from transcription through to export (Zhao et al. 2002), and dsRNAi knockout of the *C.elegans* homologue has no apparent affect (see Chapter Three, Discussion). So far as has been determined, *Dbp80* is only heterochromatic in *D.melanogaster,* and this chromatin context may have changed its regulatory properties and therefore its functional role.

In sum, *vDbp80* has been relocated from a euchromatic to a heterochromatic position within the same chromosomal element at least once during a period of 40-60 million years, possibly through the action of transposable elements. It has undergone dramatic reorganization in its structure, while remaining highly conserved at the protein level. What its function is, and whether this has also been conserved or lost as a consequence of this move, is not known.

vRpL15 and mRpL15 share both gene organization and chromosomal location

In contrast to *Dbp80, RpL15* has remained very tightly conserved both in terms of gene structure and location. It shares its organization and size with *Anopheles gambiae,* (www.ensembl.org/Anopheles_gambiae/geneview?gene=ENSANGG00000018869) which is separated from both *Drosophila* species by 250 million years. *RpL15* encodes a

component of the ribosome, one of the most ancient and essential structures in life, so this extremely high order of conservation is perhaps not surprising. What is notable is that in both species, this gene appears to be located in heterochromatin. Given the exceptionally rapid rate with which intrachromosomal rearrangements evidently take place in *Drosophila* (Ranz et al.2001), this might suggest that pericentric heterochromatin is resistant to this kind of genomic plasticity.

To my knowledge, *vRpL15* is the first known heterochromatic gene in *D.virilis* to be cloned. The only other *D.melanogaster* heterochromatic gene outside of this study to be cloned in *D.virilis* is the *light* gene, which is reported to be euchromatic in this species (Nurminsky et al. 1996). *Su(f)* has also been cloned in *D.virilis* (Audibert and Simonelig 1998), but it has not been cytologically mapped. *RpL15's* small size poses a puzzling question: how does a gene stay so compact in a genomic environment in which introns are known to grow very large due to the insertion and expansion of repetitive sequences? It is clear from both homologs that repetitive sequences have found their way into the introns (Table 4.4), but there is clearly a strong selective pressure to keep this gene small. Also, there is evidence which suggests that some ribosomal protein gene regulatory sequences are located within the first intron (Chung and Perry 1989), and DiALIGN alignments of the noncoding regions between these homologues indicate a number of well conserved regions which increase in number and significance after the polypyrimidine tract (which serves as an initiator) and through the first intron (Figure 4.16a). In general, genes that must be expressed at high levels, either because they play a housekeeping role, or are activated under conditions of stress, are small, sometimes intronless (Singh et al. 2000, Castillo-Davis et al. 2002), and ribosomal protein genes certainly fall into this category. In addition, while they are linked into a series of operons in bacteria, they are widely scattered around the genome of eukaryotes. Therefore they must have evolved regulatory features which render them resistant to position effects. So far, very few of these regulatory signals have been identified. In yeast, ribosomal protein genes appear to have insulator-like sequences (Bi and Broach 1999), and a common feature of their promoters across taxa is the polypyrimidine tract. This tract is not

absolutely universal, but it has been demonstrated to play a critical role in establishing their transcriptional activity (Hariharan and Perry 1990).

There is one other intriguing feature of the *Drosophila RpL15* homologues. A BLASTN analysis using repetitive DNA databases indicates the presence of HeT-A-like elements in the second intron of the *D.pseudoobscura* and *D.virilis* homologues (Table 4.4). The degree of identity is weak, but notable, and even higher in *D.virilis* than in *D.pseudoobscura.* HeT-A elements are co-opted transposable elements, which in *Drosophila melanogaster* play an essential role in telomere maintenance (Fanti and Pimpinelli 1999). They have also been found in many members of the *Drosophila* genus including *D.virilis* (Casacuberta and Pardue 2003). It has long been thought that these elements, and others like them, are confined to telomeric DNA, but as the heterochromatic portions of the genome are more fully sequenced it is becoming clear that they can turn up in pericentric heterochromatin as well (Losada et al. 1999, Agudo et al. 1999). Degenerating HeT-A or TART sequences near the centromeres may simply reflect ancestral copies that have undergone further insertion/deletion events during the course of their inactivation in heterochromatin. But they may also point to the possibility that the kinds of chromosomal fusions which produce metacentric configurations from rods may include telomere-centromere fusions, as well as the centromere-centromere kind. During interphase in polytene nuclei, the pericentric heterochromatin of all the chromosomes appears to aggregate in a body known as the chromocentre. Similar arrangements of chromatin within the nuclei of primordial germ cells may also take place, and under these circumstances, it is conceivable that fusion events between telomeres and centromeres could take place during this time, leading to the accumulation of telomeric sequences in pericentric heterochromatin. In support of this view, it is intriguing to note that the position of *RpL15* in *Anopheles gambiae* is relatively close to the telomere of Muller's element D which is homologous to chromosome 3L in *D. melanogaster* (www.ensemble.org/Anopheles_gambiae/).

In sum, *vRpL15* and *mRpL15* share both gene size, organization, and heterochromatic chromosomal environment. Since there are neither genetic tools nor extensive

sequencing projects for *D.virilis*, this physical location can only be determined by chromosomal *in situ* hybridization. Heterochromatic signals are notoriously difficult to assay in this way, due to the aggregation of the heterochromatin from all the chromosomes, making it virtually impossible to identify which chromosome arm the signal recognizes. It is perhaps illuminating that the spread shown in Figure 4.14 is in fact missing Muller's element D, which suggests that either *vRpL15* is very close to the centromere in this element, or that it in fact resides on another arm. As for *vDbp80,* an analysis of *RpL15's* location in a variety of intermediate *Drosophila* species would also be informative.

Sequence Analysis

Having genomic sequence for the homologues of two contrasting genes in three related *Drosophila* species, provides an opportunity to look for sequence features that might explain the contrasting chromatin environments in which these genes are found. However, such an analysis reveals little, and raises more questions than it answers. The only relatively indisputable sequence difference between a heterochromatic and euchromatic environment concerns the number and density of repetitive elements in and around the coding material (Table 4.4) and the AT:GC content (Table 4.5), although in this latter case, since the *Drosophila* genome is rather AT rich, this trend is not striking. Heterochromatin replicates later in the cell cycle than euchromatin (Ahmad and Henikoff 2001), but this may not include those potentially scattered and rare domains in heterochromatin which contain protein-coding genes, and which are surrounded by middle repetitive elements upon which they may have become dependent. In other words, transcription of heterochromatic genes may depend on co-opted sequences from middle repetitive elements, or be linked in some way to replication timing.

A search for conserved non-coding elements (principally in the upstream regions) is even more nebulous, but this is an endemic problem in genome analysis. There are as yet no unambiguous criteria by which non-coding regulatory elements might be defined, as there are for exons, like codons and open reading frames (Dermitzakis et al. 2003). In addition,

the definition of a sequence motif is necessarily fluid: does it bind a transcription factor? Does it provide a way of defining transcriptional domains (boundary elements)? Does it create a specific steric effect in the DNA molecule (bending etc.)? It is also now clear that both boundary elements and transcription factors respond to a range of sequence motifs, in a combinatorial fashion, leading to a graded rather than an on-off effect.

Numerous algorithms have been written which attempt to address all or some of these problems, and have been tested on known transcriptional networks, with varying degrees of success (Bergman and Kreitman 2001, Dermitzakis et al. 2003). For this work, I chose to use DiALIGN, which is a global alignment algorithm. The default settings (T=0; regions of maximum similarity denoted by 5 "*") yielded significant alignments both with randomized and non-randomized sequences. However, in the latter case, and specifically for *RpL15*, these were concentrated after (and including) the promoter region and through the first intron, a region which has been shown to contain significant regulatory signals in other taxa (Chung and Perry 1989). Therefore when combined with a knowledge of a gene's biology, these programmes can be an effective way of finding potential targets for directed mutagenesis studies. This in turn may lead to an identification of those trans-acting factors that interact with components of heterochromatin to regulate the expression of genes located therein. Similarly, comparing known transcription or boundary factor binding sites with conserved, aligned non-coding DNA might similarly identify critical regulatory proteins.

In the present analysis, there is no obvious coincidence between the homology to transcription factor binding sites, aligned non-coding DNA, and potential boundary regulators. In particular, the latter would be of interest, in helping to understand how a gene can be active in an otherwise repressive environment. However, to date, very few of these elements have been defined. The initial discovery in *Drosophila* that the scs and scs' elements interacted with the trans-acting factors SBP and BEAF to mark the periphery of a heat shock domain was followed by an intense study of the *Gypsy* insulator and its regulators Su(Hw) and mod(mdg4) (Zhan et al 2001). Recently, BEAF has been shown to interact with the non-histone chromosomal protein D1, that binds to an AT-rich

DNA element called BE28 which is localized to the pericentric regions of 2L, 2R and X (Cuvier et al 2001). There are likely as many variant boundary elements and trans-acting factors that control them as there are transcriptional regulators, and analysis of non-coding DNA would greatly benefit from a database of these motifs.

CONCLUSIONS

The two linked genes *Dbp80* and *RpL15* in *D. melanogaster* 3L heterochromatin were cloned and characterized in the related species *D.virilis*. *vDbp80* is a euchromatic gene which maps to the third chromosome (homologous to 3L in *D.melanogaster)*, and is vastly reduced in genomic size. It is flanked by a large retrotransposon called *Ulysses*, which is very likely still active, and may have played a role in the kinds of intrachromosomal rearrangements that made *Dbp80* a heterochromatic gene in *D. melanogaster*. *vRpL15* in contrast appears to be conserved both in gene organization and chromosomal location. A cross-species Southern using a *D.melanogaster* probe suggests that at least *Dbp80* may still be heterochromatic in the sibling species *D.simulans*, and the recently released draft sequence for the *D.pseudoobscura* genome suggests that both *Dbp80* and *RpL15* are linked and heterochromatic in this species. Thus, *RpL15* has been heterochromatic in *Drosophila* for at least 40-60million years, but *Dbp80* appears to have been relocated from euchromatin sometime after the split from *D.virilis* and before the evolution of *D.pseudoobscura* lineage. A more precise resolution would require an analysis of both genes in a range of intermediate species. This may soon become feasible, since the announcement of an initiative to sequence at least 10 more *Drosophila* genomes (http://flybase.org/.data/news/announcements/WhitePaperInfo.html). This initiative will also provide a wealth of comparative sequence information for non-coding DNA analysis algorithms, hopefully leading to a richer understanding of genome evolution.

APPENDIX 1: *RpL15* TRANSGENIC LINES

TRANSGENE LINE	DESCRIPTION	V/L[1]	MAP	REPORTER STRENGTH [2]
A16-3	RpL15 HIII genomic in pCaSpeR	V	II	Moderate
A16-7	RpL15 HIII genomic in pCaSpeR	V	X	Moderate
A16-12	RpL15 HIII genomic in pCaSpeR	V	X	Strong
A23-17	RpL15 HIII genomic in pCaSpeR	V	X	Moderate
A23-32[2]	RpL15 HIII genomic in pCaSpeR	V	Possible III	Variegates (strongly)
A23-36	RpL15 HIII genomic in pCaSpeR	V	II	Strong
A24-1	RpL15 HIII genomic in pCaSpeR	V	X	Strong
A28-2	RpL15 HIII genomic in pCaSpeR	L	II	Weak
A28-9	RpL15 HIII genomic in pCaSpeR	V	III	Variegates (weakly)
A32-2	RpL15 HIII genomic in pCaSpeR	L	X	Moderate
A32-4	RpL15 HIII genomic in pCaSpeR	V	II	Weak
A38-1	RpL15 HIII genomic in pCaSpeR	L	II	Weak
A38-8	RpL15 HIII genomic in pCaSpeR	V	II	Strong
A39-3	RpL15 HIII genomic in pCaSpeR	V	X	Weak
A41-2	RpL15 HIII genomic in pCaSpeR	V	II	Weak
Rg1.1	RpL15 BGL-HIII genomic in pUAST	V	III	Weak
Rg1.2	RpL15 BGL-HIII genomic in pUAST	V	X	Strong
Rg5	RpL15 BGL-HIII genomic in pUAST	V	X	Variegates (weakly)
Rg5.2[1]	RpL15 BGL-HIII genomic in pUAST	V	II	Weak
Rg5.2[3]	RpL15 BGL-HIII genomic in pUAST	V	II	Strong
Rg20	RpL15 BGL-HIII genomic in pUAST	V	II	Strong
Rg20a	RpL15 BGL-HIII genomic in pUAST	V	II	Moderate
Rg30c	RpL15 BGL-HIII genomic in pUAST	V	II	Strong
1373UAS15-1	RpL15 cDNA in pUAST	V	II	Strong
1373UAS17-2[1]	RpL15 cDNA in pUAST	V	X	Moderate
1373UAS23-1	RpL15 cDNA in pUAST	V	X	Weak
1373UAS24-1	RpL15 cDNA in pUAST	V	X	Strong
1373UAS40-3	RpL15 cDNA in pUAST	V	X	Strong
1373UAS133-1	RpL15 cDNA in pUAST	V	II	Moderate
1373UAS138-1	RpL15 cDNA in pUAST	V	Possible IV	Variegates
1373UAS156-1	RpL15 cDNA in pUAST	V	II	Moderate
1373UAS180-5	RpL15 cDNA in pUAST	V	II	Weak
1373CSP5-2	RpL15 cDNA in pCaSpeR	V	X	Moderate
1373CSP17-2[1]	RpL15 cDNA in pCaSpeR	V	X	Moderate
1372CSP17-10	RpL15 cDNA in pCaSpeR	V	X	Moderate
1373CSP23-1	RpL15 cDNA in pCaSpeR	V	X	Moderate
1373CSP36-5	RpL15 cDNA in pCaSpeR	V	X	Moderate
1373CSP40-1	RpL15 cDNA in pCaSpeR	V	X	Weak
1373CSP40-3	RpL15 cDNA in pCaSpeR	V	II	Weak
1373CSP123-1	RpL15 cDNA in pCaSpeR	V	II	Strong

[1]V=Viable, L=Lethal; [2]Strong = red; Moderate = orange; Weak = yellow

APPENDIX 2: PRIMER SEQUENCES

5'-3' seq & name	+/-	Tm	Location
RpL15 01: GGA TAC AGA GCC AAA CAG G	+	59.3	RpL15 CDS
RpL15 02: TCT ATA TCC CTT GCC AAT G	-	57.5	RpL15 CDS
RpL15 03: CAC ATC GCT CTG CTT CTT C	-	60.5	RpL15 CDS
RpL15 04: CAA TTG GTG GAT CTA GGC G	+	62.9	RpL15 CDS
RpL15 05: CCG TGC TGT AAG TTG GTT GT	+	63	RpL15 CDS
RpL15 06: AGG AAG AAG CAG AGC GAT GT	+	63	RpL15 CDS
RpL15 07: GTA CCG ATA AGC CCC CAT C	-	66	RpL15 CDS
RpL15 08: CCA CGC AAT TCA CGA TGC T	-	66	RpL15 CDS
RpL15 09: GCA TGT GCT CAC GGT TCT TG	-	67	RpL15 CDS
RpL15 GSP10: AGC GTC CAG TTC CCA AAG G	+	65	RpL15 CDS
RpL15 GSP11: TTG AAC TCG TAT TGG ATT GCG	+	62	RpL15 CDS
DvRpL15 01: TCC GTG TTT GGC AGT ATC G	+	64.4	vRpL15 CDS
DvRpL15 02: GCT TAC GAC CAC CAC GGC	-	66.2	vRpL15 CDS
DvRpL15 03: AGA TCC ACC AAT AGT TTG CG	-	62.0	vRpL15 CDS
DBP80 GSP01: GGC TTT GTG CTA TCA TAT	-	52.3	DBP80 exon 4
DBP80 GSP02: TGG ATC TGC GAG AAG AGT T	-	60.1	DBP80 exon 4
DBP80 GSP03: AGC TTA GAA ACC TCC TGG TC	+	58.9	DBP80 exon 3
DBP80 GSP05: GGA TCT GCA ACA ATC AGC C)	+	63.3	DBP80 exon 7
DBP80 GSP04: AGG TTA TGG ACT TTG CTC G	-	59.3	DBP80 exon 7
DBP80 GSP 06: ACA TGG GTG ACT GGG CTA A	+	63	DBP80 exon 1
DBP80 GSP 07: CGT AGG GGA TAA ACA CAG AAC	-	61	DBP80 exon 5
DBP80 GSP 08: AAC GTT GAC TCG GCT AAG CAT	-	66	DBP80 exon 4
DBP80 GSP09: AGC GAC GTC GAA ATC GGT T	-	65	DBP80 exon 2
DBP80 GSP10: GAT CCA GCT GAA ACA AGT CTC	+	63	DBP80 exon 3
DBP80 GSP11: CAG AAC CTG AGG ATG ATT TAA ACA	-	64	DBP80 exon 4
DBP80 GSP12: TGT TTA TCC CCT ACG TAC GAG T	+	64	DBP80 exon 5
DBP80 GSP13: CAT GAT GAC CTT GGG TGG CTA T	-	65	DBP80 exon 6
DBP80 GSP14: ATG CTG AAT CCA CAT TGC CAA A	+	64	DBP80 exon 7
DBP80 GSP15: TTA ATG CGT GAA GAG GAA TCA	+	61	DBP80 exon 8
DBP80 GSP16: ATT ATA GCT TGA CCA ACG CTT ATG	-	65	DBP80 exon 8
DBP80 GSP17: GAA CTG CTG CTT GGC TTG C	+	64	DBP80 exon 9
DBP80 GSP18: ATA TTT GTA GTG ATA AGC ACC TTC	-	62	DBP80 exon 9
DBP80 GSP19: GAT ATT GAA CAA TTA CAA GTT GTA	+	60	DBP80 exon 10?
DBP80 GSP20: CGA TTT GCC CGT AGA CCT T	+	63	DBP80 exon 10?
DBP80 GSP21: TTC CAA ACC TCC CAG TCC G	-	64	DBP80 exon 10?
DBP80 GSP22: GGA ATC GCC ATA AAT CTT ATA ACG	+	65	DBP80 exon 10?
DBP80 GSP23: CTA TAT CAT CGG CAC TAT CGG TAT	-	66	DBP80 exon 10?
DvDBP GSP 01: AGC TGA GAT TTG CTG TTC G	+	60.3	DvDBP CDS
DvDBP GSP 02: ACA GGA CAT TCG TGG TGA T	+	59.2	DvDBP CDS
DvDBP GSP 03: GTT CAT TGT CCT TGC TGG C	-	62.7	DvDBP CDS
DvDBP GSP 04: ATC ATT CCC CTC ATT TAG C	+	58.2	DvDBP CDS
DvDBP GSP 05: GGA CGA GGC TAA AAG GGA A	+	63.6	DvDBP CDS
DvDBP GSP 06: AAA TGT TGC TCA CCT AAT GC	-	59.6	DvDBP CDS
DvDBP GSP 07 GCG TAA ATG CCC TTC AGT AGT	-	63	DvDBP CDS

APPENDIX 3 (continued) PRIMER SEQUENCES

5'-3' seq & name	+/-	Tm	Location
DvDBP GSP 08: CGG ATC TGC AAG GAG TGT T	-	62	DvDBP CDS
DvDBP GSP10: CTG GTG GTG TTG CTG CGA A		64	5' DvDBP80 upstream
DvDBP GSP11: TAT GTA TCC GCC ACC TCT GC	+	64	Seq downstr. vDbp80
E20 GSP 01: TCA AAC TCG GCT TCT TCA TAG		59	vGrip63
E20 GSP 02: ATA ACC GGA AGC GGC AAC C		61	vGrip63
E20 GSP 03: ACA TGG ACA CAA CAT CGG AT	+	60	vGrip63
E20 GSP 04: CGC CAA GTT TCG GGA GAC C	+	63	vGrip63
E20 GSP 05: AAG CCG TTT GTT TTG GAT GA	+	59	vGrip63
E20 GSP 06: AGG CAA TGG TGG GAA GAA C	-	61	vGrip63
E21 GSP 01: ATT TCC GTT AGC ATA CAC CT		59	vG7129
E21 GSP 02: GTG GGA AAT AAT CTG CCT AAG		60	vG7129
E21 GSP 03: TGG GGA ATG TCC TTG ATG AAC	+	63	vG7129
E22 GSP 01: CGA GAA GAC CAG CGT TGC C	+	63	vG7129
E23 GSP 01: GGC TGT TGC TGT TGC TCT TG	+	65	vG7129
3058 BGL GSP01 AGG TTT GCG TGC GTT AGA G	-	63	RpL15 genomic upstream
3058 BGL GSP02: GCT ATG CGT TAT TTC TAT GC	-	58	RpL15 genomic upstream
3058 HIII GSP01 CCA ATT TGA TGG TGG GAA G	-	63	RpL15 genomic upstream
3058 HIII GSP02 TAA CAA CAG TCA GAG CCA GA	-	58	RpL15 genomic upstream
3058 Rp01: GAA TAC GCA GTG AAA CGA ATG A	+	64	RpL15 genomic upstream
3058 Rp02: AGC TTT ACT CAA ACG GAT TAG A	+	59	RpL15 genomic upstream
3058 Rp03: TCA TAT CTG GAG GCA CTG AGT	+	61	RpL15 genomic upstream
3058 Rp04: AGA TCT CAG CAT ACA GGT TCT	+	57	RpL15 genomic upstream
3058 Rp05: CGG GAA AAC GAA ACA GGC AG	+	69	RpL15 genomic upstream
3058 Rp06: GGA GTC TTG CCG AAA AAT GGT	-	66	RpL15 genomic downst
3058 Rp07: GGA GTC TTG CCG AAA AAT GGT	-	65	RpL15 genomic downst
3058 Rp08: TCT CGC AGT CGC CAG CAG T	-	70	RpL15 genomic downst
3058 Rp09: TCT GGC CAT GCT CGT TTG AAC GTA	-	73	RpL15 genomic downst
3058 Rp10: CGT ATC GCA ATT ACT ATT CGA G		65	RpL15 genomic upstream

REFERENCES

Adams, M.D. et.al. (2000) The Genome sequence of *Drosophila melanogaster*. *Science* 287:2185-2195

Agudo, M., Losada, A., Abad, J.P., Pimpinelli, S., Ripoll, P. and Villasante, A. (1999) Centromeres from telomeres? The centromeric region of the Y chromosome of *Drosophila melanogaster* contains a tandem array of telomeric HeT-A- and TART-related sequences. *Nucl. Acids Res.* 27:3318-3324

Ahmad, K. and Henikoff, S. (2001) Centromeres are specialized replication domains in heterochromatin. *J. Cell. Biol.* 153(1):101-109

Aitchison, J.D. and Rout, M.P. (2000) The road to ribosomes: filling potholes in the export pathway. *J. Cell. Biol.* 151(5):F23-F26

Akhtar, A., Zink, D. and Becker, P.B. (2000) Chromo domains are protein-RNA interaction modules. *Nature* 407:405-409

Arkhipova, I.R. (2001) Transposable elements in the animal kingdom. *Mol. Biol.* 35(2):157-167

Ashburner, M. (1989). *Drosophila*: A laboratory Manual. Cold Spring Harbour Laboratory Press

Atchison, M.L., Meyuhas, O. and Perry, R.P. (1989) Localization of transcriptional regulatory elements and nuclear factor binding sites in mouse ribosomal protein gene rpL32. *Mol. Cell. Biol.* 9(5):2067-2074

Audibert, A. and Simonelig, M. (1998) Autoregulation at the level of mRNA 3' end formation of the suppressor of forked gene of *Drosophila melanogaster* is conserved in *Drosophila virilis*. *Proc. Natl. Acad. Sci.* 95:14302-14307

Baldwin, M.C. and Suzuki, D.T. (1971) A Screening procedure for detection of putative deletions in proximal heterochromatin of *Drosophila*. *Mutation res.* 11:203-213

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, A.C., (2001) Selective recognition of methylated lysine 9 on Histone H3 by the HP1 chromo domain. *Nature* 410:120-124

Barakat,A., Szick-Miranda, K., Chang, I., Guyot, R., Blanc, G., Cooke, R., Delseny, M., and Bailey-Serres, J. (2001) The organization of cytoplasmic ribosomal protein genes in the *Arabidopsis* genome. *Plant Physiology* 127:398-415

Baron, M., Aslam, H., Flasza, M., Fostier, M., Higgs, J.E., Mazaleyrat, S.L. and Wilkin, M.B. (2002) Multiple levels of Notch signal regulation. *Mol. Membr. Biol.* 19:27-38

Bateson, W. (1894) Materials for the study of variation treated with especial regard to discontinuity in the origin of species. Macmillan, London.

Belfort, M., Reaban, M.E., Coetzee, T. and Dalgaard, J.Z. (1995) Prokaryotic introns and inteins: a panoply of form and function. *J. Bact.* 177(14):3897-3903

Bergman, C.M. and Kreitman, M. (2001) Analysis of conserved non-coding DNA in *Drosophila* reveals similar constraints in intergenic and intronic sequences. *Genome Res.* 11:1335-1345

Bi, X. and Broach, J.R. (1999) UASrpg can function as a heterochromatin boundary element in yeast. *Genes & Dev.* 13:1089-1101

Biggs, W.G.III., Zavitz, K.H., Dickson, B., var der Straten, A., Brunner, D., Hafen, E. and Zipursky, S.L. (1994) The *Drosophila rolled* locus encodes a MAP kinase required in the sevenless signal transduction pathway. *Embo J.* 13:1628-1635

Bolshakov, V.N., Topalis, P., Blass, C., Kokoza, E., della Torre, A., Kafatos, F.C. and Louis, C. (2002) A comparative genomic analysis of two distant Diptera, the fruit fly *Drosophila melanogaster* and the malaria mosquito, *Anopheles gambiae*. *Genome Res.* 12:57-66

Bone, J.B. and Kuroda, M.I. (1996) Dosage compensation regulatory proteins and the evolution of sex chromosomes in *Drosophila*. *Genetics* 144:705-713

Boudet, N., Aubourg, S., Toffano-Nioche, C., Kreis, M. and Lecharny, A. (2001) Evolution of intron-exon structure of DEAD helicase family genes in Arabidopsis, Caenorhabditis, and *Drosophila*. *Genome Res.* 11:2101-2114

Bowen, N.J. and Jordan, I.K. (2002) Transposable elements and the evolution of eukaryotic complexity. *Curr. Issues Mol. Biol.* 4:65-76

Brown, S.J., Rhoads, D.D., Stewart, M.J., van Slyke, B., Chen, I.T., Johnson, T.K., Denell, R.E. and Roufa, D.J. (1988) Ribosomal protein S14 is encoded by a pair of highly conserved adjacent genes on the X chromosome of *Drosophila melanogaster*. *Mol. Cell. Biol.* 8:4314-4321

Casacuberta, E. and Pardue, M. (2003) Transposon telomeres are widely distributed in the *Drosophila* genus: TART elements in the *virilis* group. *Proc. Natl. Acad. Sci.* 100(6):3363-3368

Castillo-Davis, C.I., Mekhedov, S.L., Hartl, D.L., Koonin, E.V. and Kondrashov, F.A. (2002) Selection for short introns in highly expressed genes. *Nature Genetics* 31:415-418

Celniker, S.E. et.al. (2002) Finishing a whole genome shotgun: Release 3 of the *Drosophila melanogaster* euchromatic genome sequence. *Genome Biology* 3(12): research0079.1-0079.14

Cho, G. and Doolittle, R.F. (1997) Intron distribution in ancient paralogs supports random insertion and not random loss. *J. Mol. Evol.* 44:573-584

Chung, S. and Perry, R.P. (1989) Importance of introns for expression of mouse ribosomal protein gene rpL32. *Mol. Cell. Biol.* 9(5):2075-2082

Clegg, N.J., Honda, B.M., Whiteheaad, I.P., Grigliatti, T.A., Wakimoto, B., Brock, H.W., Lloyd, V.K. and Sinclair, D.A.R. (1998) Suppression of position effect variegation in *Drosophila melanogaster* affect expression of the heterochromatic gene *light* in the absence of a chromosome rearrangement. *Genome* 41:495-503

Colot, H.V., Hall, J.C. and Rosbash, M. (1988) Interspecific comparisons of the period gene of *Drosophila* reveals large blocks of non-conserved coding DNA. *Embo J.* 7(12):3929-3937

Cramton, S.E. and Laski, F.A. (1994) string of pearls encodes *Drosophila* ribosomal protein S2, has Minute-like characteristics, and is required during oogenesis. *Genetics* 137:1039-1048

Csink, A.K. and Henikoff, S. (1998) Something from nothing: the evolution and utility of satellite repeats. *Trends Genet.* 14(5):200-204

Cuvier, O., Hart, C.M., Käs, E. and Laemmli, U.K. (2002) Identification of a multicopy chromatin boundary element at the borders of silenced chromosomal domains. *Chromosoma* 110:519-531

de la Cruz, J., Kressler, D. and Linder, P. (1999) Unwinding RNA in *Saccharomyces cerevisiae*: DEAD-box proteins and related families. *Trends Bioch. Sci.* 24:192-198

Dermitzakis, E.T., Bergman, C.M. and Clark, A.G. (2003) Tracing the evolutionary history of *Drosophila* regulatory regions with models that identify transcription factor binding sites. *Mol. Biol. Evol.* 20(5):703-714

Devlin, R.H., Holm, D.G., Morin, K.R. and Honda, B.M. (1990a) Identifying a single copy DNA sequence associated with the expression of a heterochromatic gene, the *light* locus of *Drosophila melanogaster*. *Genome* 33:405-415

Devlin, R.H., Bingham, B. and Wakimoto B.T. (1990b) The organization and expression of the *light* gene, a heterochromatic gene of *Drosophila melanogaster*. *Genetics* 125:129-140

Dimitri, P. (1999) Revising the selfish DNA hypothesis. *Trends in Genetics* 15(4):123-124

Dorer, D., Anane-Firempong, A. and Christensen, A.C. (1991) Ribosomal protein S14 is not responsible for the Minute phenotype associated with the *M(1)7C* locus in *Drosophila melanogaster*. *Mol. Gen. Genet.* 230:8-11

Dorer, D. and Henikoff, S. (1994) Expansions of transgene repeats cause heterochromatin formation and gene silencing in *Drosophila*. *Cell* 77:993-1002

Dorn, R., Kraus, V., Reuter, G. and Saumweber, H. (1993a) The enhancer of position-effect variegation of *Drosophila E(var)3-93D* codes for a chromatin protein containing a conserved domain common to several transcriptional regulators. *Proc. Natl. Acad. Sci.* 90:11376-11380

Dorn, R., Morawietz, H., Reuter, G. and Saumweber, H. (1993b) Identification of an essential *Drosophila* gene that is homologous to the translation initiation factor eIF-4A of yeast and mouse. *Mol. Gen. Genet.* 237:233-240

Dudick, M.E., Wright, T.R. and Brothers, L.L. (1974) The developmental genetics of the temperature sensitive lethal allele of the suppressor of forked, *l(1)su(f)ts67g*, in *Drosophila melanogaster*. *Genetics* 76(3):487-510

Eberl, D.F., Duyf, B.J. and Hilliker, A.J. (1993) The role of heterochromatin in the expression of a heterochromatic gene, the *rolled* locus of *Drosophila melanogaster*. *Genetics* 134:277-292

Eberl, D.F., Lorenz, L.J., Melnick, M.B., Sood, V., Lasko, P. and Perrimon, N. (1997) A new enhancer of position-effect variegation in *Drosophila melanogaster* encodes a putative RNA helicase that binds chromosomes and is regulated by the cell cycle. *Genetics* 146(3):951-963

Eisen, A., Sattah, M., Gazitt, T., Neal, K., Szauter, P., Lucchesi, J. (1998) A Novel DEAD-box RNA helicase exhibits high sequence conservation from yeast to humans. *Bioch. Biophys. Acta* 1397:131-136

Eisen, J.A. (1998) A phylogenomic study of the MutS family of proteins. *Nucl. Acids. Res.* 26(18):4291-4300

Eissenberg, J.C. and Elgin, S.C.R. (2000) The HP1 protein family: getting a grip on chromatin. *Curr.Opin.Genet. Dev.* 10:204-210

Elgin, S.C.R. (1996) Heterochromatin and gene regulation in *Drosophila*. *Curr.Opin.Genet.Dev.*6:193-202

Eskes, R., Yang, J., Lambowitz, A.M. and Perlman, P.S. (1997). Mobility of yeast group II introns: engineering a new site-specificity and retrohoming via full reverse splicing. *Cell.* 88:865-874

Evgen'ev, M.B., Zelentsova, H., Poluectova, H., Lyozin, G.T., Veleikodvorskaja, V., Pyatkov, K.I., Zhivotovsky, L.A. and Kidwell, M.G. (2000) Mobile elements and chromosomal evolution in the *virilis* group of *Drosophila*. *Proc. Natl. Acad. Sci.* 97(21):11337-11342

Evgen'ev, M.B., Zelentzova, H., Shostak, N., Kozitsina, M., Barskyi, V., Lankenau, D. and Corces, V.G. (1997) Penelope, a new family of transposable elements, and its possible role in hybrid dysgenesis in *Drosophila virilis*. *Proc. Natl. Acad. Sci.* 94:196-201

Fanti, L. and Pimpinelli, S. (1999) The peculiar organization of telomeres in *Drosophila melanogaster. Gene Therap. Mol. Biol.*4:1-10

Felsenfeld, A.L. and Kennison, J.A. (1995) Positional signaling by hedgehog in *Drosophila* imaginal disc development. *Development* 121:1-10

Fujiyama, A., Watanabe, H., Toyoda, A., Taylor, T.D.,Itoh, T., Tsai, S.F., Park, H.S., Yaspo, M.L., Lehrach, H., Chen, Z., Fu, G., Saitou, N., Osoegawa, K., de Jong, P.J., Suto, Y., Hattori, M. and Sakaki, Y. (2002) Construction and analysis of a human-chimpanzee comparative clone map. *Science* 295:131-134

Gatfield, D., Le Hir, H., Schmitt, C., Braun, I.C., Köcher, T., Wilm, M. and Izaurralde E. (2001) The DExH/D box protein HEL/UAP56 is essential for mRNA nuclear export in *Drosophila. Curr. Biol.* 11:1716-1721

Ghosh D. (2000) Object-oriented transcription factors database (ooTFD). *Nucl. Acids Res.* 28:308-310

Gray, Y.H.M., Tanaka, M.M. and Sved, J.A. (1996) P-element induced recombination in *Drosophila melanogaster*: hybrid element insertion. *Genetics* 144:1601-1610

Gunawardane, R.N., Martin, O.C., Cao, K., Zhang, L., Dej, K., Iwamatsu, A. and Zheng, Y. (2000) Characterization and reconstitution of *Drosophila* gamma-tubulin ring complex subunits. *J. Cell. Biol.* 151(7):1513-1523

Hanai, S., Uchida, M., Kobayashi, S., Miwa, M. and Uchida, K. (1998) Genomic organization of *Drosophila* poly (ADP-ribose) polymerase and distribution of its mRNA during development. *J. Biol. Chem.* 273(19):11881-11886

Hariharan, N. and Perry, R.P. (1990) Functional dissection of a mouse ribosomal protein promoter: significance of the polypyrimidine initiator and an element in the TATA-box region. *Proc. Natl. Acad. Sci.* 87:1526-1530

Hart, K., Klein, T., and Wilcox, M. (1993) A Minute encoding a ribosomal protein enhances wing morphogenesis mutants. *Mech. Dev.* 43:101:110

Hatton, L.S. and O'Hare, K. (1999). Deficiency mapping of genes in *Drosophila* using single embryo PCR. *Tech. Tips Online* 1:TO1816

Heitz, E. (1928) Das Heterochromatin der Moose I. *Jahrbeucher Wiss.Bot.* 69:762-818

Hilliker, A.J. (1976) Genetic analysis of the centromeric heterochromatin of chromosome 2 of *Drosophila melanogaster*: deficiency mapping of EMS induced lethal complementation groups. *Genetics* 83:765-782

Hilliker, A.J. and Holm, D.G. (1975) Genetic analysis of the proximal region of chromosome 2 of *Drosophila melanogaster*. I. Detachment products of compound autosomes. *Genetics* 81:705-721

Howe, M., Dimitri, P., Berloco, M. and Wakimoto, B.T. (1995) *Cis*-effects of heterochromatin on heterochromatic and euchromatic gene activity in *Drosophila melanogaster*. *Genetics* 140:1033-1045

Hoskins, R.A., Smith, C.D., Carlson, J.W.,Carvalho, A.B., Halpern, A., Kaminker, J.S., Kennedy, C., Mungall, C.J., Sullivan, B.A., Sutton, G.G., Yasuhara, J.C., Wakimoto, B.T., Myers, E.W., Celniker, S.E., Rubin, G.M., and Karpen, G.H. (2002) Heterochromatic sequences in a *Drosophila* whole-genome shot-gun assembly. *Genome Biology* 3(12): research0085.1-0085.16

Huang, S.L. and Baker, B.S. (1976) The mutability of the Minute loci of *Drosophila melanogaster* with ethyl methanesulfonate. *Mut. Res.*34:407-414

Hwang, K., Eissenberg, J.C., and Worman, H.J. (2001) Transcriptional repression of euchromatic genes by *Drosophila* Heterochromatin Protein 1 and histone modifiers. *Proc. Natl. Acad. Sci.* 98(20):11423-11427

Jenuwein, T. and Allis, D. (2001) Translating the histone code. *Science* 293:1074-1080

Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., Ashburner, M., and Celniker, S.E. (2002) The transposable elements of *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biology* 3(12): research/0084.1-0084.19

Kay, M.A., Zhang, J.Y., and Jacobs-Lorena, M. (1988) Identification and germline transformation of the ribosomal protein rp21 gene of *Drosophila*: Complementation analysis with the Minute QIII locus reveals nonidentity. *Mol. Gen. Genet.* 213:354-358

Kennison, J.A. and Tamkun, J.W. (1988) Dosage-dependent modifiers of Polycomb and Antennapedia mutations in *Drosophila*. *Proc. Natl. Acad. Sci.* 85:8136-8140

Koc, E.C., Burkhart, W., Blackburn, K., Moseley, A., Koc, H. and Spremulli, L.L. (2000) A proteomics approach to the identification of mammalian mitochondrial small subunit ribosomal proteins. *J. Biol. Chem.* 275(42):32585-32591

Kongsuwan, K., Yu, Q., Vincent, A., Frisardi, M.C., Rosbash, M., Lengyel, J.A. and Merriam, J. A (1985) *Drosophila* Minute encodes a ribsomal protein. *Nature* 317:555-558

Koryakov, D.E., Zhimulev, I.F. and Dimitri, P. (2002) Cytogenetic analysis of the third chromosome of *Drosophila melanogaster*. *Genetics* 160:509-517

Kress, H. (1993) The salivary gland chromosomes of *Drosophila virilis*: a cytological map, pattern of transcription and aspects of chromosome evolution. *Chromosoma* 102:734-742

Kurek, R., Reugels, A.M., Glätzer, K.H. and Bünemann, H. (1998) The Y *Chromosoma*l fertility factor Threads in *Drosophila* hydei harbours a functional gene encoding and axonemal Dynein beta heavy chain protein. *Genetics*. 149:1363-1376

Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., Jenuwein, T. (2001) Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* 410:116-120

Lambertsson, A. (1998) The Minute genes in *Drosophila* and their molecular functions. *Advances in Genetics* 38:69-134

Lasko, P. (2000) The *Drosophila melanogaster* genome: translation factors and RNA binding proteins. *J. Cell. Biol.* 150(2):F51-56

Levy, S., Avni, D., Hariharan, N., Perry, R.P., Meyuhas, O. (1991) Oligopyrimidine tract at the 5' end of mammalian ribosomal protein mRNAs is required for their translational control. *Proc. Natl. Acad. Sci.* 88:3319-3323

Lindsley D.L. and Hardy R.W. (1992) Cytology of In(3LR)TM8 and In(3L)TM9. *D.I.S.* 71:154

Lintermann, K., Roth, G.E., King-Jones, K., Korge, G. and Lehmann, M. (1998) Comparison of the GAGA factor genes of *Drosophila melanogaster* and *Drosophila virilis* reveals high conservation of GAGA factor structure beyond the BTB/POZ and DNA binding domains. *Dev. Genes Evol.* 208:447-456

Lohe, A.R., Hilliker, A.J. and Roberts, P.A. (1993) Mapping simple repeated DNA sequences in heterochromatin of *Drosophila melanogaster*. *Genetics* 134:1149-1174

Losada, A., Agudo, M., Abad, J.P. and Villasante, A. (1999) HeT-A telomere-specific retrotransposons in the centric heterochromatin of *Drosophila melanogaster* chromosome 3. *Mol. Gen. Genet.* 262:618-622

Lu, B.Y., Emtage, P.C.R., Duyf, B.J., Hilliker, A.J., and Eissenberg J.C. (2000) Heterochromatin Protein 1 is required for the normal expression of two heterochromatic genes in *Drosophila*. *Genetics* 155:699-708

Lucchesi, J.C. (1998) Dosage compensation in flies and worms: the ups and downs of X chromosome regulation. *Curr. Opin. Genet. Dev.* 8:179-184

Lyamouri, M., Enerly, E., Kress, H. and Lambertsson, A. (2002) Conservation of gene order, structure and sequence between three closely linked genes in *Drosophila melanogaster* and *Drosophila virilis*. *Gene* 282:199-206

Mager, W.H. (1988) Control of ribosomal protein gene expression. *Biochim. Biophys. Acta* 949:1-15

Marchant, G.E. and Holm, D.G. (1988a) Genetic analysis of the heterochromatin of chromosome 3 in *Drosophila melanogaster*. I. Products of compound-autosome detachment. *Genetics* 120:503-517

Marchant, G.E., Holm D.G. (1988b) Genetic analysis of the heterochromatin of chromosome 3 in *Drosophila melanogaster*. II. Vital loci identified through EMS mutagenesis. *Genetics* 120:519-532

Mardis, E., McPherson, J., Martienssen, R., Wilson, R.K. and McCombie, W.R. (2002) What is finished and why does it matter. *Genome Res.* 12:669-671

McCollum, A.M., Ganko, E.W., Barrass, P.A., Rodriguez, J.M. and McDonald, J.F (2002) Evidence for the adaptive significance of an LTR retrotransposon sequence in a *Drosophila* heterochromatic gene. *BMC Evol. Biol.* 2:5-13

Miklos, G.L.G. and Rubin, G.M. The role of the genome project in determining gene function: insights from model organisms. *Cell* 86:521-529

Morgenstern, B. (1999). DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* 15:211-218

Moriyama, E.N., Petrov, D.A. and Hartl, D.L. (1998) Genome size and intron size in *Drosophila*. *Mol. Biol. Evol.* 15(6):770-773

Mount, S.M., Burks, C., Hertz, G., Stormo, G.D., White, O., and Fields, C. (1992) Splicing signals in *Drosophila*: intron size, information content and consensus sequences. *Nucl. Acids Res.* 20(16):4255-4262

Muller, H.J. (1930) Types of visible variations induced by X-rays in *Drosophila*. *J.Genet.* 22:299-334

Muller, H.J. (1940) Bearings of the *Drosophila* work on Systematics. *New Systematics* 185-268

Neufeld, T.P., Carthew, R.W. and Rubin, G.M. (1991) Evolution of gene position: Chromosomal arrangement and sequence comparison of the *Drosophila melanogaster* and *Drosophila virilis* sina and Rh4 genes. *Proc. Natl. Acad. Sci.* 88:10203-10207

Nielson, S.J., Schneider, R., Bauer, U., Bannister, A.J., Morrison, A., O'Carroll, D., Firestein, R., Cleary, M., Jenuwein, T., Herrera, R.E. and Kouzarides, T. (2001) Rb targets histone H3 methylation and HP1 to promoters. *Nature* 412:561-565

Nurminsky, D.I., Moriyama, E.N., Lozovskaya, E.R. and Hartl, D.L. (1996) Molecular phylogeny and genome evolution in the *Drosophila virilis* species group: duplications of the *Alcohol Dehydrogenase* gene. *Mol. Biol. Evol.* 13(1):132-149

O'Hare, K., and Rubin, G.M. (1983) Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell* 34:25-35

O'Neil, M.T. and Belote, J.M. (1992) Interspecific comparison of the transformer gene of *Drosophila* reveals an unusually high degree of evolutionary divergence. *Genetics* 131:113-128

Palmer, J.D., Adams, K.L., Cho, Y., Parkinson, C.L., Qiu, Y. and Song, K. (2000) Dynamic evolution of plant mitochondrial genomes: mobile genes and introns, and highly variable mutation rates. *Proc. Natl. Acad. Sci.* 97(13):6960-6966

Patterson, J.T., and Stone, W.S. (1952) Evolution in the genus *Drosophila*. The MacMillan Company, New York.

Pearson , W.R. (1990) Rapid and Sensitive Sequence Comparison with FASTP and FASTA. *Methods in Enzymology* 183:63-98

Petrov, D.A. (2002) DNA loss and the evolution of genome size in *Drosophila*. *Genetica* 115:81-91

Petrov, D.A., Schutzman, J.L., Hartl, D.L. and Lozovskaya, E.R. (1995) Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. *Proc. Natl. Acad. Sci.* 92:8050-8054

Piacentini L., Fanti, L., Berloco, M., Perrini, B. and Pimpinelli, S. (2003) Heterochromatin Protein 1 (HP1) is associated with induced gene expression in *Drosophila* euchromatin. *J. Cell. Biol.* 161(4):707-714

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

Pimpinelli, S., Santini, G. and Gatti, M. (1976) Characterization of *Drosophila* heterochromatin. II. C- and N-banding. *Chromosoma*. 57:377-386

Preston, C.R. and Engels, W.R. (1996) Flanking duplications and deletions associated with P-induced male recombination in *Drosophila*. *Genetics* 144:1623-1638

Qian, A., Hongo, S. and Jacobs-Lorena, M. (1988) Antisense ribosomal protein gene expression specifically disrupts oogenesis in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci.* 85:9601-9605

Ranz, J.M., Casals, F. and Ruiz, A. (2001) How malleable is the eukaryotic genome? Extreme rate of chromosomal rearrangement in the genus *Drosophila*. *Genome Res.* 11:230-239

Ranz, J.M., Segarra, C. and Ruiz, A. (1997) Chromosomal homology and molecular organization of Muller's elements D and E in the *Drosophila repleta* species group. *Genetics* 145:281-295

Reynaud, E., Bolshakov, V.N., Barajas, V., Kafatos, F.C. and Zurita, M. (1997). Antisense suppression of the putative ribosomal protein S3A gene disrupts ovarian development in *Drosophila melanogaster*. *Mol. Gen. Genet.* 256:462-467

Richards, E.J. and Elgin, S.C.R. (2002) Epigenetic Codes for heterochromatin formation and silencing: rounding up the usual suspects. *Cell*. 108:489-500

Risinger, C., Deitcher, D.L., Lundell, I., Schwarz, T.L. and Larhammer, D. (1997) Complex gene organization of synaptic protein SNAP-25 in *Drosophila melanogaster*. *Gene* 194:169-177

Ritossa, F.M., Atwood, K.C. and Speigelman, S. (1966) A molecular explanation of the bobbed mutants of *Drosophila* as partial deficiencies of "ribosomal" DNA. *Genetics* 54:819-834

Robert, V., Prud'homme, N., Kim, A., Bucheton, A. and Pélisson, A. (2001) Characterization of the *flamenco* region of the *Drosophila melanogaster* genome. *Genetics* 158:701-713

Roberts, D.B. Ed. (1998) *Drosophila*: a practical approach. Second edition, Oxford University Press.

Robertson, H.M., Preston, C.R., Phillis, R.W., Johnson-Schlitz, D.M., Benz, W.K. and Engels, W.R. (1988) A stable genomic source of P element transposase in *Drosophila melanogaster*. *Genetics* 118:461-470

Rubin, G.M. (1986). Procedures for *Drosophila* Molecular Biology. (Non commercial publication)

Russo, C.A., Takezaki, N. and Nei, M. (1995) Molecular phylogeny and divergence times of Drosophilid species. *Mol. Biol. Evol.* 12 (3):391-404

Rzhetsky, A. and Ayala, F.J. (1999) The enigma of intron origins. *Cell Mol. Life Sci.* 55:3-6

Saebøe-Larssen, S. and Lambertsson, A. (1996) A novel *Drosophila Minute* locus encodes ribosomal protein S13. *Genetics* 143:877-885

Saebøe-Larssen, S., Lyamouri, M., Merriam, J., Oksvold, M.P. and Lambertsson, A. (1998) Ribosomal Protein insufficiency and the *Minute* syndrome in *Drosophila*: a dose-response relationship. *Genetics* 148:1215-1224

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: a laboratory manual. Second Edition. Cold Spring Harbour Laboratory Press.

Sass, G.L. and Henikoff, S. (1998) Comparative analysis of position-effect variegation mutations in *Drosophila melanogaster* delineates the targets of modifiers. *Genetics* 143:733-741

Scheinker, V.S., Lozovskaya, E.R., Bishop, J.G., Corces, V.G. and Evgen'ev, M.B. (1990) A long terminal repeat-containing retrotransposon is mobilized during hybrid dysgenesis in *Drosophila virilis*. *Proc. Natl. Acad. Sci.* 87: 9615-9619

Schmidt, A., Hollmann, M. and Schafer, U. (1996) A newly identified Minute locus M(2)32D, encodes the ribosomal protein L9 in *Drosophila melanogaster*. *Mol. Gen. Genet.* 251:381-387

Schmitt, C., von Kobbe, C., Bachi, A., Panté, N., Rodrigues, J.P., Boscheron, C., Rigaut, G., Wilm, M., Séraphin, B., Carmo-Fonseca M. and Izaurralde, E. (1999) Dbp5, a DEAD-box protein required for mRNA export, is recruited to the cytoplasmic fibrils of nuclear pore complex via a conserved interaction with CAN/Nup159p. *Embo J.* 18(15):4332-4347

Schotta, G., Ebert, A., Dorn, R. and Reuter, G. (2003) Position-effect variegation and the genetic dissection of chromatin regulation in *Drosophila*. *Sem.Cell. Dev. Biol.* 14:67-75

Schultz, J. (1929) The Minute reaction in the development of *Drosophila melanogaster*. *Genetics* 14:366-419

Schulze, S., Sinclair, D.A.R., Silva, E., Fitzpatrick, K.A., Singh, M., Lloyd, V.K., Morin, K.A., Kim, J., Holm, D.G., Kennison, J.A., and Honda, B.M. (2001) Essential genes in proximal 3L heterochromatin of *Drosophila melanogaster*. *Mol. Gen. Genet.* 264:782-789

Sinclair, D.A.R., Clegg, N.J., Antonchuk, J., Milne, T.A., Stankunas, K., Ruse, C., Grigliatti, T.A., Kassis, J.A. and Brock, H.W. (1998) *Enhancer of Polycomb* is a suppressor of position-effect variegation in *Drosophila melanogaster*. *Genetics* 148:211-220

Sinclair, D.A.R., Grigliatti, T.A. and Kaufman, T.C. (1984) Effects of a temperature-sensitive *Minute* mutation on gene expression in *Drosophila melanogaster*. *Genet. Res., Camb.* 43:257-275

Sinclair, D.A.R., Mottus, R.C. and Grigliatti, T.A. (1983) Genes which suppress position-effect variegation in *Drosophila melanogaster* are clustered. *Mol. Gen. Genet.* 191:326-333

Sinclair, D.A.R., Suzuki, D.T. and Grigliatti, T A. (1981) Genetic and Developmental analysis of a temperature sensitive *Minute* mutation of *Drosophila melanogaster*. *Genetics* 97:581-606

Singh, M., Silva, E., Schulze, S., Sinclair, D.A.R., Fitzpatrick, K.A. and Honda, B.M. (2000) Cloning and characterization of a new theta-class glutathione-S-transferase (GST) gene, *gst-3,* from *Drosophila melanogaster. Gene* 247:167-173

Snay-Hodge, C.A., Colot, H.V., Goldstein, A.L., Cole, C.N. (1998) Dbp5p/Rat8p is a yeast nuclear pore-associated DEAD-box protein essential for RNA export. *Embo J.* 17(9):2663-2676

Struhl, K. (1999) Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* 98:1-4

Sun, F-L., Cuaycong, M.H., Craig, C.A., Wallrath, L.L., Locke, J. and Elgin, S.C.R. (2000) The fourth chromosome of *Drosophila melanogaster*: interspersed euchromatic and heterochromatic domains. *Proc. Natl. Acad. Sci.* 97(10):5340-5345

Tamkun, J.W., Kahn, R.A., Kissinger, M., Brizuela, B.J., Rulka, C., Scott, M.P. and Kennison, J.A. (1991) The arflike gene encodes an essential GTP-binding protein in *Drosophila. Proc. Natl. Acad. Sci.* 88:3120-3124

Thompson, J.D., Higgins, D.G. and Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* 22:4673-4680

Thummel, C. (1993) Compilation of *Drosophila* cDNA and genomic libraries. *D.I.S.* 72:180-183

Torok, I., Herrmann-Horle, D., Kiss, I., Tick., G., Speer, G., Schmitt, R. and Mechler, B.M. (1999) Down-regulation of *RpS21*, a putative Translation initiation factor interacting with *P40*, produce viable *Minute* imagos and larval lethality with overgrown hematopoietic organs and imaginal discs. *Mol. Cell. Biol.* 19(3):2308-2321

Tseng, S.S.I., Weaver, P.L., Liu, Y., Hitomi, M., Tartakoff, A.M. and Chang, T. (1998) Dbp5p, a cytosolic RNA helicase is required for poly(A)+ RNA export. *Embo J.* 17(9):2651-2662

Tulin, A., Stewart, D., Spradling A.C. (2002) The *Drosophila* heterochromatic gene encoding poly(ADP-ribose) polymerase (PARP) is required to modulate chromatin structure during development. *Genes & Dev.*16:2108-2119

Vieira, J., Vieira, C.P., Hartl, D.L. and Lozovskaya, E.R. (1997a) A framework physical map of *Drosophila virilis* based on P1 clones: applications in genome evolution. *Chromosoma.* 106:99-107

Vieira, J., Vieira, C.P., Hartl, D.L. and Lozovskaya, E.R. (1997b) Discordant rates of chromosome evolution in the *Drosophila virilis* species group. *Genetics* 147:223-230

Vilardell, J.and Warner, J.R. (1994) Regulation of splicing at an intermediate step in the formation of the spliceosome. *Genes & Dev.* 8:211-220

Vilinsky, I., Stewart, B.A., Drummond, J., Robinson ,I and Deitcher, D.L. (2002) A *Drosophila SNAP-25* null mutant reveals context dependent redundancy with *SNAP-24* in neurotransmission. *Genetics* 162:259-271

Voelker, R.A., Huang, S.M., Wisely, G.B., Sterling, J.F., Bainbridge, S.P. and Hiraizumi, K. (1989) Molecular and genetic organization of the *suppressor of sable* and *Minute (1) 1B* region in *Drosophila melanogaster. Genetics* 122:625-642

Volpe, T.A., Kidner, C., Hall, I.M., Teng, G., Grewal, S.I.S. and Martienssen, R.A. (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. *Science* 297:1833-1837

Warren, W.D., Lin, E., Nheu, T.V., Hime, G.R. and McKay, M.J. (2000) Drad21, a *Drosophila rad21* homologue expressed in S-phase cells. *Gene* 250:77-84

Weijers, D., Franke-van Dijk, M., Vencken, R., Quint, A., Hooykaas, P. and Offringa, R. (2001) An *Arabidopsis* Minute-like phenotype is caused by a semi-dominant mutation in a Ribosomal protein S5 gene. *Development* 128:4289-4299

Weiler, K.S. and Wakimoto, B.T. (1995) Heterochromatin and gene expression in *Drosophila. Annu. Rev. Genet.* 29:577-605

Wildman, D.E. (2002) A map of the common chimpanzee genome. *Bioessays* 24: 490-493

Wittkopp, P.J., Vaccaro, K. and Carroll, S.B. (2002) Evolution of *yellow* gene regulation and pigmentation in *Drosophila. Curr. Biol.* 12:1547-1556

Wool, I.G. (1996) Extraribosomal functions of ribosomal proteins. *Trends Bioch. Sci.* 21(5):164-165

Wustmann, G., Szidonya, J., Taubert, H. and Reuter, G. (1989) The *Genetics* of position-effect modifying loci in *Drosophila melanogaster. Mol. Gen. Genet.* 217:520-527

Yacoub, A., Augeri, L., Kelley, M.R., Doetsch, P.W. and Deutsch, W.A. (1996) A *Drosophila* ribosomal protein contains 8-oxoguanine and abasic site DNA repair activities. *Embo J.* 15:2306-2312

Yacoub, A., Kelley, M.R. and Deutsch, W.A. (1996) *Drosophila* ribosomal protein PO contains apurinic/apyrimidinic endonuclease activity. *Nucl. Acids Res.* 24:4298-4303

Yokokura, T., Tei, H. and Yamamoto, D. (1993) Sequence and expression of a gene encoding a ribosomal protein S4 homolog from *Drosophila melanogaster.* *Gene* 132(2): 251-254

Yoshihama M., Uechi, T., Asakawa, S., Kawasaki, K., Kato, S., Higa, S., Maeda, N., Minoshima, S., Tanaka, T., Shimizu, N., and Kenmochi, N. (2002) The human ribosomal protein genes: sequencing and comparative analysis of 73 genes. *Genome Res.* 12:379-390

Yu, Q., Qiu, R., Foland, T.B., Griesen, D., Galloway, C.S., Chiu, Y., Sandmeier, J., Broach, J.R. and Bi, X. (2003) Rap1p and other transcriptional regulators can function in defining distinct domains of gene expression. *Nucl. Acids Res.* 31(4): 1224-1233

Zaffran, S., Chartier, A., Gallant, P., Astier, M., Arquier, N., Doherty, D., Gratecos, D. and Sémériva, M. (1998) A *Drosophila* RNA helicase gene, *pitchoune*, is required for cell growth and proliferation, and is a potential target of d-Myc. *Development* 125: 3571-3584

Zelentsova, H., Poluectova, H., Mnjoian, L., Lyozin, G., Veleikodvorskaja, V., Zhivotovsky, L., Kidwell, M.G. and Evgen'ev, M.B. (1999) Distribution and evolution of mobile elements in the *virilis* species group of *Drosophila*. *Chromosoma* 108:443-456

Zhan, H., Liu ,D. and Liang, C. (2001) Insulator: from chromatin domain boundary to gene regulation. *Hum. Genet.* 109:471-478

Zhang, Z., Harrison, P. and Gerstein, M. (2002) Identification and analysis of over 2000 ribosomal protein pseudogenes in the human genome. *Genome Res.* 12:1466-1482

Zhao, J., Jin, S., Björkroth, B., Wieslander, L. and Daneholt, B. (2002) The mRNA export factor Dbp5 is associated with Balbiani ring mRNP from gene to cytoplasm. *Embo J.* 21(5):1177-1187