Characterisation of the genomic region in and around the *shaking-B* locus of *Drosophila melanogaster*.

By

Alan Griffin

Thesis submitted for the degree of Doctor of Philosophy

Department of Genetics

University of Glasgow

1995.

Contents:

List of Contents.

Chapter One - Introduction.

Chapter Two - Materials and Methods.

Chapter Three - Cloning of the *shak-B* Region and Localisation of Several Chromosomal Aberration Endpoints.

Chapter Four - Characterisation of Repetitive DNA From the 19E Region.

Chapter Five - Identification and Characterisation of Transcribed Regions Within the Chromosomal Walks.

Chapter Six - Analysis of Several Transcribed Regions from the 952 (19E3) Region.

Chapter Seven - Discussion.

Appendices.

List of Figures.

Chapter One.

Chapter Three.

Chapter Four.

Chapter Five.

Chapter Six.

List of Tables.

Chapter Two.

Chapter Four.

Chapter Five.

Chapter Six.

Abbreviations.

B1	*Bam* HI
bp	Base pairs
BSA	Bovine Serum Albumen
Df	Deficiency
DMF	Dimethylformamide
Dp	Duplication
DTT	Dithiothreitol
EMS	Ethyl-methanesulphonate
HIII	*Hin* dIII (restriction enzyme).
IPTG	Isopropyl-b-thiogalctose
kb	Kilobase pairs
MOPS	Morpholino propane sulphuric acid
O.R.F.s	Open Reading Frames
p.s.i.	pounds per square inch
RI	*Eco* RI (restriction enzyme).
RNase	Ribonuclease
S1	*Sal* I (restriction enzyme).
SDS	Sodium dodecyl sulphate
TEMED	N,N,N', N-tetramethylethylenediamine
Tris	Tris (hydroxy) aminoethane
Xgal	5-Bromo-4-chloro-3-indoyl-b-galactoside

Acknowledgements.

I would like to thank several people for their help and encouragement in this project. Firstly and foremost, Dr. Catherine Wild who has always been ready with her advice and encouragement on all subjects scientific (or otherwise!). To the 'original' members of the group, Dr. Claire Taylor and Irene Durant, who made the first years so enjoyable that I decided to stay on even when events seemed to be conspiring against me. Also thanks to those who joined later and made the rest of my Ph.D. liveable, especially Margaret Adam and Marian Wilkin. To those from other groups from the madhouse that is the genetics department, I would like to thank, Dr. Simon Thomlinson, who may one day give me back my £20, to Dr. Steven Goodwin for making me laugh and for Dr. Arvind Hundal for keeping me human(ish). I would also like to thank Dr. Eric Frei for sharing his (lack of!) results with *Passover* transcripts when Northern blotting with me and also those other 'friends' who have encouraged me and kept me informed about what's happening in the *Passover* world.

I would like to single out Dr. John Morrison from the Molecular Medicine Unit at St. James in Leeds for special thanks for his help and encouragement in the completion of this thesis as well as Adel Mansur, David Robertson and of course Grange and Rosalin for making my Post Doc so much fun. To Michele Cummings for being a good neighbour and to Mr. Gwyn Jones for being a superhero. I must especially thank the ladies in the Prep. room especially Doris, Sadie, Janet, Mattie and of course Agnes, for taking an interest in other things besides the amount of fly food I wanted!

I would like to reserve special thanks for my parents, David and Vera Griffin, who can now say "At last you've submitted the bloody thing!" and to my Grandfather Harry Percival Turner, for helping me keep my feet on the ground and always being ready for a chat. As well as these I have to thank all the other members of my family who have taken an interest in my project and which part of the country I'm at now. The last person I would like to thank is Dr. Jane Davies, my project supervisor for giving me the chance to do this project in the first place and for teaching me several things about science and life.

The Road goes ever on and on
Down from the door where it began.
Now far ahead the Road has gone,
And I must follow, if I can.
Pursuing it with weary feet,
Until it joins some larger way,
Where many paths and errands meet.
And whither then? I cannot say.

J.R.R.T.

Abstract.

The *shaking-B* locus of *Drosophila melanogaster* (formerly known as *Passover*) is a complex locus situated at polytene subdivision 19E3, near the base of the X chromosome. Some alleles of *shak-B* are embryonic or first instar larval lethals, whilst others cause defects in the nervous system of the adult fly. To clone the *shak-B* locus, four chromosomal walks were performed. Three walks were initiated using clones generated by a microdissection of the *mal* to *su(f)* region at the base of the X chromosome whereas a fourth was started using a cloned fragment from the *runt* locus. The walks encompass over 400kb of cloned DNA.

Genetic analysis revealed *shak-B* to reside, at least in part, between the distal breakpoint of *Df(1)LB6* and the proximal breakpoint of *Df(1)16-3-35.* This study used deficiency and duplication mapping to localise these, and other breakpoints in the walks, allowing them to be orientated and positioned onto the genetic map of the region. In addition, a transcriptional analysis was performed using reverse Northern and Northern blots. This analysis identified 28 transcription units, a number of which were shown to be repetitive. These repetitive regions have been characterised and several found to have homology to transposable elements. A large number were shown to contain a short repetitive sequence, the *opa* repeat. These *opa* repeats are associated with a large number of important, developmentally regulated genes in many organisms.

The chromosomal walk that contains the *shak-B* locus (the 952 walk) was shown to contain less repetitive DNA and to have many more single copy transcribed regions than the others. The majority of these transcripts were analysed by Northern analysis. Transcripts were not detected in two cases. One of these is a region from which the 3' ends of *shak-B* transcripts originate. It is felt that the lack of signal from both regions was due either to the rarity of, or to the non-polyadenylated state of, both sets of transcripts. *In situ* hybridisations were performed using genomic fragments from a region of the 952 walk as a probe. This region, which is likely to contain part of the *shak-B* locus, gave discrete hybridisation to the lamina / retina border of the eye. Sequence was obtained from several transcribed genomic regions from the 952 walk, and several tentative homologies noted.

Chapter One

Chapter One. Introduction.

1.0 Introduction.

The nervous system is the most complex organ that animals posses. It is the means by which an animal is able to detect, decipher and act upon both exogenous stimuli, (such as the detection of light via the retina of the eye), and endogenous stimuli, (such as the control of respiration via the detection of CO_2 levels by the carotid and aortic bodies [Vander *et al.,* 1994]). The mechanisms by which the nervous system is able to process and then act upon and / or store information are not fully understood, though many advances have been made, (for example see reviews by Albright, 1991; Goodwin, 1993). One of the major problems that developmental biologists face, is explaining how such a complex organ with its enormous diversity of neurones and glia, develops from the morphologically indistinct cells of the early embryo to the complex fully functioning system of a mature eukaryote, be it in an invertebrate or a vertebrate system.

Drosophila has proved invaluable as an aid to understanding the molecular mechanisms that control this development in both the central nervous system (CNS) and the peripheral nervous system (PNS). This is largely because of the sophisticated genetics, *in vivo* transformation methods and other molecular genetic techniques that can be bought to bear on developmental questions. It has also been demonstrated that many of the mechanisms and cues used by simple organisms, such as insects, are often analogous to those occurring in higher organisms *e.g.* Goodman (1994). Therefore work in *Drosophila* can often be applied to studies on vertebrates. For example the homeobox genes originally described and cloned in *Drosophila melanogaster* (*e.g.* Bender *et al.,* 1983b;reviewed by Morata, 1993), have now been shown to control major events in the development of many species including mammals (reviewed by Scott *et al.,* 1989). Similarly, other genes known to play an important role in the development of the *Drosophila* nervous system are also known to be expressed in related tissues in other organisms. For example homologous genes to those in the *Drosophila achaete-scute* complex have been discovered in the mouse (*MASH1*, Del Franco *et al.,* 1993); in the frog, *Xenopus laevus* (*XASH1*, Ferreiro *et al.,* 1992); and in humans (*hASH1*, Ball *et al.,* 1993).

1.1 Development of the Nervous System in *Drosophila melanogaster.*

The development of the nervous system in *Drosophila* and most other organisms can be divided into a series of overlapping stages. Firstly, inductive events must occur which bring about the creation of tissue(s) that will go onto form the nervous system; secondly, neurones and their supporting cells, the glia, are born; thirdly, there is a generation of specific cell fates followed shortly thereafter by axonal outgrowth. Axonal outgrowth comprises the events beginning with the organising of the growth cone and its extension from the neuronal soma and ends when the growth cone makes contact with the final target cell(s)and ceases motility. This is followed finally by the formation of correct synaptic connections, which eventually results in the production of morphologically specialised contacts for transmitter release. Many of the gene products involved with control of development and / or regulation of the nervous system have complex expression patterns, whilst the genes themselves are often structurally quite complex, *e.g. connectin*; the neurogenic loci *etc.* (Lehmann *et al.,* 1983; Hartenstein and Campos-Ortega 1986; Jimenez and Campos-Ortega, 1990; Campos-Ortega and Haenlin, 1992; Greenwald and Rubin, 1992; Knust *et al.,* 1992; Ghysen *et al.,* 1993; Price *et al.,* 1993; Nose *et al.,* 1992; 1994; Busseau *et al.,* 1994; Ruohola-Baker *et al.,* 1994). It is thought likely that the locus currently being characterised in this laboratory, *shak-B* (formerly known as *Passover*) is involved with the final stages of nervous system development *i.e.* with events involved in synaptogenesis. It is a complex locus at both the genetic and the molecular level (see below).

1.2 The *shaking-B* Locus.

The neural mutation *Passover* (now renamed *shak-B*^{Pas}) was isolated in a screen for flies with defects in the visually induced escape response of *Drosophila*
The rational behind this approach was that a behavioural defect may reflect small changes in a well characterised neural circuit that mediates this response, namely the Giant Fibre System (GFS). The GFS (Figure 1.1 for schematic diagram) was chosen as the system to be mutagenised as it fulfilled several criteria for study; i) it has been extensively characterised, both anatomically and physiologically; ii) it has a small number of identifiable synapses; iii) the nerves are large enough for cytological examination; iv) the synapses are essential for a specific behaviour, the escape response. Therefore, a disruption of these synapses could be scored and characterised

relatively easily (King and Wyman, (1980); Rheuben and Kammer, (1980); Koto *et al.,* (1981); King and Tanyouye, 1983; Strausfeld and Bassemir, 1983; Thomas and Wyman, 1984b; Wyman *et al.,* 1985; Bacon and Strausfeld, 1986).

1.2.1 Morphology of the Giant Fibre System.

The GFS is a neuronal circuit composed of eight, well characterised neurones that mediate the escape response of *Drosophila.* (Analogous systems also exist in other species of fly [King and Valentino, 1983; Bacon and Strausfeld, 1986]). The Giant Fibres (GFs) which are so called because of their large size relative to other *Drosophila* neurones, are a bilaterally symmetrical pair of descending interneurones, the cell bodies of which are situated in the posterior, lower protocerebrum of the fly's brain (Koto *et al.,* 1981; King and Valentino, 1983). The GFs themselves were originally identified by Power (1948) and since this date the GFS has been studied using a variety of different approaches, (for example see Coggshall, 1978; Stausfeld and Bassemir, 1983).

Each GF descends through the cervical connective and enters the mesothoracic neuromere of the thoracic ganglion where each axon synapses with two identified neurones (King and Wyman, 1980). First, the GF forms an extensive contact with the ipsilateral peripherally synapsing interneurone (PSI). This exits the ganglion and enters the peripheral nerve where it makes reciprocal chemical synapses (Tanouye and Wyman, 1980; King and Wyman, 1980) with the five motor neurones (DLMns) of the contralateral dorsal longitudinal (DLM) muscles (Ikeda *et al.,* 1980). The DLM are the main wing depressor muscles that function to power the down stroke of the wing beat during flight. The GF turns posteromedially to contact the motorneurone (TTMn) of the ipsilateral tergotrochanteral muscle (TTM). The TTM is the largest 'twitch' muscle in the fly's body and provides the majority of the thrust for the jump, by extending the mesothoracic leg. This jump is the first part of the escape response. The TTMn also activates the wing elevator muscles. Electrophysiological studies (discussed below) suggest that the GF - TTMn synapses are electrical (Tanouye and Wyman, 1980).

Branches of the GF in the brain send out fine dendritic processes into several regions. These include the protocerebral lobe, the central commisure and the central body. Contacted fibres in these areas include interneurones from the eyes and antennae as

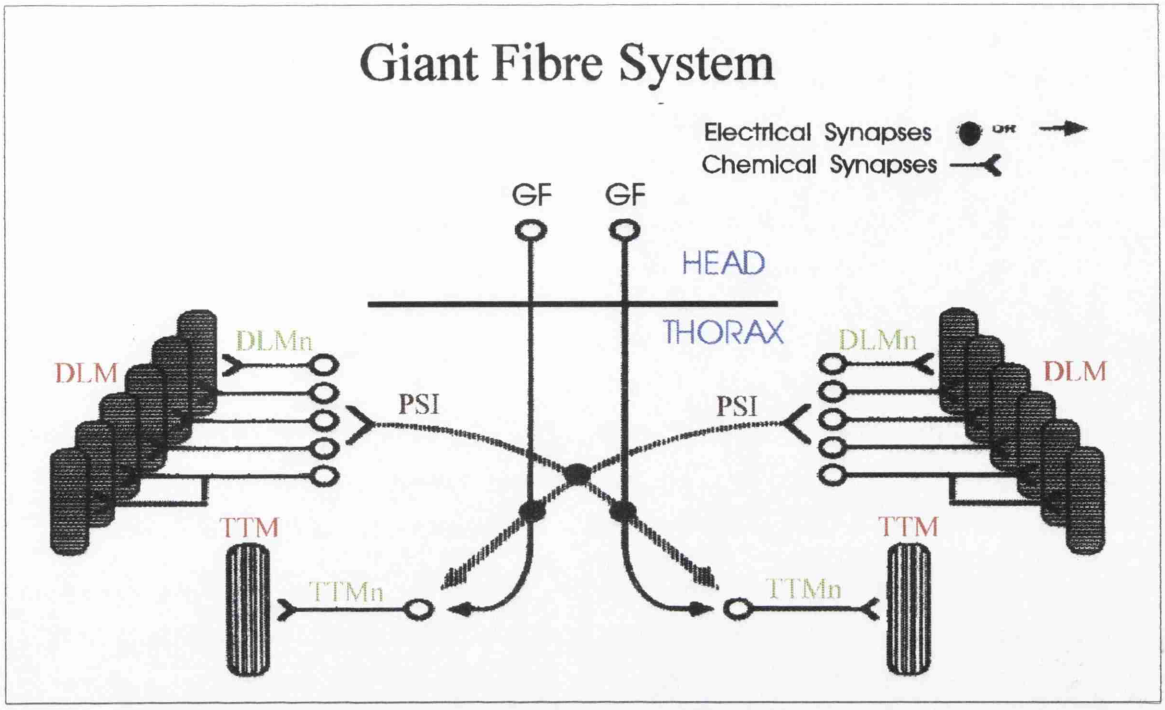


Figure 1.1 -Schematic representation of the Giant Fibre System. The eight neurones on each side (GF, TTMn, PSI, and five DLMs) are depicted with their synaptic interconnections. The GF cell bodies are in the brain; all other neurones are thoracic. GFs, giant fibres; DLM, dorsal longtitudinal muscle; DLMn, dorsal longtitudinal motorneurone; TTM, tergotrochanteral muscle; TTMn, tergotrochanteral motorneuron; PSI peripherally synapsing interneuron. From Baird *et al*., 1990.

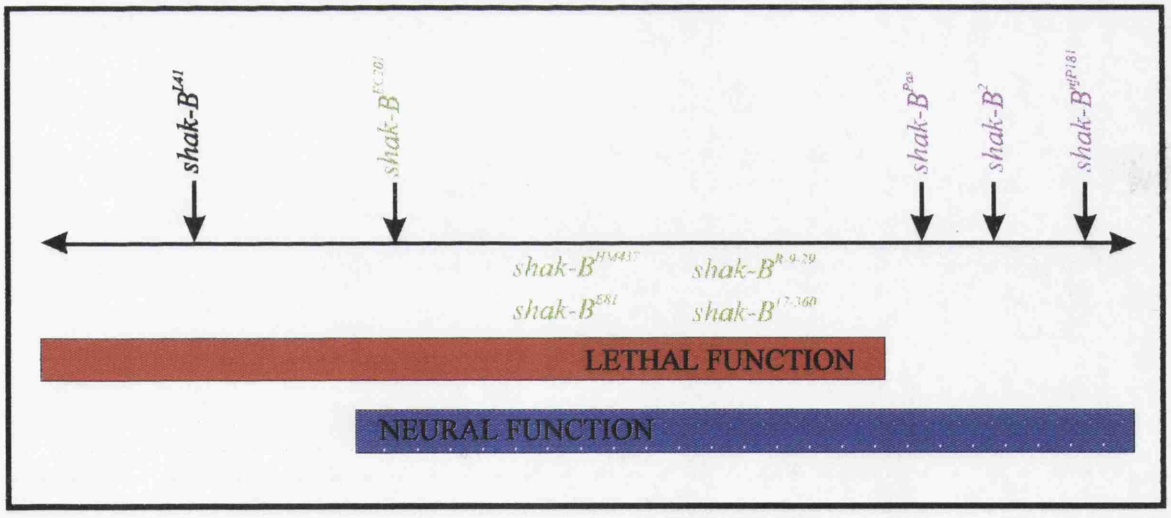


Figure 1.2 -Schematic diagram of the location of the *shak-B* alleles. The neural alleles are indicated purple, the lethal allele in black and those alleles that fail to complement the lethal or the neural functions are indicated green. Those above the line have been localised, the position of those beneath the line is still unknown. The diagram is not to scale as the neural alleles are located within 400bp of each other whereas the site of the lethal allele is some 10kb away from them. The alleles beneath the line have been positioned merely for convenience, they could be in some other region that has a joint role in the lethal and neural functions of the locus. See text (Section 1.27) for more details.

well as from association centres in the brain. (Other areas besides these are also contacted by the GF dendritic processes, Phelan *et al.*, 1996). This means that the GFS receives information from a variety of different stimuli, hence small field visual motion, puffs of air onto the head and body, displacement of the antennae, flickers, and wide field visual movement of gratings can all induce the escape response of the fly (Thomas and Wyman, 1984b).

Koto *et al.*, (1981) suggested that the GF is the command neurone for the escape response, thus implying that firing of the GF would automatically and exclusively lead to the response. For a variety of reasons however this appears not to be the case. The TTMn is not the only neurone that innervates the TTM. Two smaller motorneurones target the TTM (Ikeda *et al.*, 1980) and in the related fly species *Musca* and *Calliphora* homologous neurones to these have been shown to synapse onto the TTM. The TTMn also receives a variety of inputs (Bacon and Strausfeld, 1986).

Coggshall (1978) showed that in *Drosophila* the PSI receives other contacts than just from the GF, and speculated that these other neurones may regulate the PSIs responses. It appears that independent firing of the PSI can activate both the TTM and the DLM and hence instigate the escape response. King and Wyman (1980) suggest that "the mutual contacts in the thorax between PSI and TTMn could provide for bilateral activation of the muscles even if only a single PSI or GF is stimulated". So although the GFs may play an important role in the initiation of the escape response evidence suggests that it does not command the behaviour.

1.2.2 Electrophysiology of the Giant Fibre System.

Electrophysiological studies of wild type flies have shown that a light off stimulus or direct stimulation of the GF in the fly's brain causes a highly stereotyped temporal response (Tanouye and Wyman, 1980; Wyman and Thomas, 1983). When a stimulus is given to the GFS by a direct, electrical stimulation of the GFs in the brain of the fly, it takes 20ms for an action potential to be seen in the GF. 0.9ms later a single spike is elicited in the TTMn, causing the muscle to twitch and the fly to jump. 1.3ms after the GF spike the DLM start to contract (Tanouye and Wyman, 1980).

The latency from GF stimulation to the TTMn spike is 0.81ms whereas the latency from TTMn stimulation to the TTM spike is 0.66ms, thus only 0.15ms is allowed for the action potential to cross the synapse between the GF and the TTMn. This suggests that the synapse is electrical as this period of time is too short for chemical transmission. This is supported by refractory period and following frequency analyses which have shown that the pathway is extremely stable to high frequency stimulation. Also transynaptic cobalt migration has been shown to occur, which in other systems is known to be diagnostic for electrical synapses, as the dye can migrate across narrow electrical synapses but not across chemical synapses which are wider (Stewart, 1978; Koto *et al.,* 1981; Strausfeld and Bassemir, 1983).

The difference between the response of the DLM after direct stimulation of the DLMn in the thoracic ganglion, and the response of the DLM after GF stimulation is *ca.* 0.42ms. Time constraints fail to leave enough time for two chemical synapses, so it is concluded that at least one of the GF - PSI - DLMn synapses is electrical (Phelan *et al.,* 1996).

1.2.3 GFS Mutagenesis.

The GFS mutagenesis was performed by feeding ethyl-methanesulphonate (EMS) to male flies that had the wing disabling mutation *Curly* (Baird *et al.,* 1990). EMS causes base substitutions and small deletions in the DNA of developing spermatocytes. The progeny were then screened for non-escapees (*i.e.* non-jumping flies) by placing them on an inverted flask and then inducing the escape response by switching a light on and off. Those flies with mutations causing defects pre-synaptic to the GFS *e.g.* blind flies; those with GFS defects and those that had musculo-skeletal defects did not jump and so remained on the inverted flask. These were then set up as breeding lines and those that bred true for non-jumping were electrophysiologically tested to eliminate flies with defects outwith the GFS. Several mutations were produced that affected the escape response. Two that were isolated affect connectivity of the GFS. The *bendless* (*ben*) mutation affects the GF which, as the name implies, fails to bend in the thoracic ganglion at the appropriate point and so fails to contact the TTMn. The *ben* gene has been shown to encode a protein that is closely related to ubiquitin-conjugating enzymes (Muralidhar and Thomas, 1993; Oh *et al.,* 1994). *shak-B*^{Pas} (*Passover*) was so called because the TTMn was observed to 'pass over' the GF at the ganglion midline and not to synapse with it. This 'passing over' phenotype however, has since been

shown not to be caused by the *shak-B*^{Pas} background but is due to the presence of a particular deficiency chromosome, *Df(1)16-3-22* (Baird *et al.*, 1993). Another mutation, called *shak-B*^{2} was isolated in a different EMS mutagenesis screen that selected for mutants that amongst other phenotypes 'shook when etherised' (Homyk *et al.*, 1980).

1.2.4 Phenotypes of *shak-B* Mutants.

When compared to wild-type *Drosophila*, *shak-B*^{Pas} homozygous flies have abnormally long TTM response latencies of 1.5 +/- 0.8ms. No response can be evoked whatsoever from the DLM. The GF / TTM pathway is also defective in that normally the TTM can be driven 1:1 with the GF but in *shak-B*^{Pas} and *shak-B*^{2} homozygous flies this is only the case with frequencies less than 1Hz. It has also been shown that the defect occurs centrally and is not due to defective motor neurones or neuromuscular junctions as wild type values were obtained when the motor neurones were stimulated extracelluarly in the ganglia whilst recording from the muscles (Tanouye and Wyman, 1980; Baird *et al.*, 1990).

When the neurones of the GFS in *shak-B* mutants were stained with horse radish peroxidase or back filled with lucifer yellow, the GF was found to be normal. However, in *shak-B*^{Pas} flies the medial branch of the TTMn was shown to stop short of its normal site of contact with the GF and also to be reduced in diameter (King and Wyman, 1980; Wyman and Thomas, 1983; Baird *et al.*, 1990). It has since been shown that the anterior-posterior extent of the TTMn medial branch is reduced, which in turn reduces the amount of the membrane the TTMn has available for contact with the GF and the PSI (Baird *et al.*, 1993). The PSI is less amenable to study and has yet to be characterised to the same extent as the TTM so the defect has not been identified in this region Egger *et al.*, 1989; Swain *et al.*, 1990).

Besides the escape response phenotype, several other defects have been shown to occur in *shak-B* mutants. In sections of wild-type fly heads where the antennal nerve has been back filled with cobalt, it is possible to see three axon bundles which link the antennal glomerelli, (Stocker, 1981 . In *shak-B* mutants, one of these axon tracts is missing or much thinner than that observed in the wild-type (Aceves-Pina, personal communication). This has now been shown to be due to the loss of, (or defect in) the

gap junctions by which the molecule normally gains entry to the cells (Phelan *et al.*, 1996). *shak-B* mutants also have an abnormal electroretinogram (ERG). This recording is a measure of the change in electrical potential of the retina and lamina of the fly's eye. Alterations in it suggest that transduction is being affected in some way, (Coombe, 1986). *shak-B*^{Pas} flies also have an abnormal gustatory response. They exhibit increased thresholds of detection of sucrose and fructose and have lost completely their attraction response to 0.1M sodium chloride (Balakrishnan and Rodrigues, 1991). They also display an abnormal courtship behaviour (K O'Dell, personal communication). *shak-B*^{2} flies have leg tremors under ether anaesthesia. This shaking is weak and does not continue when the legs are severed from the body. This is different from the phenotype observed with the potassium channel mutant *Shaker*, which displays a generalised hyperexcitablity as when its legs are severed they continue to shake. This would seem to indicate that *shak-B*^{2} flies posses a CNS defect, possibly the disconnection of an inhibitory input which may be connected with lesions in the GFS. This idea of a CNS defect is supported by results which show flies with the *shak-B*^{Pas} mutation, have a reduced ability to groom themselves. A defective grooming response is thought to be correlated with defects in the CNS as grooming requires 'quite considerable sensory-motor co-ordination' (Phillis *et al.*, 1993).

It is known that olfaction is normal in *shak-B* flies (R. Stocker, personal communication), however most of the phenotypes identified to date have been quite subtle, *e.g.* the courtship behaviour, the gustatory response and the grooming behaviour. Therefore, without specific paradigms to score for them, these phenotypes would be undetectable in normal mutagenic screens. Thus other, as yet undetected phenotypes may be present in *shak-B* flies.

1.2.5 Genetics of *shak-B*.

shak-B^{Pas} was originally mapped to the base of the X chromosome to *Df(1)16-3-22* (Koto *et al.*, 1981). This deficiency removes approximately eighteen polytene bands extending from the *maroonlike* locus to the *extra organs* locus (Schalet and Lefevre, 1976). The *shak-B*^{Pas} mutation was localised to polytene band 19E3 using deletion mapping and complementation analysis (Baird *et al.*, 1990). When *shak-B*^{Pas} was made heterozygous with deficiencies that uncover the 19E3 region the heterozygote behaved like a *shak-B*^{Pas} homozygote. The *shak-B*^{Pas} mutation also failed to

complement some alleles of a previously isolated group of lethal and viable mutations, the *R-9-29* (now renamed *shak-B*) locus which lies in 19E3. The *shak-B* locus was initially shown to have nine alleles of which seven are lethal. The viable allele *shak-B*^{2} (Homyk *et al.,* 1980) was localised to the *shak-B* locus by Miklos *et al.,* (1987). Besides the uncoordinated leg movements under anaesthesia Baird *et al.,* (1990) found that *shak-B*^{2} flies have more or less the same electrophysiological phenotype as *shak-B*^{Pas} does, but that it does not act as a dominant mutation in the elimination of the jump response. Homozygous *shak-B*^{2} mutants also have the abnormal ERG phenotype. Flies homozygous for the lethal alleles, die after a prolonged first instar stage of two to three days (Baird *et al.,* 1990).

The *shak-B*^{Pas} allele dominantly disrupts the eye to GF pathway and so eliminates the escape response. This dominance however, is incomplete. In *shak-B*^{Pas} /+ heterozygotes the DLM following frequency can be as low as 10 - 25Hz as compared to wild-type values of 75Hz (Baird *et al.,* 1990). The TTM response in *shak-B*^{Pas} homozygotes is weak or more often absent, whilst in *shak-B*^{Pas} /+ heterozygotes the response has wild-type values. As *shak-B*^{Pas} homozygotes have a more severe effect than *shak-B*^{Pas} / deficiency heterozygotes the allele cannot be a null. Dosage compensation experiments have indicated that *shak-B*^{Pas} is neither a simple hyper- or hypomorph and is probably antimorphic, *i.e.* the mutant product is worse for the fly than if there was no product at all (D. Baird, 1988). *shak-B*^{2} / + flies yield electrophysiological values that are wild-type indicating *shak-B*^{2} is totally recessive.

The complementation pattern of alleles at the *shak-B* locus is complex. The existing alleles are, *shak-B*$^{R\text{-}9\text{-}29}$, *shak-B*^{HM437}, *shak-B*^{E81}, *shak-B*^{L41}, *shak-B*^{EC201}, *shak-B*$^{17\text{-}360}$, *shak-B*^{2} and *shak-B*^{Pas}. All except the *shak-B*^{2} and *shak-B*^{Pas} alleles are homozygous lethal and are also lethal when *trans* heterozygous with each other. The alleles may be grouped into three classes based upon their complementation pattern. The *shak-B*^{L41} allele fully complements the neural phenotype of *shak-B*^{2}and *shak-B*^{Pas} and yet fails to complement the lethality of the other alleles, thus *shak-B*^{L41} can be thought of as a purely lethal allele. A second lethal allele, *shak-B*^{EF535}, that also complements *shak-B*^{2}and *shak-B*^{Pas}, has been shown to posses the same molecular lesion as *shak-B*^{L41}, and so it is thought the *shak-B*^{EF535} allele is lost (Krishnan *et al.,* 1995; Crompton, *et*

al., 1995). The second group comprising *shak-B*$^{R\text{-}9\text{-}29}$, *shak-B*^{HM437}, *shak-B*^{E81}, *shak-B*^{EC201} and *shak-B*$^{17\text{-}360}$, fail to complement the *shak-B*^{L41} lethality but also fail to complement the *shak-B*^{2} and *shak-B*^{Pas} neural phenotypes, thus these are thought of as lethal and neural alleles. *shak-B*^{2}and *shak-B*^{Pas} comprise a third group of alleles with only a neural phenotype (Figure 1.2 for summary).

The complementation results suggest that two separate functions reside at the *shak-B* locus. One is an essential function and the other is required for neural function(s) (Baird *et al.,* 1990). These two functions must overlap as lethal / neural alleles exist which disrupt both functions. The simplest molecular explanation for this would involve differential splicing at the locus with at least one exon being common, however several other models are also plausible (see Section 1.2.6).

Df(1)A118 which impinges upon the *shak-B* locus proximally complements the lethality of the *shak-B* alleles but fails to complement the neural alleles. The *16-3-35* deficiency which impinges upon the locus distally, fails to complement either the lethal or neural phenotypes. As wild type recombinants can be obtained using these deficiencies (*i.e.* there is DNA between the proximal endpoint of *Df(1)16-3-35* and the distal endpoint of *Df(1)A118*), this suggests the region that is required for an essential function is more distal than the region required for the neural function(s).

1.2.6 *Trans* Effects of 19E5 - 6 Deficiencies.

The situation is further complicated by the interaction of the *shak-B* locus when combined in *trans* with proximal deficiencies in the 19E region. When in *cis i.e.* when both regions are normal on at least one chromosome, no effect on either the DLM or TTM latencies is observed. The deficiency *Df(1)16-3-35* (see Figure 1.3) or *shak-B*^{2} when heterozygous for the deficiencies *Df(1)T2-14A*, *Df(1)A53*, *Df(1)LB7* or *Df(1)17-489* (see Figure 1.3) have a mild effect on the TTM latency but no discernible effect on the DLM response. In contrast when *shak-B*^{Pas} is combined with the same deficiencies the TTM latency is increased and the DLM latency is also affected thus it appears *shak-B*^{2} and *16-3-35* have a less severe effect than *shak-B*^{Pas} when heterozygous for deficiencies deleting the 19E5-E6 region, thus illustrating the antimorphic nature of the *shak-B*^{Pas}.

The heterozygotes that produce the mild phenotype are deficient for two non-adjacent regions with at least one complementation group (*R-9-28*) intervening. All alleles of *R-9-28* fully complement the 19E3 mutants and the *16-3-35* deficiency. The proximal deficiency *Df(1)Q539* (see Figure 1.3) fully complements the neural phenotype and so marks the proximal limit of the interaction, however, as the *EC235* alleles fully complement the neural phenotype and no *little fly* alleles were available for testing, Baird *et al.,* (1990), note these two loci may also intervene. The gene may thus be split by an intervening gene (or genes) sitting in an intron (or introns) of *shak-B*. Several genes within genes have been identified in *Drosophila e.g.* the tRNA genes in the *decapentaplegic* gene complex (Gelbart *et al.,* 1985); a pupal cuticle protein in the *Gart* locus (Henikoff *et al.,* 1986); and at least seven genes are nested within two separate and large introns of the *dunce* gene, six of which have been characterised (*Sgs4*, *Pig-1*, and *nested genes 1, 2, 3* and *4*) and shown to have no relation to *dunce* function (Chen *et al.,* 1987; Furia *et al.,* 1991;1993). An alternative to the gene within a gene theory, is the possibility that the coding region could be at 19E3 with a long range interaction due to an enhancer or other control region in 19E5-6. Still other models are possible if the 19E5-6 deficiencies disrupt the 19E3 function due to their rearrangement of the chromosome, *i.e.* they may be juxtaposing novel DNA sequences close to the *shak-B* locus that affects the locus in some way. This is less likely than the other possibilities however, as the deficiencies have differing proximal endpoints, and thus juxtapose different sequences and yet have the same electrophysiological effect. If the reduction in expression was due to position effects, it would be expected that the *R-9-28* locus would also be affected as it is nearer to the 19E5-6 breakpoints than *shak-B*. When *R-9-28 / Df(1)T2-14A* heterozygotes were created however, no change in viability was detected (A Schalet, unpublished data, Baird *et al.,* 1990).

1.2.7 Interactions of *shak-B* with Surrounding Loci.

A strong hypomorphic allele of *runt*, *VE726*, appears to increase the latency of the TTM when heterozygous with the *shak-B*^{Pas} allele (Baird *et al.,* 1990). *VE726* is the only allele from the *runt* locus and *shak-B*^{Pas} is the only allele from the *shak-B* locus which yields this result. This may be a 'statistical blip' or it may indicate an actual interaction between these alleles as the *runt* gene is now known to encode a transcriptional regulatory protein that is extensively expressed in the developing CNS

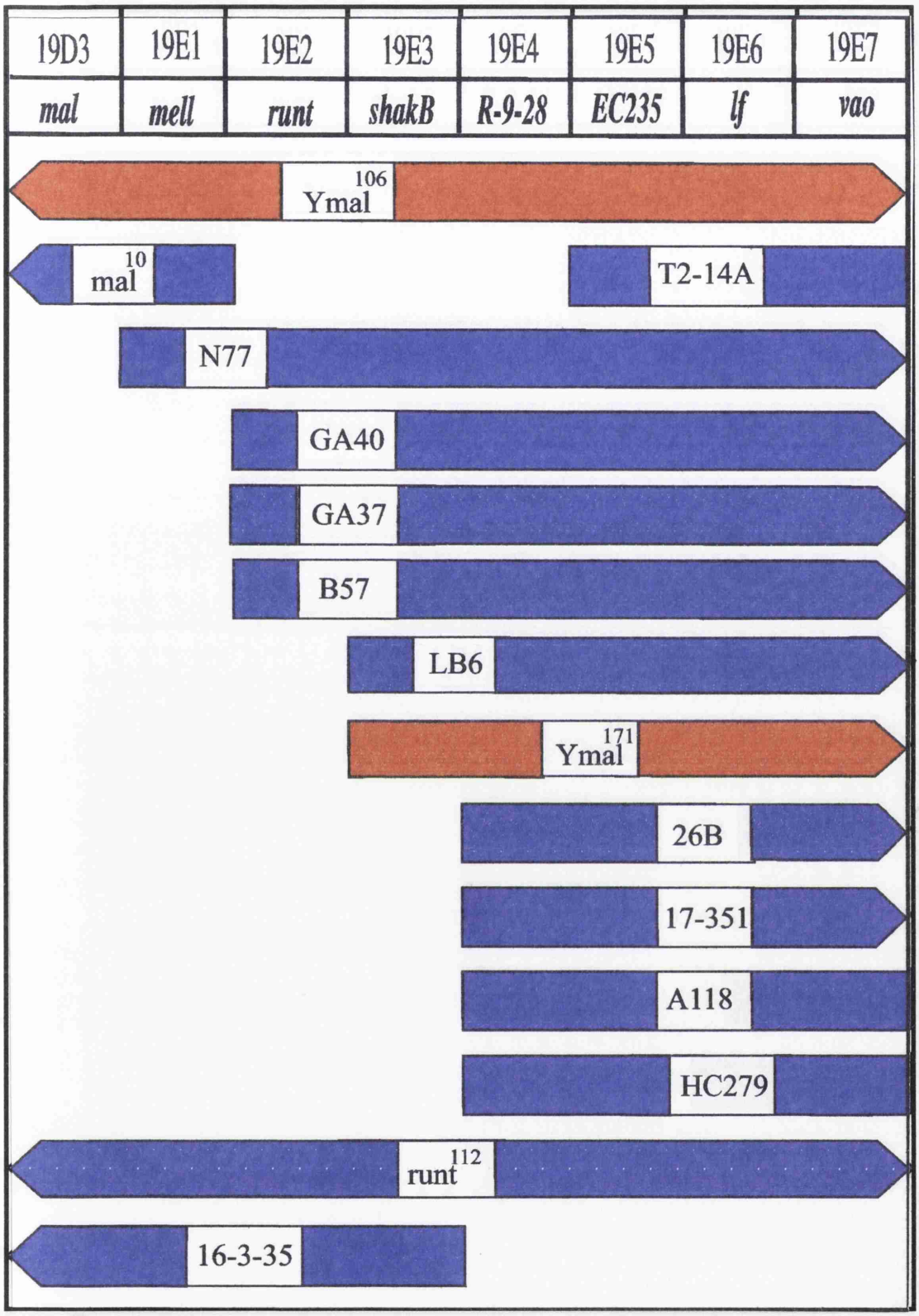


Figure 1.3. Schematic diagram of the genetic organisation of the deficiencies and duplications at the base of the X chromosome which were used and / or characterised in this study.

Duplication Deficiency

and PNS after its earlier function as one of the pair rule genes (Kania *et al.,* 1990). It has also been reported that some of the *R-9-28* alleles interact with *shak-B*^{Pas} with respect to the gustatory response (Balakrishnan and Rodrigues, 1991).

1.2.8 Molecular Analysis of the *shak-B* Locus.

1.2.8.1 cDNA Clones Isolated from the Locus and the Location and Effects of the Molecular Lesions Associated with the Mutant Alleles.

Several cDNA clones have been isolated from the region bounded by the proximal endpoint of the *16-3-35* deficiency and the distal endpoint of the *LB6* deficiency. As discussed above (Section 1.2.5) this region is known to be required for a vital function in the fly. The cDNA clones isolated represent alternatively spliced variants from the *shak-B* locus (Krishnan *et al.,* 1993; 1995; Crompton *et al.,* 1992; 1995). Three clones, B10 (now renamed S2.2, M. Wilkin, this laboratory; Crompton *et al.,* 1995), KE2 (Crompton *et al.,* 1992) and N52 (S. Ji, this laboratory), have been isolated from embryonic cDNA libraries using conventional and PCR based screening approaches. The cDNA isolated by Krishnan *et al.,* 1993 (called P2.4), was isolated from a pupal cDNA library, and is thought to be involved with the neural function. The cDNA isolated by them that is thought to be involved with the lethal function at the locus is similar too, but shorter than S2.2 (Figure 1.4 shows the location of exons from the transcripts isolated to date from the locus).

An analysis of the molecular lesions associated with the different alleles at the locus shows that the different transcripts mentioned above are responsible for the different functions of the gene (Krishnan *et al.,* 1995; Crompton *et al.,* 1995). The proteins from the locus are coded for by similar but distinct 5′ exons joined to a common set of 3′ exons. Neural only alleles (*shak-B*^{njp181}, *shak-B*^{2} and *shak-B*^{Pas}) and the lethal only allele (*shak-B*^{L41}) map to the 5′ exons that code for the neural and lethal proteins respectively. The molecular lesions associated with the lethal / neural alleles (*shak-B*$^{R\text{-}9\text{-}29}$, *shak-B*^{HM437}, *shak-B*^{E81} and *shak-B*$^{17\text{-}360}$) are found in the common 3′ exons. The *shak-B*^{E81} allele has a T>A transversion; *shak-B*$^{R\text{-}9\text{-}29}$ and *shak-B*^{EC201} both have G>A transitions whereas *shak-B*^{HM437} has a T>A transversion. All of these alterations produce stop codons. *shak-B*$^{17\text{-}360}$ is a large rearrangement affecting the common exons of the locus. The allele *shak-B*^{njP181}, (created by a P element insertion)

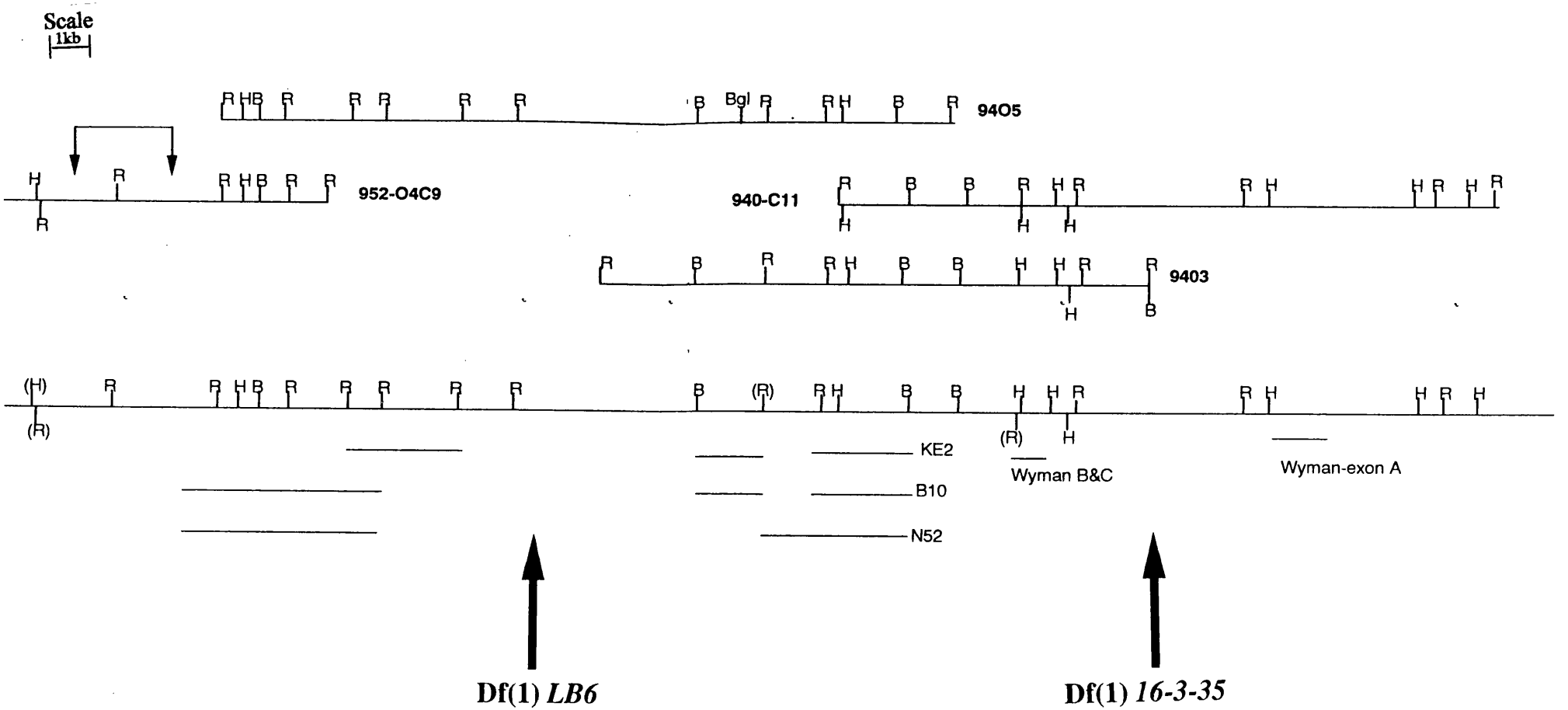


Figure 1.4 Schematic representation of the areas of the 952 chromosomal walk from which the *shaking-B* exons originate. The references for the cDNA exons shown are; Wyman (Krishnan *et al.*, 1993; 1995); B10 and KE2 (Crompton *et al.*, 1992; 1995); N52 (S. Ji, 1995). The deficiency breakpoints Df(1) *16-3-35* and Df(1) *LB6* are shown as the region between them is known genetically to contain sequences necessary for a vital function in the fly.

was shown to be located in the 5′ leader sequence 28bp from the start of the P2.4 cDNA. In *shak-B*^{2} alleles there is a T>A transition that results in a stop codon (TAA) replacing a leucine in the mature protein. The *shak-B*^{Pas} allele has been shown to be the caused by a C>T transition that results in a tryptophan replacing an arginine. Both of these alleles occur in the putative signal sequence of the P2.4 neural cDNA (see below; Krishnan *et al.*, 1993).

1.2.8.2 The Predicted Protein Products from the *shak-B* Locus.

The predicted protein from the P2.4 transcript is 361 amino acids in length with a putative extracellular domain of 256 amino acids followed by a transmembrane domain of 22 amino acids. Using the hydrophobic moment algorithm of Eisenberg *et al.*, (1984), Krishnan *et al.*, (1993), suggested that the predicted transmembrane domain is one that is usually found as a part of a multimeric complex. This is also supported by gene dosage experiments performed by Baird, (1988). Flies with a 1 *Shak-B*^{Pas} / 1 + ratio are mutant whilst those with a 1 / 2 ratio are phenotypically normal. This suggests some competion between the two alleles. One possibility (mentioned above) is that *Shak-B*^{Pas} competes with the one wild type product for binding of a ligand. However the change from 1/2 the binding sites being occupied to 2/3 of the sites being occupied (as would happen with a 1 / 1 to 1 / 2 dosage difference) might not be expected to be sufficient for a change in phenotype. These phenotypic differences observed with the different gene dosages can be explained quite easily if it is mutimeric *i.e.* the probability that a homomultimer will be composed solely of normal monomers depends upon gene dosage. For example the chance of forming a completely wild type hexamer (like a hemi-gap junction) increases six fold with the change in gene dosage described ($6^{1/2}$= only 1 chance in 64 of forming a wild type hexamer whereas $6^{2/3}$= a 1 in 11 chance of forming a wild type protein). Of course other explanations are also possible, for example the kinetics of binding could be altered so that the *shak-B*^{Pas} protein binds a ligand more tightly which therefore produces the phenotype.

The predicted protein from the KE2 transcript is 120 amino acids in length and the S2.2 transcript has a predicted protein length of 372 amino acids (Figure 1.4 shows the relationship between S2.2, p2.4, KE2 and N52. It is proposed that the S2.2 transcript has four transmembrane domains (Crompton *et al.*, 1995).

The predicted protein products of cDNAs isolated from the *shak-B* locus have structures similar to integral membrane proteins with at least one (Krishnan *et al.,* 1993), but probably four (Barnes, 1994; Crompton *et al.,* 1995), transmembrane domains. The structure of the proteins encoded by the *shak-B* transcripts (and the other proteins reported in the group) is of three major and two minor hydrophobic regions. Most of the residues that are conserved between the different members of the family are associated with the major hydrophobic domains whilst the loops between them vary in their length and composition (Barnes, 1994; Crompton *et al.,* 1995). Some controversy remains over the putative signal sequence. Krishnan *et al.,* (1993), suggest is a valid sequence because it bears all the signal sequence characteristics, *i.e.* it has a typical sequence with a hydrophobic core, and the predicted cleavage site (an arginine that is replaced by a tryptophan in the *shak-B*^{Pas} allele) is within the distance expected from the *N* terminus (von-Henje, 1986). However, the sequence of the *unc7* gene has an additional 120 residues at the *N* terminus and has no signal sequence. Similarly the l(1)Ogre and S2.2 proteins do not appear to have a signal sequence (Crompton *et al.,* 1995).

The finding that the *shak-B*^{Pas} allele substitutes a neutral tryptophan for an arganine in the proposed cleavage site of the signal peptide may explain its partial dominance and its antimorphic nature (Baird, 1988). If the transmembrane portion of the neural (P2.4) transcript binds into a multimeric complex (as is suggested by the transmembrane sequence), and if the extracellular domain does mediate a homo- or heterophillic interaction, then the antimorphic nature of *shak-B*^{Pas} could be explained by the mutant *shak-B*^{Pas} product competing for binding to (one or more) of its normal ligands, but the resulting complex being functionally defective (Krishnan *et al.,* 1993). Under an alternative model proposed by Crompton *et al.,* (1995) however, the substitution of tryptophan for arganine is proposed to exert its effects by altering the properties of the transmembrane domain. Under this model the alteration would lie at the extracystolic end of the first transmembrane helix of the protein. Distinguishing between the signal peptide and signal anchor models for the neural protein is not yet possible with the available data, and so must await further study.

1.2.8.3 Homologies of the Transcripts and the Proteins.

The cDNAs from the *shak-B* locus (Krishnan *et al.,* 1993; 1995; Crompton *et al.,* 1992; 1995) and the predicted proteins they yield, have homology to several other genes. One is the *Drosophila* locus, *lethal (1) optic ganglion reduced* (*l(1)ogre*, Watanabe and Kankel, 1990); and two others, the *unc7* gene (Starich *et al.,* 1993) and the *eat 5* gene (Avery, 1993; Starich *et al.,* 1995) both come from the nematode worm *C. elegans* Other sequence homologies were identified in *C. elegans* (ten to date) but some of these are to uncharacterised cDNA sequences (Barnes, 1994; Starich *et al.,* 1995). It is thought likely that the *shak-B* transcripts, together with these others, comprise a novel gene family which it has been proposed be called OPUS (Barnes, 1994).

It has been suggested that the OPUS family of proteins may be the invertebrate equivalents of the vertebrate gap junction proteins, the connexins (Barnes, 1994). Connexins have long been implicated in developmental mechanisms in vertebrates (*e.g.* Dealy *et al.,* 1994; Goodenough and Musil, 1993; Kanter *et al.,* 1994 for references). The evidence however, for the relationship between the OPUS and Connexin family of proteins is based solely on topological data implied from the presumed protein products of the transcripts. The fact that no significant extended similarity between the vertebrate connexins and the OPUS protein sequences has been identified (Barnes, 1994) makes the relationship very tentative at best. However, the structural similarities between the OPUS and the vertebrate connexins would appear to be quite compelling. The similarities include four transmembrane regions with the C terminal and N terminal tails intracellular; The connexins are conserved throughout their sequence except for the cytoplasmic tails which are variable in both sequence and size. This is also true for the OPUS family; There are six completely conserved cysteines on the extracellular loops (these cysteines in connexins have been shown to be crucial for junctions to form, changing any of them to serine results in an absolute loss of function); There is also a positively charged cytoplasmic loop between the second and third transmembrane domains whilst there is a proline residue in the middle of the second transmembrane domain which is known to be crucial in voltage gating in the connexins (Suchnya *et al.,* 1993). This proline residue is completely conserved in both the connexin and OPUS family; Members of the OPUS family, like the connexin family, do not have a good consensus signal sequence although there is a

clear transmembrane domain; Finally, gap junction are mutimeric structures (hexamers) and *shak-B*^{Pas} has a transmembrane sequence typical of mutimeric proteins (Krishnan *et al.,* 1993).

There are several things make this theory less plausible however. Firstly essential *shak-B* transcripts appear in the embryo after most of the cells become uncoupled. They also have been shown to appear in cardioblasts which are known not to form many gap junctions (Tepass and Hartenstein, 1994). Secondly the sizes of the *Drosophila melanogaster* gap junction proteins have been analysed biochemically and their sizes do not agree with the predicted sizes of any of the *shak-B* transcripts (Ryerse, 1989a). Finally considering gap junctions occur at the cell surface, the intracellular location of one member of the OPUS family (*l(1)ogre*) together with the formation of ectopic gap junctions in the absence of Unc-7 protein (see below) again suggests that the OPUS family may not be the invertebrate equivalent of the connexins. There are other (vertebrate) protein families that the OPUS family have similar topologies to *e.g.* the transmembrane 4 (TM4) family (*e.g.* Horejsi and Vlcek, 1991; Dong *et al.,* 1994).

The *l(1)ogre* gene and the *shak-B* P2.4 transcript, both affect postembryonic neurones (Lipshitz and Kankel, 1985; Krishnan *et al.,* 1993). Both genes cause neuronal defects but these affect connectivity rather than neurogenesis (Thomas and Wyman, 1984; Baird *et al.,* 1990; Baird *et al.,* 1993; Starich *et al.,* 1993). In *unc7* mutants, the worm is defective in forward locomotion. When it tries to move forward the body forms into irregular kinks. It has been demonstrated that this is due to miswiring of neurones in the nematode (Starich *et al.,* 1993). Electron micrographs indicate that AVA interneurones which will normally form synapses only with motorneurones for backward locomotion, now form gap junctions with motorneurones for forward motion. This miswiring has been put forward as the explanation for the behavioural phenotype. The Unc7 protein is also thought to be involved in the nematode's response to certain anaesthetics as do the *unc79* and *unc80* gene products which cause hypersensitivity to halothane and chloroform when mutant. When double mutants are constructed with *unc7* these mutants lose their hypersensitivity to these anaesthetics. It has also been shown that the *unc7* mutation gives increased resistance to the antihelmintic, ivermectin (Morgan *et al.,* 1990). Anaesthetics are thought to have their

effect by acting on ion channels or ion channel associated proteins (Franks and Lieb, 1991; Matthews, 1992). Although *unc7* is not the ivermectin sensitive chloride channel in nematodes, it is believed that invermectin has its effects by acting on chloride channels. It has thus been suggested that the *unc7* product may be an ion channel associated protein (Starich *et al.,* 1993). *eat-5* mutations disrupt pharyngeal pumping (Avery and Horvitz, 1989). In normal worms the corpus and the terminal bulb of the pharynx contract nearly simultaneously. In *C. elegans* laser ablation of the pharyngeal nervous system has shown that contraction synchrony of pharyngeal muscle cells can be attributed to gap junctions. In *eat-5* mutants parts of the pharynx, the corpus and terminal bulb, contract approximately normally, but they are no longer synchronised (Avery, 1993). Connections are normal within the corpus and within the bulb, but abnormal or missing between the two. It is therefore possible that *eat-5* mutations also affect a particular gap junctional connection whilst apparently leaving others unaffected (Starich *et al.*, 1995).

1.2.8.4 Expression Patterns of the Various cDNAs.

It has been demonstrated that the P2.4 cDNA (or a closely related transcript that cross hybridises with it) is expressed in the brain of the newly emerged adults in two cells that are almost certainly the GF cell bodies (Krishnan *et al.,* 1993; Crompton *et al.,* 1995). In the thorax expression is also detected at the posterior border of the wing neuromere, which is the location at which the cell bodies of both the TTMn and the PSI are located. It was not reported which part of P2.4 was used as probe(s) for the *in situ* hybridisation; it is not known whether the probe used would cross hybridise to the rest of the transcripts discovered to date. It was not possible to detect which of these was the P2.4 expressing cell (Krishnan *et al.,* 1993). Other *in situs* suggest that S2.2 is present in the thorax at much higher levels than P2.4 (Jane Davies, personal communication; Crompton *et al.,* 1995).

Crompton *et al.,* (1995) found that the S2.2 and KE2 are expressed in embryos, initially in precursors of a subset of the larval musculature and in the midgut visceral mesoderm. Later expression is detected in the foregut visceral musculature, in the dorsal pharageal musculature as well as in cardioblasts. In pupation a dynamic expression of *shak-B* transcripts was described. At the end of the third larval instar faint hybridisation was seen to cells which will go onto form part of the

suboesophogael ganglia in the adult and to a pair of cells in the central brain. Much stronger expression is seen in the same pair of cells 12 hours after puparium formation (APF) which are probably the GFs. In the thoracic neuromeres variable numbers of ventrally located cells express *shak-B* whilst more dorsally a group of 3-4 cells (which probably include the PSI and TTMn cell bodies) displayed strong hybridisation. Faint expression was also seen in most cells of the central brain and thoracic neuromeres of the ventral ganglion at this stage which then intensified over the next 15 hours. At 25-30 hours APF they show that the developing optic lobes start to express *shak-B* and by 48 hours APF the majority of cells in the optic ganglia, brain, and thoracic neuromeres are expressing *shak-B* at high levels. This expression then remains high for several hours. At 75 hours APF extensive expression of *shak-B* was seen in the medulla, in some cells of the lobular complex, in the thoracic ganglion, and in the central brain, including several cells which are again probably the GF cell bodies. Two days after eclosion the only *shak-B* expressing cells described were those cells that are presumed to be the GF cell bodies. Hybridisation with probes specific for the S2.2 transcripts failed to detect expression in the GFs, the lamina and the medullary cortex of the optic lobes. The signal seen in these structures must therefore be due to hybridisation to the neural transcript (P2.4, [or other members of the OPUS family to which it cross hybridises]) and not the vital transcript.

1.2.8.5 Possible Mechanisms of Action of the *shak-B* Products.

1.2.8.5.1 The Neural Function.

Krishnan *et al.,* (1993) proposed that, as the expression of P2.4 occurs in the GFs and in a cell that is probably its postsynaptic target (either the TTMn or the PSI), then the simplest explanation for the neural function of *shak-B* is that it acts as a homophilic cell adhesion molecule. The synapse between the GF and the TTMn is known to be electrical and so some support for an adhesive role of Shak-B proteins comes from observations that a primary cell adhesion event may be needed prior to the formation of gap junctions (*e.g.* Jongen *et al.,* 1991). Several points argue against this however. Crompton *et al.,* (1995) showed that the expression of P2.4 is low in the TTMn cell body whilst S2.2 is expressed at high levels. It is also suggested that as it is likely "that neural and essential protein species play similar roles, the lack of an obvious morphological phenotype in flies carrying *shak-B* alleles" *i.e.* similar to the TTMn defect described earlier (Section 1.2.4) "is difficult to equate with a purely adhesive

function". Finally as Shak-B and Ogre proteins are so related it would be difficult to explain how Ogre, with its intracellular location, can have a role in cell - adhesion.

It seems likely that Shak-B proteins, although probably not the structural gap junction proteins themselves, must play a role either in gap junction formation or act to stabilise GF junctions once they are formed. This is because lucifer yellow filling of the GFs have shown that the GF and TTMn are uncoupled in *shak-B*^{2} flies *i.e.* the dye does not pass from the GF into the TTMn in these flies whereas it does in those with wild-type genotypes (Crompton *et al.,* 1995). High expression of *shak-B* in cells that do not normally have very many gap junctions (*e.g.* cardioblasts) would imply a similar role in the formation or stabilisation of other specialised membrane structures. That is, it is possible that *shak-B* has its effects by altering the membranes of affected cells. This is a role that has been proposed to explain the effect *unc7* mutants have on neuronal physiology (Starich *et al.,* 1993).

1.2.8.5.2 The Vital Function.

The mechanism behind the lethality that resides at the *shak-B* locus is still unknown. It is proposed (Crompton *et al.,* 1995) that essential Shak-B proteins are required for the function of some embryonic mesodermal derivatives and that the higher levels of hybridisation (described above) to the brain and ventral ganglion suggests they may also play a general role in CNS development, while the neural proteins endow a subset of neurones with more specialised properties. It may be significant that both the essential *shak-B* and *unc7* transcripts are expressed when postembryonically derived neurones are generated and extending processes (Starich *et al.,* 1993; Crompton *et al.,* 1995). It is also known that *l(1)ogre* has a role in the development and maintenance of postembryonic neuroblasts (Lipshitz and Kankel, 1985; Singh *et al.,* 1989).

Although many suggestions have been made the nature of the actual lethal defect remains elusive. The theory that some of the Shak-B proteins whilst not actually being the gap junctions may be associated with them, or have a role in membrane stabilisation is suggestive. It requires more members of the OPUS family to be identified in *Drosophila* (and in other invertebrates), and a cellular location determined for the *shak-B* proteins to prove it. Only a few gap junction associated proteins have been identified in *Drosophila*. For example Finbow *et al.,* (1994) have

identified the ductin component of gap junctions based upon its homology to sequences from the *Manduca sexta.*

All three clones isolated from this laboratory are thought to be involved with the lethal function that resides at the *shak-B* locus. An analysis of the alleles at this location (Krishnan *et al.,* 1995; Crompton *et al.,* 1995) supports this. *shak-B*^{L41}, a lethal allele, has a 17bp deletion that removes the translation start site from the open reading frame of both S2.2 and KE2. No alternative start sites are available for use by the mutant protein. This implies that at least some of the proteins initiated from the vital transcript isolated by Krishnan *et al.,* 1995 (pas^{vital}); by S2.2 and by the KE2 start sites are needed for a vital function in the fly.

1.3 19E and Other *Drosophila* Genomic Regions.

The *shaking-B* locus lies upon the border of a fascinating area in which the DNA changes from the conventional X chromosome euchromatin to the β-heterochromatin at the proximal X. Many studies have been carried out in this region not only to study the many important loci situated here but also to determine the way that DNA changes as it approaches the highly repetitive α-heterochromatin at the centromere (*e.g.* Miklos *et al.,* 1984; 1988). From the viewpoint of studying the fly genome as a whole *e.g.* studying the incidence of vital and non-vital genes; the occurrence of repetitive DNAs; the study of this region is extremely important.

1.3.1 The Genomic Region Containing the *shak-B* Locus.

An estimated one and a half megabases makes up the 34 band interval in which the *shak-B* locus lies, from the *maroon-like* locus at 19D3, to the *suppresser of forked* locus at 20F2 (John and Miklos, 1988). This region is recognised as a transition zone from the conventional euchromatin within subdivision 19 to the β-heterochromatic sequences located in division 20. It is amenable to study and it is probably one of the best characterised regions in the *Drosophila* genome as it has been the subject of intensive genetic, developmental and cytogenetic analysis (Lifschytz and Falk, 1968;1969; Schalet and Lefevre, 1973; 1976; Lifschytz and Yakobovitz, 1978; Kramers *et al.,* 1983; Eeken *et al.,* 1985; Zusman and Wiechaus, 1985; Lefevre and Watkins, 1986; Miklos *et al.,* 1987;1988; Perrimon *et al.,* 1989a:b; Baird *et al.,* 1990; Perrimon *et al.,*(1989b) for review). This area contains several loci with proven and/or

putative neurological phenotypes such as *runt* (Kania *et al.*, 1990; Duffy and Gergen, 1991), *uncoordinated*, *uncoordinated-like*, *shaking-B / Passover* (Baird *et al.*, 1990; Krishnan *et al.*, 1993;1995; Crompton *et al.*, 1992; 1995), *sluggish-A* (Hayward *et al.*, 1993), *stoned* (Petrovich *et al.*, 1993) and *stress sensitive* (Perrimon *et al.*, 1989b).

The α–heterochromatin which is located proximal to the *suppresser of forked* locus at 20F2, is not easily visible in polytene chromosomes at the light microscope level. Electron micrograph studies however, have shown the central region that contains α-heterochromatin is a compact block which is almost completely devoid of activity as regards RNA synthesis (Lakhotia and Jacob, 1974). This central region is surrounded by a diffuse granular area (the β-heterochromatin) that appears to be as active in RNA synthesis as conventional euchromatin. Thus it seems that the cytological characteristics of the α– and β-heterochromatin may reflect their transcriptional properties (Miklos and Cotsell, 1990; Yammamoto *et al.*, 1990). The α-heterochromatin has been shown to be principally composed of highly repetitive simple DNA sequences such as satellite sequences (see below Section 1.4.1), interspersed with moderately repetitive elements. Some of these sequences are specific to single chromosomes whilst others are dispersed amongst all the chromosomes (*e.g.* Miklos *et al.*, 1984; 1988). The β-heterochromatin also seems to be rich in repetitive sequences but it is not composed of significant amounts of the major satellite DNAs (Miklos *et al.*, 1988). A large proportion of these repetitive sequences are middle repetitive and have been shown to cross hybridise with transposable elements (Yammamoto *et al.*, 1990; see Section 1.4.2.1) whilst a significant proportion have been shown to be other middle repetitive non-satellite sequences (Young *et al.*, 1983; Donnelly and Keifer 1986, 1987; Miklos *et al.*, 1988), some of which are over represented relative to subdivisions 19E and F (Yammamoto *et al.*, 1990). A description of the major repetitive sequences described in *Drosophila* and from this region is given below and in Chapter 4.

Euchromatin makes up only an estimated 50% of the mitotic length of the chromosome and yet it contains 99% of the X chromosome loci (Hilliker *et al.*, 1980a) Many studies have looked at the gene frequency at the base of the X chromosome and tried to relate this to the chromatin state. It appears however that the gene frequency of β-heterochromatin is little different to normal euchromatin (Schalet and Lefevre,

1973, 1976; Lefervre, 1981), whilst the α-heterochromatin has a reduced gene frequency, this being approximately one gene every megabase of DNA (Hilliker *et al.*, 1980a)

A large number of the genes in the 19D - 20F2 region have been cloned or are in the process of being so, *e.g. unc*, (Miklos *et al.*, 1984); *su(f)*, (Mitchelson and O'Hare, 1991); *l(1)B214*, (Russell *et al.*, 1992); *sluggish-A*, (Hayward *et al.*, 1993); *flightless-1*, (Campbell *et al.*, 1993); *stoned*, (Petrovich *et al.*, 1993). Several of these genes are of interest from both a phenotypic and also a cytological viewpoint. The nearest distal neighbour to *shak-B* is the *runt* locus, which is known to encode a novel DNA binding protein that has been shown to have roles in a number of important events in the *Drosophila* life cycle (*e.g.* Gergen and Butler, 1988; Kania *et al.*, 1990; Duffy and Gergen, 1991). Other loci *e.g. extra-organs* at 20A, are of interest from a cytogenetic viewpoint as they contain a very large amount of non-repetitive single copy DNA relative to some of the other single lethal complementation groups in this region (Yammamoto *et al.*, 1990).

1.3.2 The Numbers of Essential and Non-Essential Genes in *Drosophila*.

The genome of *Drosophila melanogaster* has been estimated to contain *ca.* 1.65×10^6bp of DNA, of which 40 megabases makes up the X chromosome (Yammamoto *et al.*, 1990). Estimates for the total number of genes in the 5059 bands (Bridges, 1938) of the polytene chromosomes of the *Drosophila* genome have ranged from 2500 to 6400 (see Lefevre and Watkins, 1986). Recent estimates making use of an up to date compilation of *Drosophila* mutants (Lindsley and Zimm, 1992) and of studies that have systematically saturated specific chromosomal regions for mutations, suggests that there are in fact some 3600 genes that can be mutated to recessive zygotic lethality, with up to a further 900 genes required for an essential function during oogenesis or spermatogenesis. This latter figure may be an overestimate as Perrimon and Mahowald (1986) estimated the number of essential loci needed for oogenesis at about 50. The difference between the two figures is due to the assumption that sterility mutants are only required for gameteogenesis. In fact it is known that a large percentage of them are alleles of lethally mutable loci. Thus this figure of 4500 essential genes in *Drosophila* is probably a maximum (Brizuela *et al.*, 1994).

1.3.3 Number of Transcripts in Other Regions of the Genome and at 19E3.

It has been estimated that there are between 10 000 and 14 000 different RNA transcripts encoded by the *Drosophila* genome (Turner and Laird, 1973; Levy and McCarthy, 1975; Levy *et al.*, 1975; Levy and Manning, 1981) thus suggesting that more than half of the genes in *Drosophila* may be non-essential. The differences in the numbers of essential genes and the number of different RNA transcripts has also been noted in studies that have characterised chromosomal regions at the molecular level. For example the *ace - rosy* walk at subdivision 87DE contains 11 essential genes (plus the non-essential *rosy* gene) and encodes 43 transcripts in a region of 315kb (Hilliker *et al.*, 1980; Spierer *et al.*, 1983; Hall *et al.*, 1983; Bossy *et al.*, 1984). The *el-noc* complex walk contains only one essential gene and four other non-essential ones in 250kb (Davis *et al.*, 1990). The *bramha* walk at subdivision 72AB contains 5 essential genes and 14 transcripts in a region of 50kb (Brizuela *et al.*, 1994). The *dunce* walk at subdivision 3C11-12 contains 5 different transcribed regions belonging to the *dunce* locus, which encode a minimum of eight transcripts. There are also at least seven other transcription units that exist within introns of the gene (Qiu and Davis, 1993; Furia *et al.*, 1993). Over half of the genes located in a 170kb walk at the 26A region, have been shown to be non-essential insofar as they were not identified in saturation screens for recessive lethals using standard genetic methods (Knipple *et al.*, 1991). A locus near to *shak-B*, *l(1)B214* at 19F1-2 has also been characterised at the molecular level and has been show to contain one lethal gene and at least eighteen different polyA$^+$ RNA species in 50kb (Russell *et al.*, 1992). This compares with one essential gene at 19E3 which encodes at least five transcripts (Crompton *et al.*, 1992;1995; Krishnan *et al.*, 1993;1995) with a further eleven transcribed regions surrounding it, six of which are known to encode at least eight transcripts (this study, Table 5.1).

1.4 Repetitive DNA Characterised in Chromosomal Walks in *Drosophila*.

A number of chromosomal walks have been carried out in *Drosophila* from which repetitive DNA cloned in the walks has been characterised to at least some degree. The walks that have cloned *uncoordinated* (*unc*) at 19E8 (Healy *et al.*, 1988,1990); *suppresser of forked* (*su{f}*) at 20E / F (Mitchelson *et al.*, 1993; Langley *et al.*, 1993) and *l(1)B214* at 19F1-F2 (Russell *et al.*, 1992) on the X chromosome, and, for example the *light* gene at 40C-40F on chromosome 2L (Devlin *et al.*, 1990a, b); have

all identified and, to at least some extent, partially characterised repetitive sequences within their respective walks. About 72kb, of the 80kb of sequence encompassing the *lt* gene, consists of middle repetitive sequences (Devlin *et al.,* 1990a). The remaining 8kb of single copy sequence correlates well with exonic regions. A similar situation exists in the *su(f)* region where every recombinant phage clone isolated contained repetitive sequences and where the single copy sequences correlated well with the *su(f)* exons (Mitchelson *et al.,* 1993; Langley *et al.,* 1993). The region containing the *unc* gene does not contain as many repetitive sequences (Healy *et al.,* 1988) and although the relationship between the repetitive sequences and the *unc* transcripts has yet to be reported, it is known that both unique and tandemly repeated sequences exist, separated by a member of a highly repeated sequence family part of which is homologous to the ribosomal type I insertion sequence (Healy *et al.,* 1988).

Other walks, which have cloned large areas of euchromatin, are either free of repetitive DNA or contain fewer repetitive regions than those identified in walks near to centromeres. For example the 315kb of DNA that was isolated in 23 overlapping phage clones in the *ace-rosy* walk (Bender *et al.,* 1983a) was shown to contain two mobile elements and only a few moderately repetitive elements, between which there was considerable sequence divergence (Hall *et al.,* 1983). Similar situations were shown to exist in the *bithorax*-complex walk (Bender *et al.,* 1983b; Hall *et al.,* 1983); in the *achaete-scute* complex (Campuzano *et al.,* 1985), and in the *el noc* complex (Davis *et al.,* (1990) and T. Davis personal communication). Only a small number of genomic phage clones in these walks contained repetitive DNA. This paucity of repetitive sequences within euchromatic walks does not hold true for every region. A 200kb walk carried out at 3B2-3C2, to clone the *white* locus (Pirrotta *et al.,* 1983), contained a large number of repetitive elements (16% of the microclones obtained from a microdissection of the region were repetitive as judged by reverse Southern analysis). A large number of these sequences were homologous to transposable elements *e.g. B104*, *copia*, and *412*, (Pirotta *et al.,* 1983). However, this region, may be exceptional as the proximal end of the walk flanks a region with the characteristics of what was previously termed 'intercalary heterochromatin'. Intercalary heterochromatin is a highly repetitive area that is under-replicated in polytene chromosomes and that can undergo ectopic pairing). Another feature of intercalary heterochromatin (increased susceptibility to X ray induced chromosomal breakage and

reunion) is probably not only associated with repetitive DNA but may also occur within long stretches of single copy DNA. Other walks also located close to regions that used to be classed as 'intercalary heterochromatin' are also more repetitive than ones situated in euchromatin *e.g. Kruppel* (Preiss *et al.,* 1985) and *lethal (2) giant larvae* (Mechler *et al.,* 1985). Due to their association with 'intercalary heterochromatin' these walks are probably not typical of euchromatin.

1.4.1 Repetitive DNA at the Base of the X Chromosome in *Drosophila melanogaster.*

The repetitive DNA found at the base of the X chromosome in *Drosophila melanogaster,* can be divided into three main classes. These are highly repetitive sequences *e.g.* satellite DNAs; moderately repetitive sequences *e.g.* transposable sequences; and short repetitive sequences such as the *opa* repeat. Within the first two groups there are four subgroups which are summarised below.

1.4.2 Highly Repetitive Simple Sequence (Satellite) DNA.

Large amounts of repetitive DNA occur at the centromeres and telomeres of *Drosophila* chromosomes. Hundreds to many thousands of copies of a repeating unit may be found within one chromosomal region. These long tandem arrays of repeated sequences frequently form separate 'satellite' bands when total DNA is centrifuged to equilibrium on density gradients (Peacock *et al.,* 1973 and Endow *et al.,* 1975). In *Drosophila melanogaster* these satellite bands fall into four distinct buoyant density classes 1.672, 1.686, 1.688 and 1.705gml^{-1} in CsCl. They have now also been distinguished by the length and the nucleotide sequence of the repeat unit (Peacock *et al.,* 1977; Endow *et al.,* 1975; Sederoff *et al.,* 1975 and Carlson and Brutlag 1977). Satellites are amongst the last DNA segments to be replicated in S phase and are often under represented in polytene chromosomes when compared to other DNA sequences.

Each *Drosophila* species has a distinctive set of centromeric satellites (Endow *et al.,* 1975). *Drosophila melanogaster* possesses four different major satellite families with at least fourteen other minor families of highly repeated satellite DNA (Lohe and Brutlag, 1986). Satellite I contains two simple and closely related domains, with one a pentameric and one a heptameric repeat unit, (5'-$[AATAT]_n$-3' and 5'-$[AATATAT]_n$-3'). About 80% of Satellite II is composed of a tandem repeats of a

decamer (5′-[AATAACATAG]$_n$-3′). Satellite IV, like satellite I, includes both a pentameric and a heptameric domain which posses an obvious relationship to satellite I, (5′-[AAGAG]$_n$-3′ and 5′-[AAGAGAG]$_n$-3′) whereas Satellite III has a unique structure. It has a repeat length of 359 base pairs (bp) with there being an estimated 16,000 copies of the repeat of the Satellite III monomer (Hsieh and Brutlag, 1979). The chromosomal locations of these satellites differ from one another as is shown when satellites were hybridised *in situ* to polytene chromosomes (Steffenson *et al.*, 1981). Satellite I and IV sequences can be found in all centromeres though specific locations differ between them as does the number of repeats. Satellite III hybridises to only the X and Y chromosome centromeres.

1.4.2.1 Functions of Heterochromatin.

For many years the heterochromatin was thought of as merely junk DNA with few genes and no detectable function. The primary reasons for this were; firstly there was no phenotypic effect (and thus no ascribable function) which could be observed following the deletion of satellite DNA: *Drosophila* mutants that lack all or most of their centromeric heterochromatin survive and seem to function well (Hilliker, 1982): Secondly, there is substantial freedom for satellite DNA to adopt quite different primary sequences even in closely related species *e.g. Drosophila melanogaster* and *Drosophila virilis* (Endow *et al.*, 1975). This suggests their is little reason to preserve any particular sequence. Thirdly, the amount and sequence organisation of a centromeric satellite can vary extensively even amongst individuals of the same species, for example the α-satellite of humans (Waye and Willard, 1985). It is only comparatively recently, that work has suggested that there may actually be important functions for the tandem repeats in heterochromatin of *Drosophila melanogaster.* Pairing of chromosomes at meiosis has been shown to require homology between the 240bp spacer which is present in several tandemly repeated copies in the intergenic spacer. This is a region of DNA that separates each of the rDNA repeats (McKee and Karpen, 1990; Mckee *et al.*, 1992). One copy of the 240bp spacer can significantly rescue X-Y pairing, however proper recognition and segregation improves with increased numbers of it. Thus it has been suggested that the ability of heterochromatin to maintain a high copy number of dispersed repetitive elements maybe an evolutionary response to such additive features. Meiotic segregation of the rest of the chromosomes has also been shown to be dependent upon repeated sequences in the

pericentric heterochromatin (Hawley *et al.*, 1992). These sequences are responsible for the co-orientation of the homologues for correct meiotic segregation. This may define a general role for heterochromatin (Irick, 1994).

The *Responder* locus (*Rsp*) of the *Segregation Distorter* (*SD*) system has also been shown to depend upon heterochromatin. Haploid spermatids with a sensitive allele of *Rsp* abort, without producing functional sperm, when in the presence of *trans*-acting products of the dominant *Segregation distorter* (*Sd*) and / or *the Enhancer of Segregation distorter* [*E(SD)*] loci (Temin *et al.*, 1991). Sensitivity is known to be correlated with the copy number of a 240bp repeat unit in centromeric heterochromatin of chromosome 2R. Mechanisms of action put forward to explain how these loci work include the repeats forming a binding site that in the presence of *Sd* or the *E(SD)* products, causes a failure in normal chromosomal condensation (Ganetzky, 1977). *Cis*-acting elements in the proximal heterochromatin have also been implicated in the regulation of precocious X chromosome inactivation, a process that is thought to be necessary for normal spermatogenesis (Rhaman and Lindsley, 1980). The regulation and / or function of several other genes are also probably intimately linked with heterochromatin *e.g.* the *Mst40* sequences at polytene subdivision 40 on chromosome 2L (Russell and Kaiser, 1994).

Other work involving the *rolled* locus and five other heterochromatin linked genes on 2R, has suggested that at least some loci located in the heterochromatin are peculiar in that they need a heterochromatic environment to function correctly. It has been suggested that although the specific sequence of satellite DNAs can be very variable, the sequences themselves are unimportant other than to create the correct genomic environment for genes like *rolled* (Eberl *et al.*, 1993) and the *light* gene on chromosome 2L to function (Devlin *et al.*, 1990a, b) *i.e.* the factors controlling gene expression *etc.* may need a hetrochromatic enviroment to work. Thus satellite DNA sequences function to create heterochromatin. Why genes like *rolled* and *light* need a heterochromatic environment or why they evolved this need in the first place is unknown.

1.4.3 Middle Repetitive Sequences

1.4.3.1 Transposable Elements in *Drosophila melanogaster.*

Evidence suggests that most dispersed repeated DNA sequences in *Drosophila* are transposons (Young 1979) and it has been estimated that these sequences make up approximately 10% ($1.6x10^8$bp) of the *Drosophila* genome (Manning *et al.,* 1975). As many as fifty transposable element families have been estimated to exist in *Drosophila melanogaster* (Finnegan and Fawcett, 1986), and a single individual may contain up to 50 transposon families with an average of 50 copies. Recent studies suggest that this may be an overestimation of the number of element families as recently isolated elements share identity with previously characterised identified transposable elements. *e.g. Sancho* 1, *Sancho* 2, *jockey* and *wallaby* are probably the same element (Mizrokhi *et al.,* 1988; de Frutos *et al.,* 1992). The wide range of transposable elements can be subdivided into distinct classes, each reflecting their mechanism of transposition and their organisation in terms of repeat structures.

All eukaryotic transposable elements may be divided into two classes based on their transposition mechanism (Finnegan 1989). Two groups of element fall into Class I: both Class 1 groups transpose via RNA intermediates and encode a reverse transcriptase and are termed retroelements. One type are flanked by two long terminal repeats (LTRs), which show strong similarity to retroviruses. These retroelements are termed retrotransposons. The second group of Class I elements lack terminal repeats and strongly resemble the L1 family of elements found in mammalian genomes (Brookfield, 1993), they are termed retroposons. Only a few retroposon families have been identified in *Drosophila melanogaster. e.g. F* elements (Dawid *et al.,* 1981); *G* elements (Di Nocera and Dawid 1983); *I* elements (Fawcett *et al.,* 1986) and *jockey* (Mizrokhi *et al.,* 1988), whereas many well characterised retroelement families have now been described *e.g. 412* (Will *et al.,* 1981); *mdg1*, (Ilyin *et al.,* 1980b); *mdg3*, (Ilyin *et al.,* 1980c); *297*, (Potter *et al.,* 1979); *mdg4(gypsy)*, (Bender *et al.,* 1983a; Ilyin *et al.,* 1980d); *copia*, (Mount and Rubin 1985) and *B104 / roo* (Scherer *et al.,* 1981,1982; Mererowitz and Hogness, 1982). Many retrotransposons are known to undergo some form of developmental regulation of gene expression, such as the elements *297, 17.6 ,copia, 412* and *mdg1*. The latter three elements share a similar expression pattern throughout development, with high expression at embryo and pupal stages and lower at the adult stage. Cavarec and Heidmann (1993) have identified a

region within *copia* which appears to be capable of binding homeoproteins, (which are responsible for many developmentally regulated stages of gene expression in *Drosophila*), Other studies have shown that *copia* expression is indeed affected by the presence of various homeoproteins. This regulatory region is found downstream of the LTR, between the enhancer and the LTR, which is a region where the regulatory sequences are known to lie in most retrotransposons (*e.g.* Parkhurst and Corces, 1987). The retrotransposons are thought to make up approximately 1.8% or $2.8x10^6$ bp of the *Drosophila melanogaster* genome (Zachar and Bingham, 1989).

Class II elements can be divided into two groups also, both groups transpose directly from DNA to DNA. Only a few members of the Class II families have been described in *Drosophila melanogaster*. These include the hybrid dysgenesis determinants *P* and *hobo* which have short inverted repeats at their termini. A second group of Class II elements are internally repetitious and posses long terminal inverted repeats, these include the *fold-back* (*FB*) elements.

1.4.4 Clustered Scrambled Repeats of *Drosophila melanogaster*.

This type of repetitive DNA has been shown to occur some 15 times in the *Drosophila* genome (Wensink *et al.*, 1979). In organisation it consists of a complex arrangement of short moderately repetitive elements that are densely spaced in large clusters. Each cluster therefore has a scrambled arrangement of a subset of elements from some larger population.

The sizes of identified clusters range from 3kb (*e.g.* pDM1, Wensink *et al.*, 1974; Wensink, 1978) to 15kb (Wensink *et al.*, 1979). The size of the internal repeated regions varies from 100bp to 1000bp with an average of *ca.* 200bp and these have very different arrangements across the clusters as a whole, being repeated both within the same cluster and between different clusters. It is estimated that copies of the internal regions from within one clustered scrambled repeat are scattered in at least 1000 chromosomal regions (Wensink *et al.*, 1979). An estimate of the minimum number of internal repeated regions necessary to account for the observed homologies between different regions of the clustered scrambled repeats is 52 though this was reported to be a very conservative estimate.

The clusters themselves are scattered across the genome, being found in both euchromatic and heterochromatic locations. They have been shown to be present in several previously characterised regions of *Drosophila* DNA such as one of the heatshock genes (Livak *et al.,* 1978) and in the *cDm412* clone that contains the mobile element *412* (Finnegan *et al.,* 1978; Potter *et al.,* 1979). The homology with the *412* element occurs both within and without the element itself, and this observation may suggest the source of the clusters. If transposable elements were to transpose into each other (which is thought to be a relatively common event *e.g.* Dawid *et al.,* 1981; Scherer *et al.,* 1982; Di Nocera and Dawid, 1983; Young *et al.,* 1983) then when internal deletions occur the result would be a cluster of small fragments of moderately repetitive DNA.

1.4.5 Short Repetitive Sequences in *Drosophila melanogaster.*

The *Drosophila* genome contains short repetitive sequences that are often but not exclusively located within transcribed genes (McGinnis *et al.,* 1984; Wharton *et al.,* 1985; Haynes *et al.*; Magoulas and Hickey, 1992). One class of short DNA repeats are those that encode polyamino domains and which are found predominantly within developmentally regulated transcripts. This class of repetitive DNA comprises distinct families of repeats such as the *opa* repeat (see Section 6.5.1.4) consisting of CAG and CAA triplets which encode a polyglutamine domain (Wharton *et al.,* 1985); the *pen* repeat which is rich in GGN triplets and encodes a glycine-rich domain (Haynes *et al.,* 1987); the *PRD* repeat encoding a sequence of CA(T/C)CCG, which is translated as histidine - proline repeats (Frigerio *et al.,* 1986) and the *ala* repeat consisting of GCN triplets which encodes a polyalanine rich domain (Magoulas and Hickey, 1992).

1.4.5.1 The *opa* Repeat.

The *opa* repeat (also known as the *strep* repeat) was originally identified in the *Notch* gene of *Drosophila melanogaster.* As mentioned above the repeat consists of a $(CAX)_n$ repeat, where n is normally less than 30 and where X is either an A or G and very occasionally a C (Steward, 1987). It is found in a large number of developmentally regulated genes in *Drosophila e.g. Antennapaedia* (Laughton *et al.,* 1986-unpublished data); *twist* (Thisse *et al.,* 1988); *single-minded* (Crews *et al.,* 1988), *hunchback* (Tautz *et al.,* 1987) and *Notch* Kidd *et al.,* (1983); Wharton *et al.,* (1985); Kidd *et al.,* (1986). The *opa* repeat has also been found in a large number of

other organisms such as humans (*e.g.* the androgen receptor gene [Lubahn *et al.*, 1988) and in mice *e.g.* the *Mop* repeats (Duboule *et al.*, 1987). Work using *opa* repeat containing fragments as probes on both *Drosophila* and murine Northern blots has shown a high degree of cross-hybridisation with many different size-classes of polyA$^+$ RNA. This has also been demonstrated with various cDNA libraries, where over a third of recombinant clones were shown to contain *opa* repeats (Duboule *et al.*, 1987; Wharton *et al.*, 1985). In transcripts the repeat has been found in the 3′ or 5′ untranslated regions as well as in open reading frames where it tends to encode a polyglutamine domain (*e.g.* Wharton *et al.*, 1985; Grabowski *et al.*, 1991). It appears that most, if not all, of the messages in which the *opa* repeat is found, are developmentally regulated (Duboule *et al.*, 1987). This finding has led to suggestions that there may possibly be an (unidentified) role for the repeat in transcripts. However, studies in which the *opa* repeat was deleted from the *Notch* gene (Lieber *et al.*, 1993) argue against this, as these studies could detect no phenotypic variation in *Notch opa*$^-$ flies suggesting that the repeat may not have a role (at least in the *Notch* gene), or that if it does, it may have a subtle phenotype. The origin of the *opa* repeat within transcripts is also a matter of controversy. It is possible that the genes containing *opa* repeats all evolved from a common ancestor which now has many different roles and functions (Scott and Carroll, 1987). From an evolutionary point of view however, and bearing in mind the large numbers of very different genes which contain *opa* repeats this seems to be unlikely although it cannot be ruled out.

The significance, of these repetitive sequences in either protein function (if they are translated) and / or evolution is unknown. As they are found in a large number of developmentally important transcripts, and indeed have been used as a means of isolating novel developmentally regulated genes from *Drosophila* (Magoulas and Hickey, 1992), it would seem logical to assume some function for them. The number of repeats however can vary quite considerably within the same gene in different populations *e.g.* 5-30 copies of *opa* repeat exist in *Notch* (Wharton *et al.*, 1985). They can also be eliminated from coding regions with no apparent phenotypic effects on the fly. They can also occur in non-coding DNA (Section 1.4.5.1). It may be that short repetitive regions in genes have no function or one that is not readily detectable.

Chapter Two

Chapter Two - Materials and Methods.

2.1 Materials.

2.1.1 Chemicals and Biochemicals.

General chemicals of analytical reagent grade were obtained from BDH Ltd., Poole, U.K.; Boehringer Mannheim, Lewes, U.K.; Formachem Ltd, Strathaven, U.K.; FSA Laboratory Supplies, Loughborough, U.K.; Koch-Light Ltd, Haverhill, U.K.; May and Baker, Dagenham, U.K..

Ampicillin, dithiothreitol (DTT), ethidium bromide, Ficoll, Mops buffer, benzamidine hydrochloride, bovine serum albumin, Tris buffer, Triton X-100, sodium dodecyl sulfate (SDS) and diethyl pyrocarbonate (DEPC) were obtained from Sigma (London) Chemical Co., Poole, Dorset, U.K..

Acrylamide, ammonium sulphate (enzyme grade), hydrochloric acid (HCL, Aristar), 2-mercaptoethanol, N,N'-methylenebisacrylamide, N,N,N',N'-tetramethylenediamine and ATP, were obtained from Boehringer Mannheim, Lewes, U.K..

Bromophenol blue and caesium chloride were obtained from BDH Chemicals Ltd., Poole, Dorset, U.K..

Xylene cyanol was obtained from IBI Ltd., Cambridge, U.K..

Agarose, 5-bromo-4-chloro-3-indoyl-β–galactoside (X-gal), isopropyl-β–thiogalactoside (IPTG), phenol (ultrapure) and urea (ultrapure) were obtained from Gibco, BRL Ltd., Paisley, U.K..

Bactotryptone, yeast extract and bactoagar (agar) were obtained from Difco, Detroit, USA. Oxoid No.1 agar and trypticase soy broth were obtained from Oxoid Ltd., London, U.K..

[α-^{32}P]CTP and [α-^{35}S]-dATP (code SJ.304) were obtained from Amersham International plc., Amersham, U.K..

The Sequenase version 2.0 sequencing kit was obtained from United States Biochemical Corporation (distributed by Cambridge BioScience, Cambridge, U.K.).

All other chemicals were of analytical reagent grade and were obtained from one of the following suppliers: BDH Ltd., Poole, U.K.; Formachem Ltd., Strathaven, U.K.; FSA Laboratory Supplies, Loughborough, U.K.; Koch-Light Ltd., Haverhill, U.K..

2.1.2 Enzymes.

All restriction enzymes, T4 DNA ligase and T4 Polynucleotide kinase were obtained either from BRL, Gibco Ltd., Paisley, UK., or from Promega Ltd, Southampton, U.K.. Ribonuclease A (RNase A) and Lysozyme was obtained from Sigma (London) Chemical Co., Poole, Dorset, U.K.. Alkaline phosphatase (calf intestine) was obtained from Boehringer Mannheim, Lewes, U.K.

2.1.3 Chromatography Media.

Sephadex G-50 was supplied by Pharmacia, Milton Keynes, U.K.

2.2 Methods.

2.2.1 pH Measurements.

All pH measurements were made with a Corning pH meter 240 probe calibrated at room temperature using standards of pH7.0 and pH4.0 prepared from tablets obtained from the manufacturer.

2.2.2 Distilled Water.

Glass distilled water stored in polythene containers was used in all experiments.

2.2.3 Dialysis Membranes.

Dialysis Membranes were obtained from Scientific Instruments Centre Ltd., London and were boiled for 5 minutes in 1% (w/v) EDTA;pH7.0, stored in 70% ethanol and rinsed in distilled water prior to use.

2.2.4 Spectrophotometric Determination of Nucleic Acid Concentrations.

Nucleic acid concentrations were determined spectrophotometrically at 260nm and at 280nm to check for contamination. (Sambrook *et al.*, 1989). In a 1 cm path length

quartz cuvette an absorbance of 1.0 corresponds to 50μg ml^{-1}. for double stranded DNA, 40μg ml^{-1} for single stranded DNA and approximately 20μg ml^{-1} for single stranded oligonucleotides.

2.2.5 Sterilisation.

Solutions for the preparation of growth media and solutions used in manipulation of nucleic acids were autoclaved at 120°C for 15 minutes, with some supplements and buffers being autoclaved at 108°C for 10 minutes. Because of their probable heat lability, ampicillin and IPTG were sterilised by filtration through 0.22μm pore-sized Millex G.V. filters (Millipore Ltd., U.K.) into sterile bottles.

2.2.6 Gel Drying and Autoradiography.

Polyacrylamide gels were dried onto Whatman 3MM paper using an ATTO RapidDry Gel Drier connected to a HOWE Refrigerated Solvent Trap and a Brook Crompton Parkinson multi-purpose vacuum unit. ^{32}P or ^{35}S in polyacrylamide gels or on filters was detected by exposure to either Amersham Hyperfilm or Fuji RX X-ray film using intensifying screens at -70°C. ^{35}S was detected by exposure to film in the absence of intensifying screens at room temperature. Films were developed by a X-OMAT Automatic Film Processor Model Compact X2, (X-Ograph Ltd.).

2.2.7 Pre-flashing of Autoradiography Film.

This was performed as described by Hahn (1983) and the manufacturers (Amersham and Kodak). It was done to linearise the response of the film to radioactive signal intensity. Where differing films were used then all film was calibrated to give the same linear response to signal intensity. Pre-flashing was performed by flashing the film at a distance of two metres, one to three times with a flashgun (Hanimex X140) covered with a Wratten Filter Nº 23A (red) and variable layers of semi-transparent material (autoclave tape). The film was then developed under standard conditions (Section 2.2.6) and cut into strips which were then analysed spectrophotometrically. The film gave an o/d reading of 0.05 Abs units at 650nm as recommended.

2.2.8 Buffer Solutions.

2.2.8.1 Electrophoresis (DNA).

10X TBE Buffer:	0.9M Tris-borate, 0.02M $Na_2EDTA.2H_2O$,
50X TAE Buffer:	2M Tris-acetate, 0.05M $Na_2EDTA.2H_2O$, pH was adjusted to 8.2 with acetic acid.
1X Alkaline Electrophoresis Buffer:	0.05M NaOH, 0.01M EDTA;pH8.0,
Single colony gel Buffer: (in 1X TBE buffer)	2% Ficoll, 1% SDS, 0.1% Bromophenol Blue, 0.1% Orange G.
Loading Buffer: (in 1X TBE buffer)	10% Ficoll, 0.1% Bromophenol Blue, 0.1% Orange G.
[dl] Alkaline Loading Buffer:	18% Ficoll, 0.15% Bromocreosol Green, 0.25% Xylene Cyanol FF in 0.3M NaOH, 6mM EDTA;pH8.0.

2.2.8.2 Electrophoresis (RNA).

10X MOPS: (adjusted to pH 7.0 with NaOH)	0.2M Morpholinopropanesulphonic acid, 80mM NaAc, 10mM Na_2EDTA,
MMF:	50% Formamide, 6% Formaldehyde, 0.1M MOPS.
Formaldehyde gel loading buffer:	50% Glycerol, 1mM Na_2EDTA, 0.4% Bromophenol Blue, 0.4% Xylene Cyanol

2.2.8.3 DNA Manipulation.

Restriction and Ligation buffers were obtained from Gibco, BRL Ltd., Paisley, U.K.

ATP stock solution (100mM): 60mg rATP in 0.8ml distilled water, pH adjusted to 7.0 with 0.1M NaOH and made up to 1ml with sterile water (Stored at -20°C).

TE Buffer: (pH to 7.0)	10mM Tris, 1mM EDTA.
10X Klenow Buffer: (Stored at -20°C)	0.5M Tris, 0.1M $MgSO_4$, 1mM DTT, 500µg ml^{-1} bovine serum albumen.
Nick Translation Buffer:	as 10X Klenow Buffer.
10X Kinase Buffer: (pH to 7.6)	100mM KCl, 70mM $MgCl_2$, 5mM DTT

2.2.8.4 DNA and RNA Hybridisation Buffers.

20X SSC: (pH to 7.0)	3M NaCl, 0.3M NaCitrate
20X SSPE: (pH to 7.4)	3.6M NaCl, 200mM NaH_2PO_4, 20mM EDTA,
Denaturing Solution:	1.5M NaCl, 0.5M NaOH.
Neutralising Solution: (pH to 8.0)	1.0M Tris, 1.5M NaCl,
Denhardts Solution:	0.2mg ml^{-1} BSA, 0.2mg ml^{-1} Ficoll-400, 0.2mg ml^{-1} Polyvinyl pyrolidone.

Sonicated Salmon Sperm DNA: A 10mg ml^{-1} solution was prepared by dissolving dried salmon sperm DNA (Sigma (London) Chemical Co., Poole, Dorset, U.K.), by vigorous mixing overnight. This was then treated for 7 minutes with a Brandon Ultrasonic Processor (Model B-15) in a DAWE Acoustic Booth. Typically the DNA was sheared to lengths between 200 to 1000bp.

2.2.8.5 DNA / RNA Extraction, Purification and General Purpose Buffers.

Phenol: All Phenol used in the purification of DNA and RNA contained 0.1% 8-hydroxy-quinolene and was buffered against 0.25M Tris HCl;pH 8 for DNA work, or buffered against water for use with RNA to give a solution of pH 5-6

Chloroform: A mixture of chloroform and isoamyl alcohol (24:1) was used to reduce foaming during extraction and improve phase separation of the aqueous and organic phases.

Phage Buffer: 0.1M NaCl,
0.017M $MgSO_4$,
2% gelatine,
1mM Tris (pH 7.5)

Birnboim Doly Buffer I: 50mM Glucose,
25mM Tris,
10mM EDTA (pH8.0).

Birnboim Doly Buffer II: 0.2M NaOH,
1% SDS

Birnboim Doly Buffer III: 5M KAc (pH4.8),
made by mixing equal volumes of 3M CH_3COOK and CH_3COOH.

1X Phosphate Buffered Saline (PBS):0.13M NaCl,
2mM KCl,
7mM Na_2HPO_4,
3mM $KHPO_4$,

Trichloroacetic Acid (TCA) 100% (w/v) Solution: 500g TCA plus 227ml distilled water

2.2.8.6 Mounting Media.

Gelatin / Glycerol Mountant was prepared as described by Ashburner (1989). 7g of gelatin (Merck no. 4078) was allowed to swell in 42ml of dH_20 for several minutes and then dissolved by placing in a beaker of boiling water. 50ml of glycerol was then added along with a crystal of phenol to act as a bactericide. The mix was stored at 4°C when not required and remelted at 45°C for use.

2.3 Microbiological Techniques.

2.3.1 Plasmid and Bacteriophage Vectors used in this Study.

2.3.1.1) Bacteriophage Vectors.

The bacteriophage vectors used throughout this study are described in Table 2.1a.

2.3.1.2) Plasmid Vectors.

The plasmid vectors used throughout this study are described in Table 2.1b.

2.3.1.3 Bacterial Strains.

All bacterial strains are derived from *Escherichia coli* K-12 and are described in Table 2.2

2.3.1.4 Recombinant Clones.

The recombinant plasmids and bacteriophages used, including those constructed in this study are described in Tables 2.3 and 2.4.

2.3.2 *Escherichia coli* Growth Media: (Sterilised by autoclaving at 15 p.s.i. for 25 minutes):

L-Broth- (per litre)	Bactotryptone	10g
	Yeast extract	5g
	NaCl	10g
	(+5ml 20% (w/v) glucose)	
	pH to 7.5 with NaOH	
L-Agar-(per litre)	As for L-broth, plus 15g Difco-agar.	
2xTY-Medium- (per litre)	Bacto-tryptone	16g
	Yeast extract	10g
	NaCl	5g
NZCYM Medium-(per litre)	NZ amine	10g
	NaCl	5g
	Bacto-Yeast extract	5g
	casamino acids	1g
	$MgSO_4.7H_2O$	2g
	pH to 7.0 with 5M NaOH	

Table 2.1a) Bacteriophage vectors

Vector	Description	Library Constructed by	Vector reference
λ EMBL 3	Replacement vector used to construct the *CS* genomic library	Vince Pirotta. (Pirotta *et al.*, 1983).	Frischauf *et al.*, 1983; Karn *et al.* , 1980
λ EMBL 4	Replacement vector used to construct the *OrR* genomic library	Vince Pirotta. (Pirotta *et al.*, 1983).	Frischauf *et al.*, 1983; Karn *et al.* , 1980
λ GEM 11	Replacement vector used to construct the *OrR* genomic library	S. Tomlinson.	Frischauf *et al.*, 1983; Karn *et al.*, 1984
λ ZAP II	Insertion vector containing an internal plasmid sequence (phagemid). Used to construct the *Drosophila* head cDNA library	C. Milligan	Short *et al.*, 1988

Table 2.1b) Plasmid vectors

Vector	Description	Vector reference
pUC18/pUC19	2.69kb general purpose cloning vectors yielding high copy numbers. Recombinants identified by histochemical screening	Messing (1983) Norrander *et al.*, (1983) and Yanisch-Perron *et al.* ,(1985)
pBluescript SK-/KS+	2.96kb general purpose cloning plasmids based on the pUC vectors. Used in this study to subclone genomic fragments and was the vector excised from λ ZAP	Short *et al.*, (1988) and Stratagene product literature.
pBR322	4.36kb vector used to subclone several genomic fragments from the chromosomal walks.	Bolivar *et al.*, (1977b) and Sutcliffe (1979)
pBR329	4.2kb vector used to subclone several genomic fragments from the chromosomal walks.	Covarrubias and Bolivar, (1982)

Table 2.2) Bacterial strains used in this study.

Strain	Genotype	Comments	Reference
DS941	*recF*143, *proA*7, *str*31, *thr*1, *leu*6, *tsx*33, *mtL*1, *his*4, *argE*3, *lacY*1, *galK*2, *ara*14, lambda-, *lacI*^{q}, *lacZM*15, *lacY*$^{+}$	Used for general purpose cloning with blue/white selection.	Horii and Clarke (1973)
LE392	F-, *hsdR*574, (r_K^-, m_K^+), *supE*44, *supF*58, *lacY*1 or Δ(lac1ZY)6, *galk*2, *galT*22, *metB*1, *trpR*55	General purpose permissive strain.	Murray *et al.*, (1977)
NM621	*hsd*R, *mcr*A, *mcr*b, *SupE*44, *recD*1009	General purpose permissive strain.	Whittaker *et al.*, (1988)
XL1-BLUE	*recA*1, *endA*1, *gyrA*96, *thi*-1, *hsdR*17, *supE*44, *relA*1, *lac*, [F' *proAB*, $\overline{\text{lacI}^q}$ Z*ΔM15*, Tn10(tet^r)]	General purpose strain for cloning and a permissive host for lambda.	Bullock *et al.*, (1988)

Table 2.3) Plasmid clones used in this study.

Clone Name	Vector	Cloned as:	Insert Size (kb)	Description	Reference (Section Number:-)
pBluescriptII™ KS+ SK-	----------	----------	2.96	Ap^r cloning vector derived from pUC. SK and KS refer to the orientation of the polylinker (a *Sac*I-*Kpn*I fragment) + and - refer to the orientation of the M13 replication origin.	2.3.1
pI1	pBSII SK-	*Eco* RI-*Xho* I	0.65	Plasmid clone isolated from the *eya* head cDNA library	6.2.2
pI1AvaI	pBSII KS+	*Ava* I-*Ava* I	0.25	Subclone from the 3' end of pI1	6.2.2
pI1RIAvaI	pBSII KS+	*Eco* RI-*Ava* I	0.40	Subclone from the 5' end of pI1	6.2.2
pC10	pBSII SK-	*Eco* RI-*Xho* I	2.75	Plasmid clone isolated from the *eya* head cDNA library	6.2.2
pC10(4)	pBSII KS+	*Eco* RI-*Eco* RI	0.72	Subclone from the 5' end of pC10	6.2.2
pC10(5)	pBSII KS+	*Xho* I-*Ava* I	0.25	Subclone from the 3' end of pC10	6.2.2
pC10(6)	pBSII KS+	*Eco* RI-*Xho* I	0.75	Subclone from the 5' end of pC10	6.2.2
pC10(7)	pBSII KS+	*Eco* RV-*Xho* I	1.0	Subclone from the 5' end of pC10	6.2.2
pC10(8)	pBSII KS+	*Ava* I-*Eco* RV	0.75	Subclone from the 3' end of pC10	6.2.2
pC10(9)	pBSII KS+	*Xho* I- *Eco* RV	1.0	Subclone from the 3' end of pC10	6.2.2
prpC	pBSII SK-	*Eco* RI-*Xho* I	0.6	Plasmid clone isolated from the *eya* head cDNA library	6.2.3
prpEa	pBSII SK-	*Eco* RI-*Xho* I	0.85	Plasmid clone isolated from the *eya* head cDNA library	6.2.3
prpEb	pBSII SK-	*Eco* RI-*Xho* I	0.85	Plasmid clone isolated from the *eya* head cDNA library	6.2.3
prpFa	pBSII SK-	*Eco* RI-*Xho* I	0.70	Plasmid clone isolated from the *eya* head cDNA library	6.2.3
KE2(1.8)	pBSII KS+	*Eco* RI-*Eco* RI	1.8	Plasmid clone of λKE2 insert	5.10
B10.125	pBSII KS+	*Eco* RV	1.25	cDNA fragment from a 12-24 hr embryonic cDNA library	5.10
M23	pBSII SK-	*Eco* RI-*Xho* I	3.5	Plasmid excised from λM23 cDNA clone	5.9.7
1.0kb λAGO02	pBSII KS+	*Eco* RI-*Eco* RI	1.0	*opa* repeat containing subclone from the distal end of λAGO02	4.2.4
1.0Ex(1)	pBSII KS+	*Eco* RI-*Sst* I	0.97	*Exo* III deletion product of the 1.0kb λAGO02 subclone.	6.3.5
1.0Ex(2)	pBSII KS+	*Eco* RI-*Sst* I	0.75	*Exo* III deletion product of the 1.0kb λAGO02 subclone.	6.3.5

Table 2.3) Plasmid clones used in this study.

Clone Name	Vector	Cloned as:	Insert Size (kb)	Description	Reference (Section Number:-)
1.0Ex(3)	pBSII KS+	*Eco* RI-*Sst* I	0.65	*Exo* III deletion product of the 1.0kb λAGO02 subclone.	6.3.5
1.0Ex(4)	pBSII KS+	*Eco* RI-*Sst* I	0.55	*Exo* III deletion product of the 1.0kb λAGO02 subclone.	6.3.5
1.0Ex(5)	pBSII KS+	*Eco* RI-*Sst* I	0.35	*Exo* III deletion product of the 1.0kb λAGO02 subclone.	6.3.5
1.0Ex(6)	pBSII KS+	*Eco* RI-*Sst* I	0.25	*Exo* III deletion product of the 1.0kb λAGO02 subclone.	6.3.5
1.0Ex(7)	pBSII KS+	*Eco* RI-*Sst* I	0.15	*Exo* III deletion product of the 1.0kb λAGO02 subclone.	6.3.5
2.3kb λ952-04C9	pBSII KS+	*Hin*dIII-*Hin*dIII	2.3	2.3kb subclone from the genomic lambda clone λ952-04C9	5.9.6, 6.2.2
2.3Ex(1)	pBSII KS+	*Hin* dIII-*Sst* I	2.0	*Exo* III deletion product of the 2.3kb λ952-04C9 subclone.	5.9.6, 6.2.2
2.3Ex(2)	pBSII KS+	*Hin* dIII-*Sst* I	1.75	*Exo* III deletion product of the 2.3kb λ952-04C9 subclone.	5.9.6, 6.2.2
2.3Ex(3)	pBSII KS+	*Hin* dIII-*Sst* I	1.70	*Exo* III deletion product of the 2.3kb λ952-04C9 subclone.	5.9.6, 6.2.2
2.3Ex(4)	pBSII KS+	*Hin* dIII-*Sst* I	1.60	*Exo* III deletion product of the 2.3kb λ952-04C9 subclone.	5.9.6, 6.2.2
2.3Ex(5)	pBSII KS+	*Hin* dIII-*Sst* I	1.50	*Exo* III deletion product of the 2.3kb λ952-04C9 subclone.	5.9.6, 6.2.2
2.3Ex(6)	pBSII KS+	*Hin* dIII-*Sst* I	1.30	*Exo* III deletion product of the 2.3kb λ952-04C9 subclone.	5.9.6, 6.2.2
2.3Ex(7)	pBSII KS+	*Hin* dIII-*Sst* I	1.20	*Exo* III deletion product of the 2.3kb λ952-04C9 subclone.	5.9.6, 6.2.2
2.3Ex(8)	pBSII KS+	*Hin* dIII-*Sst* I	1.05	*Exo* III deletion product of the 2.3kb λ952-04C9 subclone.	5.9.6, 6.2.2
p94.R1	pBR329	*Eco* RI-*Eco* RI	6.0	Subclone from λ9405	3.7, 4.2.4
p94.R2	pBR329	*Eco* RI-*Eco* RI	2.8	Subclone from λ9405	3.7
p94.R3	pBR329	*Eco* RI-*Eco* RI	1.76	Subclone from λ9405	3.7
p94.R4	pBR329	*Eco* RI-*Eco* RI	1.7	Subclone from λ9405	3.7
p94.R5	pBR329	*Eco* RI-*Eco* RI	1.7	Subclone from λ9405	3.7
p94.R7	pBR329	*Eco* RI-*Eco* RI	1.25	Subclone from λ9405	3.7
p94.R8	pBR329	*Eco* RI-*Eco* RI	0.85	Subclone from λ9405	3.7
3.2kb λAGC01	pBSII KS+	*Eco* RI-*Bam*HI	3.2	Subclone from the distal end of the lambda clone λAGC01	4.2.4
3.5kb λ952-04C9	pBSII KS+	*Eco* RI-*Hin* dIII	3.5	Subclone from the lambda clone λ952-04C9	5.9.6, 6.2.2

Table 2.3) Plasmid clones used in this study (Cont'd).

Clone Name	Vector	Cloned as:	Insert Size (kb)	Description	Reference (Section Number:-)
1.8kb λ9915	pBSII KS+	*Sal* I-*Sal* I	1.8	Subclone from the distal end of the lambda clone λ9915	3.8
2.3kb λ952-04C9	pBSII KS+	*Bam*HI-*Hin*dIII	2.3	Subclone from the lambda clone λ952-04C9	5.9.6, 6.2.1
2.5kb λ952-04C9	pBSII SK-	*Eco*RI-*Eco*RI	2.5	Subclone from the lambda clone λ952-04C9	5.9.6, 6.2.1
3.9kb λ94C11	pBSII KS+	*Eco*RI-*Eco*RI	3.9	Subclone from the lambda clone λ94C11	3.6
4.9kb λ94C11	pBSII KS+	*Eco*RI-*Eco*RI	4.9	Subclone from the lambda clone λ94C11	3.6
4.1kb λ94C11	pBSII KS+	*Eco*RI-*Eco*RI	4.1	Subclone from the lambda clone λ94C11	3.6
1.4kb λ94C11	pBSII KS+	*Eco*RI-*Eco*RI	1.4	Subclone from the lambda clone λ94C11	3.6
1.9kb λAGCO2	pBSII KS+	*Eco*RI-*Hin* dIII	1.9	Subclone from the distal end of the lambda clone λAGCO2	3.5
p08C91	pBR329	*Eco*RI-*Eco*RI	4.8	Subclone from the lambda clone λ952-08C9	5.9.3
0.8kb λ952-08C9	pBSII KS+	*Eco*RI-*Bam*HI	0.8	Subclone from the p08C91 plasmid	5.9.3
p08C95	pBR329	*Eco*RI-*Eco*RI	1.7	Subclone from the lambda clone λ952–C5B	5.9.4
7.2kb λC5B03	pBSII KS+	*Eco*RI-*Bam*HI	7.2	Subclone from the proximal end of λC5B03	3.5
2.1kb λC5B03	pBSII KS+	*Bam*HI- *Bam*HI	2.1	Subclone from the distal end of the lambda clone λC5B03	3.5
4.3kb λ9915	pBSII KS+	*Eco*RI-*Eco*RI	4.3	Subclone from the proximal end of the lambda clone λ9915	3.10, 5.9.10, 6.4
3.2kb λ9915	pBSII KS+	*Sal* I-*Sal* I	3.2	Subclone from the proximal the lambda clone λ9915	3.8
2.45kb λC5B03	pBSII KS+	*Eco*RI- *Bam*HI	2.45	Subclone from the lambda clone λC5B03	3.5
0.5kb λAGCO2	pBSII KS+	*Eco*RI-*Eco*RI	0.5	Subclone from the proximal end of the lambda clone λAGO02	3.2.2
p94C15-R2	pBSII SK-	*Eco*RI-*Eco*RI	5.7	Subclone from the lambda clone λ94C15	4.2.1
0.6kb λ952-04C9	pBSII KS+	*Ava*I -*Xho*I	0.6	Subclone from the lambda clone λ952-04C9	6.2.2
1.4kb λ952-04C9	pBSII KS+	*Eco*RI-*Sst*I	1.4	Subclone from the distal end of the lambda clone λ952-04C9	5.9.6, 6.2.2
1.7kb λ952-04C9	pBSII KS+	*Eco*RI-*Eco*RI	1.7	Subclone from the lambda clone λ952-04C9	5.9.6
6.0kb λ952-04C9	pBSII KS+	*Eco*RI-*Eco*RI	6.0	Subclone from the lambda clone λ952-04C9	5.9.6
1.7kb λ952-08C9	pBSII KS+	*Eco*RI-*Eco*RI	1.7	Subclone from the proximal end of the clone λ952-08C9	5.9.6

Table 2.3) Plasmid clones used in this study (Cont'd).

Clone Name	Vector	Cloned as:	Insert Size (kb)	Description	Reference (Section Number:-)
1.8kb λ952-04C9	pBSII KS+	*Eco*RI-*Eco*RI	1.8	Subclone from the lambda clone λ952-04C9	5.9.6, 6.2.2
0.8kb λ952-04C9	pBSII SK-	*Eco*RI-*Eco*RI	0.8kb	Subclone from the lambda clone λ952-04C9	5.9.6, 6.2.2
1.7kb λS2	pBSII KS+	*Hin*dIII-*Hin*dIII	1.7	Subclone from the lambda clone λS2	3.22, 4.2.4
0.8kb λS2	pBSII KS+	*Sal* I-*Hin* dIII	0.8	Subclone from the lambda clone λS2	3.22, 4.2.4
1.4kb λS2	pBSII KS+	*Sal* I-*Hin* dIII	1.4	Subclone from the lambda clone λS2	3.22, 4.2.4
2.5kb λS4	pBSII KS+	*Sal* I-*Hin* dIII	2.5	Subclone from the lambda clone λS4	3.22, 4.2.4
3.8kb λS4	pBSII KS+	*Sal* I-*Hin* dIII	3.8	Subclone from the lambda clone λS4	3.2.2, 4.2.4
3.5kb λS2	pBSII KS+	*Sal* I- *Sal* I	3.5	Subclone from the lambda clone λS2	3.2.2, 4.2.4
1.0kb λS4	pBSII KS+	*Sal* I- *Sal* I	1.0	Subclone from the lambda clone λS4	3.2.2, 4.2.4
0.38kb λS5	pBSII KS+	*Hin*dIII-*Hin*dIII	3.8	Subclone from the lambda clone λS5	3.2.2, 4.2.4
6.8kb λS5	pBSII KS+	*Hin*dIII-*Hin*dIII	6.8	Subclone from the lambda clone λS5	3.2.2, 4.2.4
4.5kb λS5	pBSII KS+	*Sal* I- *Sal* I	4.5	Subclone from the lambda clone λS5	3.2.2, 4.2.4
7.2kb λS2	pBSII KS+	*Sal* I- *Sal* I	7.2	Subclone from the lambda clone λS2	3.2.2, 4.2.4
1.8kb λM283	pBSII KS+	*Eco*RI- *Sal* I	1.8	Subclone from the lambda clone λM283	4.2.2
1.7kb λM283	pBSII KS+	*Eco*RI- *Sal* I	1.7	Subclone from the lambda clone λM283	4.2.2
1.0kb λH683	pBSII KS+	*Eco*RI -*Ava*I	1.0	Subclone from the lambda clone λH683	4.2
3.5kb λH683	pBSII KS+	*Ava*I-*Ava*I	3.5	Subclone from the distal end of the lambda clone λH683	3.10
p8011R3	pBR329	*Eco*RI-*Eco*RI	1.7	Subclone from the lambda clone λM281	4.2.2
3.3kb λM281	pBSII KS+	*Eco*RI-*Eco*RI	3.3	Subclone from the proximal end of the lambda clone λM281	4.2.2
2.6kb λM281	pBSII KS+	*Eco*RI-*Eco*RI	2.6	Subclone from the lambda clone λM281	4.2.1
3.3kb λH382	pBSII KS+	*Eco*RI- *Sal* I	3.3	Subclone from the distal end of the lambda clone λH382	4.2.1
2.6kb λM283	pBSII KS+	*Eco*RI- *Sal* I	2.6	Subclone from the lambda clone λM283	4.2.1
2.4kb λM281	pBSII KS+	*Eco*RI- *Eco*RI	2.4	Subclone from the proximal end of the lambda clone λM281	4.2.1

Table 2.3) Plasmid clones used in this study (Cont'd).

Clone Name	Vector	Cloned as:	Insert Size (kb)	Description	Reference (Section Number:-)
3.5kb H683	pBSII KS+	*Ava* I-*Ava* I	3.5	Subclone from the lambda clone λH683	4.2.1
p8011R4	pBR329	*Eco*RI-*Eco*RI	1.6	Subclone from the lambda clone λM281	4.2.1
0.9kb λH382	pBSII KS+	*Eco*RI-*Eco*RI	0.9	Subclone from the lambda clone λH382	4.2.1
1.4kb λH382	pBSII KS+	*Eco*RI- *Sal* I	1.4	Subclone from the lambda clone λH382	4.2.1
p79C1R6	pBR329	*Eco*RI-*Eco*RI	2.5	Subclone from the lambda clone λ798C10O4A	3.2.1
p0505R1	pBR329	*Eco*RI-*Eco*RI	7.8	Subclone from the genomic clone λ0510705B	4.2.2
p798051H2	pBR329	*Hin*dIII-*Hin*dIII	3.7	Subclone from the genomic clone λ798051	4.2.2
pC1004R4	pBR329	*Eco*RI-*Eco*RI	2.65	Subclone from the genomic clone λ798C1004C	4.2.2
p7011R1	pBR329	*Eco*RI-*Eco*RI	4.5	Subclone from the genomic clone λ798011	4.2.2
p7051R3	pBR329	*Eco*RI-*Eco*RI	2.8	Subclone from the genomic clone λ05R308B	4.2.2
p7011R2	pBR329	*Eco*RI-*Eco*RI	4.5	Subclone from the genomic clone λ798011	4.2.2
p0505R3	pBR329	*Eco*RI-*Eco*RI	2.7	Subclone from the genomic clone λ05R308B	3.2.1
p79C1R1	pBR329	*Eco*RI-*Eco*RI	6.6	Subclone from the genomic clone λ798C1004C	4.2.2
p798051R2	pBR329	*Eco*RI-*Eco*RI	4.0	Subclone from the genomic clone λ798051	4.2.2
p798C1004AR1	pBR329	*Eco*RI-*Eco*RI	5.6	Subclone from the genomic clone λ798C1004A	3.2.1
pR4H2	pBR329	*Hin*dIII-*Hin*dIII	7.3	Subclone from the lambda clone λR209	4.2.2
1.8kb λR209	pBSII KS+	*BamH*I-*Hin*dIII	1.8	Subclone from the lambda clone λR209	4.2.2
1.4kb λR209	pBSII KS+	*BamH*I-*Hin*dIII	1.4	Subclone from the lambda clone λR209	4.2.4
pActin5C	pBSII KS+	*Eco*RI-*Eco*RI	8.7	Genomic fragment containing the Actin5C gene	Chapter 5
pRas	pBSII SK-	*Hin* dIII- *Sal* I	2.6	cDNA containing the Drosophila *ras* gene	Chapter 5

Table 2.3) Plasmid clones used in this study. The clones in pBSII KS+ or SK- were constructed for this study, whereas the clones constructed in pBR322 were constructed for chromosomal walking (Jane Davies, personal communication). The pActin5C and pRas clones were donated by Steve Russell and Colin Milligan (Glasgow) for use with the Northern analysis.

Table 2.3a) Transposable element clones used in this study.

Clone Name	Copy N^{o} per Genome	Vector	Insert size(kb)	Cloned as	Description
pSpringer	*ca.*6	pBR322	4.5	*Sal* I - *Sal* I	Internal fragment of the element
pPelement	0 - 50	pBR322	2.9	*Hin* dIII - *Sal* I	Missing the terminal inverse repeats only
pDoc	Unknown	pBSIIKS+	4.7	*Bgl* II - *Bgl* II into *Bam* HI sites	Complete element
pPogo	Unknown	pBR322	2.1	*Sal* I - *Cla* I	Missing terminal inverse repeats only
pB104	*ca.*80	pSP65	5.7	*Sal* I - *Sal* I	Complete element
pCopia	*ca.*60	pBR322	5.2	*Bam* HI - *Bam* HI	Complete element
pmdg1	*ca.*25	pBR322	8.4	*Hin* dIII digest	Complete plus flanking sequences.
pmdg4	*ca.*10	pBSIIKS+	6.8	*Xho* I - *Xho* I	Complete element
pIelement	30-50	pBR322	2.8	*Hin* dIII - *Hin* dIII	Missing the terminal inverse repeats only
p412	*ca.*40	pAT153	12.0	*Eco* RI - *Eco* RI	Complete plus flanking sequences
p297	*ca.*30	pBR322	13.5	*Bam* HI - *Bam* HI	Complete plus flanking sequences from 99D
pHobo	20 - 50	pBR322	2.7	*Xho* I - *Xho* I	Missing small region at each end
pJockey	Unknown	pUC19	1.6	*Hin* dIII - *Eco* RI	Internal fragment of the element

Table 2.3a) Transposable element clones used in this study (Section 4.2.5). See Finnegan and Fawcett, (1986) and Berg and Howe, (1989), for complete references for each element. The clones were kindly donated to this study by Dr. Kevin O'Hare and Dr. David Finnegan.

Table 2.4) Bacteriophage clones used in this study.

Clone	Vector	Cloned as:	Size(kb)	Description	Reference
λ7H7	EMBL 4	Partial Sau3A into *Bam*HI	14.7	Genomic clone from the 19E1-19E2 region.	This study (Chapter three).
λ7A1	EMBL 4	Partial Sau3A into *Bam*HI	18.2	Genomic clone from the 19E1-19E2 region.	This study (Chapter three).
λ798-C10-04A	EMBL 4	Partial Sau3A into *Bam*HI	16.5	Genomic clone from the 19E1-19E2 region.	J.Davies (pers comm) and this study.
λ798-C10-04c	EMBL 4	Partial Sau3A into *Bam*HI	15.7	Genomic clone from the 19E1-19E2 region.	J.Davies (pers comm) and this study.
λ798-011	EMBL 4	Partial Sau3A into *Bam*HI	12.7	Genomic clone from the 19E1-19E2 region.	J.Davies (pers comm) and this study.
λ798-051	EMBL 4	Partial Sau3A into *Bam*HI	17.7	Genomic clone from the 19E1-19E2 region.	J.Davies (pers comm) and this study.
λ798-051-07	EMBL 4	Partial Sau3A into *Bam*HI	16.7	Genomic clone from the 19E1-19E2 region.	J.Davies (pers comm) and this study.
λ051-07-05B	EMBL 4	Partial Sau3A into *Bam*HI	15.7	Genomic clone from the 19E1-19E2 region.	J.Davies (pers comm) and this study.
λ05R308B	EMBL 4	Partial Sau3A into *Bam*HI	12.5	Genomic clone from the 19E1-19E2 region.	J.Davies (pers comm) and this study.
λR03	EMBL 4	Partial Sau3A into *Bam*HI	18.8	Genomic clone from the 19E2-19E3 region.	J.Davies (pers comm) and this study.
λR209	EMBL 4	Partial Sau3A into *Bam*HI	18.3	Genomic clone from the 19E2-19E3 region.	J.Davies (pers comm) and this study.
λR306	EMBL 4	Partial Sau3A into *Bam*HI	16.3	Genomic clone from the 19E2-19E3 region.	J.Davies (pers comm) and this study.
λRC56	EMBL 4	Partial Sau3A into *Bam*HI	17.2	Genomic clone from the 19E2-19E3 region.	J.Davies (pers comm) and this study.
λ3	EMBL 4	Partial Sau3A into *Bam*HI	19.3	Genomic clone from the 19E4-19E5 region.	J.Davies (pers comm) and this study.
λ896-C21	EMBL 4	Partial Sau3A into *Bam*HI	16.8	Genomic clone from the 19E4-19E5 region.	J.Davies (pers comm) and this study.
λM281	EMBL 3	Partial Sau3A into *Bam*HI	19.5	Genomic clone from the 19E4-19E5 region.	J.Davies (pers comm) and this study.
λM283	EMBL 3	Partial Sau3A into *Bam*HI	15.3	Genomic clone from the 19E4-19E5 region.	J.Davies (pers comm) and this study.
λH382	EMBL 3	Partial Sau3A into *Bam*HI	19.4	Genomic clone from the 19E4-19E5 region.	J.Davies (pers comm) and this study.
λH683	EMBL 3	Partial Sau3A into *Bam*HI	17.9	Genomic clone from the 19E4-19E5 region.	J.Davies (pers comm) and this study.

Table 2.4) Bacteriophage clones used in this study. (Cont'd)

Clone	Vector	Cloned as:	Size(kb)	Description	Reference
λAGO08	GEM 11	Partial Sau3A into Sac II	16.8	Genomic clone from the 19E2-19E4 region.	This study (Chapter three).
λAGO02	EMBL 4	Partial Sau3A into *Bam*HI	18.5	Genomic clone from the 19E2-19E4 region	This study (Chapter three).
λAGO01	EMBL 4	Partial Sau3A into *Bam*HI	17.0	Genomic clone from the 19E2-19E4 region.	This study (Chapter three).
λC5B03	EMBL 4	Partial Sau3A into *Bam*HI	11.0	Genomic clone from the 19E2-19E4 region.	J.Davies (pers comm) and this study.
λ952-C5B	EMBL 4	Partial Sau3A into *Bam*HI	14.0	Genomic clone from the 19E2-19E4 region.	J.Davies (pers comm) and this study.
λ952-O8C9	EMBL 4	Partial Sau3A into *Bam*HI	15.7	Genomic clone from the 19E2-19E4 region.	J.Davies (pers comm) and this study.
λ952-08	EMBL 4	Partial Sau3A into *Bam*HI	15.7	Genomic clone from the 19E2-19E4 region.	J.Davies (pers comm) and this study.
λ952-O4C9	EMBL 4	Partial Sau3A into *Bam*HI	17.5	Genomic clone from the 19E2-19E4 region.	J.Davies (pers comm) and this study.
λ94O5	EMBL 4	Partial Sau3A into *Bam*HI	16.8	Genomic clone from the 19E2-19E4 region.	J.Davies (pers comm) and this study.
λ94C11	EMBL 4	Partial Sau3A into *Bam*HI	15.7	Genomic clone from the 19E2-19E4 region.	This study (Chapter three).
λ94C15	EMBL 4	Partial Sau3A into *Bam*HI	14.5	Genomic clone from the 19E2-19E4 region.	J.Davies (pers comm) and this study.
λM961	EMBL 3	Partial Sau3A into *Bam*HI	15.3	Genomic clone from the 19E2-19E4 region.	J.Davies (pers comm) and this study.
λH962	EMBL 3	Partial Sau3A into *Bam*HI	16.0	Genomic clone from the 19E2-19E4 region.	J.Davies (pers comm) and this study.
λH973	EMBL 3	Partial Sau3A into *Bam*HI	19.8	Genomic clone from the 19E2-19E4 region.	J.Davies (pers comm) and this study.
λH986	EMBL 3	Partial Sau3A into *Bam*HI	18.4	Genomic clone from the 19E2-19E4 region.	J.Davies (pers comm) and this study.
λ9915	EMBL 3	Partial Sau3A into *Bam*HI	17.5	Genomic clone from the 19E2-19E4 region.	J.Davies (pers comm) and this study.
λS2	EMBL 3	Partial Sau3A into *Bam*HI	18.4	Genomic clone from the 19E2-19E4 region.	This study (Chapter three).
λS4	EMBL 3	Partial Sau3A into *Bam*HI	17.5	Genomic clone from the 19E2-19E4 region.	This study (Chapter three).
λS5	EMBL 3	Partial Sau3A into *Bam*HI	18.4	Genomic clone from the 19E2-19E4 region.	This study (Chapter three).
λFD3	EMBL 3	Partial Sau3A into *Bam*HI	17.5	Genomic clone from the 19E2-19E4 region.	J.Davies (pers comm) and this study.

NZCYM Agarose Overlay: As for NZCYM Agar with the addition of 2.5g $MgSO_4.6H_2O$ before solidification with 6.5g Agarose (type 1 low EEO A6013)

2.3.3 Selection Supplements.

The following supplements were added to rich media to select and identify recombinant organisms.

(a) Antibiotics. Ampicillin (Amp) was used at a final concentration of 100μg ml^{-1}. A stock solution of 100mg ml^{-1} was filter sterilised and stored at +4°C. Tetracycline (Tet) was used at a final concentration of 12.5μg ml^{-1}. A stock solution of 12.5mg ml^{-1} in ethanol was prepared and stored at -20°C. Hot L-agar was cooled to 55°C before antibiotics were added. L-Amp plates were stable for at least 4 weeks if stored at 4°C.

(b) Chromogenic Substrates. X-gal and IPTG were used at a final concentration of 20mg ml^{-1}. A stock solution of IPTG of 20mg ml^{-1} was filter sterilised and stored at -20°C. A stock solution of X-gal (20mg ml^{-1}) was made up in DMF and stored at -20°C.

2.3.4 Growth of Bacteria.

Liquid cultures for transformation or plasmid or phage DNA preparations were routinely grown in L broth at 37°C with vigorous shaking. Plate cultures were grown at 37°C on L-agar with antibiotics added as required. 20ml of media in a 40ml boiling tube was inoculated from a loop of a single colony from a plate or from 10μl of a glycerol stock. 100ml of media in a 250ml conical flask was inoculated with 500μl of an overnight 10ml L-broth culture. When plating bacteriophage, phage particles were mixed with plating cells and incubated at 37°C for 20 minutes to allow the phage to adsorb to the bacteria. This suspension was then added to 7ml of cooled NZCYM agarose overlay then poured onto hardened NZCYM agar plates. After setting the plates were inverted and incubated at 37°C.

2.3.5 Measurement of Growth.

Bacterial cell density was measured as an apparent absorption at 600nm. The measurements were carried out in 1 cm light path cuvettes using a Beckman DU-50 series spectrophotometer (Beckman Instruments Ltd., U.K.) equipped with a digital read out. If necessary, the culture samples were diluted 1:10 to give an absorbance of less than 0.5.

2.3.6 Harvesting of Bacteria

Cells were harvested by centrifugation at 8,000 rpm for 15 minutes at 4°C in a Beckman JA 14 (6x250ml rotor).

2.3.7 Storage of Bacterial Strains.

Bacterial strains were stored on L-agar slopes at room temperature, or in 50% LB broth, 40% glycerol at -20°C.

2.4 Fly Husbandry.

All fly stocks were maintained and grown at either 18°C or 25°C, in vials or if large numbers of flies were required in plastic half pint bottles. New stocks were placed in quarantine for one month to ensure they were disease and mite free. All strains were tested against lethal loci to ensure the mutation had been maintained before use and at approximately six monthly intervals.

2.4.1 *Drosophila melanogaster* Growth Media.

Normal Fly Food (per three litres)	Agar	30g
	Sucrose	45g
	Glucose	90g
	Dried Yeast	105g
	Maize Meal	45g
	Wheat Germ	30g
	Treacle	90g
	Soya Flour	10g
	Water	3 litres

The constituents were mixed together and stirred until boiling, simmered for 15 minutes and then allowed to cool. After cooling, 30ml 10% Nipogen and 5ml propionic acid were added and then the food was poured 2-3 cm deep into plastic half pint bottles. These were then stored at 4°C for up to four weeks.

Rich Larval media (per three litres)	Glucose	300g
	Dried Yeast	300g
	Agar	60g
	Water	2 litres

The constituents were mixed together and boiled for 5 minutes before adding the remaining 1000ml of water. After cooling 30ml 10% Nipagen and 5ml propionic acid were added and then the food was poured 2 cm deep into plastic Tupperware boxes. These were then stored at 4°C for up to eight weeks.

Grape Juice Agar (per five litres) (for embryo collections)	Agar	99g
	Glucose	261g
	Sucrose	130g
	Dried Yeast	35g
	Grape juice	444ml
	Water	4 litres
	1.25M NaOH	118ml
	Nipogen	50ml

The constituents were mixed together and heated until boiling and then allowed to simmer for 30 minutes. The mixture was then allowed to cool before the addition of the Nipogen and then poured into petri dishes of the appropriate size to a depth of 0.5 - 1.0cm. It was then stored at 4°C for up to eight weeks.

2.4.2 Fly Strains.

The wild-type fly strains used in the course of this study were *Oregon-R* (*OrR*); *Canton-S* (*CS*) and *Sierra Leone* (*SL*), (Lindsley and Zimm, 1992). See Tables 2.5, 2.6, 2.7 and 2.8 for mutant strains. The *CS*, *OrR* and the mutant strains have been maintained at Glasgow for six years whereas the *Sierra Leone* strain has been maintained here for eighteen months (K.O'Dell, pers. comm.).

2.4.3 Tissue Collection.

The tissue collected for nucleic acid preparations needed to be staged *i.e.* of a known age. Before collections were started all fly strains were checked by crossing to neighbouring lethal complementation groups or to deficiencies and duplications. F1 progeny were collected. For deficiency stocks heterozygous females were collected, for duplication stocks male flies were collected (See Section 3.3.1 for explanation).

2.4.3a Embryo Collections.

Embryos were collected on grape juice agar plates at 25°C at three hourly intervals. They were then aged at 25°C for an appropriate period. To ensure an even representation of different stages across embryogenesis, embryos were collected and aged to the following times, 0-3 hours, 3-6 hours, 6-9 hours, 9-12 hours, 12-15 hours, 15-18 hours and 18-22 hours. Equivalent amounts of tissue were then added to each other to ensure no one time point was in excess. The embryos were isolated by brushing them off the grape juice agar (Section 2.4.1) into a fine meshed sieve. After

Table 2.5 Chromosome deficiencies used in the course of this study

Fly Genotype	Origin	Deficient Region	References
Df(1)A118 / In(1)FM6,y^{31d}sc^{8}dmB / y+Ymal^{106}	X-rays	*R-9-28-vao*	3, 4, 5, 6
Df(1)LB6 / In(1)FM6,y^{31d}sc^{8}dmB / y+Ymal^{106}	mitomycin-C	*shak-B - eo*	3, 4, 5, 8, 9
Df(1)16-3-35 / In(1)FM6,y^{31d}sc^{8}dmB /y+Ymal^{106}	Neutrons	*mal - shak-B*	3, 5, 6, 9, 10
Df(1)17-351 / In(1)FM6,y^{31d}sc^{8}dmB / y+Ymal^{106}	Neutrons	*R-9-28 - LB20*	5, 6, 9
Df(1)HC279 / In(1)FM6,y^{31d}sc^{8}dmB / y+Ymal^{106}	X-rays	*R-9-28-vao*	Perrimon *et al.*, 1989b
Df(1)26B / In(1)FM6,y^{31d}sc^{8}dmB / y+Ymal^{106}	Hybrid dysgenesis	*R-9-28-eo*	7, 9
Df(1)T2-14A / In(1)FM6,y^{31d}sc^{8}dmB /y+Ymal^{106}	Tritiated deoxycytidine	*EC235 - vao*	3, 4, 5, 6, 8, 9
Df(1)N77 / In(1)FM6,y^{31d}sc^{8}dmB	X-rays	*mell - su(f)*	Perrimon *et al.*, 1989b
Df(1)LB7 / In(1)FM6,y^{31d}sc^{8}dmB / y+Ymal^{106}	mitomycin-C	*R-9-28 - su(f)*	Perrimon *et al.*, 1989b
Df(1)B57 / In(1)FM6,y^{31d}sc^{8}dmB / y+Ymal^{106}	X-rays	*leg - unc*	3, 4, 5, 6
Df(1)GA40 / In(1) FM7,y^{31d}sc^{8}w^{a}B	X-rays	*leg - su(f)*	1, 9
Df(1)GA37 / In(1) FM7,y^{31d}sc^{8}w^{a}B	X-rays	*leg - LB20*	1, 9
Df(1)mal^{10}: In(1)dl49,sn^{x2}v^{of}mal^{2} / In(1)sc^{8},Df(1)mal^{10}sc^{8}B / y+Ymal^{106}	X-rays	*mal - mell*	3, 4, 5, 8, 9

Table 2.6 Chromosome duplications used in the course of this study

Fly Genotype	Origin	Duplicated Region	References
Dp(1;f)mini-ring: Dp(1)y^2*su(*w^a*)*w^aX^Y / C(1)DX,*ywf* / y^+Dp ring	Mutation-recombination induced rearrangement	*shak-B - su(f)*	Green *et al.*, 1987
Dp(1;f)mini-2: C(1) DX, *ywf* / X^Y, *ywf* / y^+mm2	Mutation-recombination induced rearrangement	*R-9-28 - su(f)*	Green *et al.*, 1987
Dp(1;Y)mal^{171}: Dp(1) mal^{171}, w^a *f su(f)* / y^+Ymal^{171} / w^a *f su(f)*	X-rays	*R-9-28 - su(f)*	5, 11, 12

Table 2.7 *Shak-B* alleles used in the course of this study

Fly Genotype	Origin	Lethal or Neural	References
wR-9-29 / In(1)FM6,$y^{31d}$$sc^8$*dmB* / y+Ymal^{106}	EMS	Lethal / Neural	Lifschytz and Falk, 1969
wR-9-29 / In(1)FM6,$y^{31d}$$sc^8$*dmB;bw;st*	EMS	Lethal / Neural	Lifschytz and Falk, 1969
shakB^2;bw;st	EMS	Neural	Homyk *et al.*, 1980
Pas / In(1)FM6,$y^{31d}$$sc^8$*dmB;bw;st*	EMS	Neural	Thomas andWyman, 1984a
EC201 / In(1)FM6,$y^{31d}$$sc^8$*dmB;bw;st*	EMS	Lethal / Neural	Lindsley and Zimm, 1990
y,f,E81 / In(1) FM7,$y^{31d}$$sc^8$,$w^a$,*B*	EMS	Lethal / Neural	Lifschytz and Falk, 1969
17-360 / In(1)scS1Lsc^8R+dl49,$sc^8$$sc$S1	Neutrons	Lethal / Neural	Lindsley and Zimm, 1990
yw^{a}HM437 / In(1)FM6,$y^{31d}$$sc^8$*dmB* / y^+Ymal^{106}	HMS	Lethal / Neural	Kramers *et al.*, 1983
EF535 / In(1)FM6,$y^{31d}$$sc^8$*dmB* /y+Ymal^{106}	EMS	Lethal	Lindsley and Zimm, 1990

Table 2.8 Other alleles used in the course of this study

Fly Genotype	Origin	Comments	References
wR-9-28 / In(1)FM6,$y^{31d}sc^{8}dmB$ / *y*+Ymal^{106}	EMS induced	Lethal allele from the *R-9-28* locus	2, 3, 5
ywrun / In(1)FM6,$y^{31d}sc^{8}dmB$ / *y*+Ymal^{106}	EMS induced	Lethal allele from the *runt* locus	Lindsley and Zimm, 1990

References:1, Perrimon *et al.,* 1989; 2, Lifschytz and Falk, 1969; 3, Schalet and Lefevre, 1976; 4, Schalet and Lefevre, 1973; 5, Lindsley and Zimm, 1986; 6, Miklos *et al.,* 1987; 7, Zusman, Coulter and Gergen, 1985; 8, Lindsley and Zimm, 1985; 9, Lindsley and Zimm, 1987; 10, Yamamoto and Miklos, 1987; 11, Gergen and Wieschaus, 1986; 12, Lifschytz and Yakobovitz, 1978.

each collection the tissue was either snap frozen in liquid nitrogen and stored at -70°C or homogenised by a polytron homogeniser in RNA isolation buffer (Section 2.11) until the solution was required for RNA isolation.

2.4.3b Larval Collections.

L1 larvae were isolated by performing a 12 hour collection and then allowing the embryos to age 24 hours at 25°C, thus the embryos when frozen were at mid-L1 stage +/- 12 hours. No L2 larvae were collected. L3 larvae were collected in a similar way to the L1s so the L3 tissue was at mid L3 +/- 12 hours.

2.4.3c Pupal Collections.

Pupae were collected as 3 batches completely covering pupation (early, mid and late pupae). Pupae were first collected as prepupae (the first four hours of pupation (Ashburner, 1989) by placing them in water and collecting those that sank. These were then aged a further 24 hours and then placed in water again. Those that sank were discarded and those that floated were collected and were therefore known to be between 6 and 30 hours into pupation. This procedure was repeated and the pupae were aged for a further 24 and 48 hours to give pupae aged between 24-48 and 48-72 hours into pupation. These were then combined with equivalent amounts of tissue isolated from the other two time periods. This was to ensure RNA isolated from any one time point was not in excess so RNA from the whole period of pupation would be represented in approximately equal amounts.

2.4.3d Adult Collections.

Adults collected were typically 1-5 days old. Head and body tissue was collected essentially as described by Ashburner (1989). Adults were vortexed in liquid nitrogen for one minute, to severe the neck and the head and body were then separated from each other and from wings, legs *etc.*, by sieving through brass sieves precooled in liquid nitrogen.

2.5 Recombinant DNA Techniques.

2.5.1 Small Scale Plasmid Purification.

The method employed for small scale plasmid DNA preparation was based on the method of Birnboim and Doly (1979) . 5ml of L-broth, containing the appropriate antibiotic, were inoculated with a single colony of plasmid transformed cells and grown overnight. 1.5ml of the overnight culture was harvested in a microfuge and the cell pellet resuspended in 100µl Birnboim Doly Buffer I (Section 2.2.8.5). 10µl of 10mg ml^{-1} lysozyme was then added and incubated at room temperature for 5 minutes. 200µl freshly made Birnboim Doly Buffer II (Section 2.2.8.5) was added, mixed gently and incubated on ice for 5 minutes. Cells were lysed by the addition of 150µl ice cold Birnboim Doly Buffer III (Section 2.2.8.5). The lysate was incubated in ice for 5-15 minutes and the cell debris removed by centrifugation in a microfuge. The supernatant was then removed into a fresh microfuge tube and the plasmid DNA was phenol / chloroform extracted (Section 2.5.4). The supernatant was again put into a fresh tube and DNA precipitated by the addition of 2 volumes of ethanol at room temperature After incubation at room temperature for 2 minutes, the plasmid DNA was recovered by centrifugation in a microfuge for 5 minutes The pellet was washed with 200µl 70% (v/v) ethanol, air dried for 5-15 minutes and resuspended in 30µl TE. A 5µl sample was sufficient for digestion by restriction enzymes. During the digestion, 1µl boiled RNase A (10mg ml^{-1}) was added to the digest for the last 15 minutes of the incubation.

2.5.2 Large Scale Plasmid Purification.

The method employed was based on the alkali lysis method of Birnboim and Doly (1979) , and was used to purify milligram quantities of plasmid DNA. 100ml of L-broth, containing the appropriate antibiotic, was inoculated with a single colony of plasmid transformed cells. This culture was incubated at 37°C on an orbital shaker overnight (16 hours). The bacterial cells were harvested by centrifugation for 10 minutes at 9,000 rpm, 4°C (Beckman JA 14, 6x250ml rotor). The cell pellet was resuspended in 5ml Birnboim Doly Buffer I (Section 2.2.8.5) containing lysozyme at a concentration of 5mg ml^{-1} and the suspension was incubated at room temperature for 5 minutes. 10ml of Birnboim Doly Buffer II (Section 2.2.8.5) was added, gently mixed by swirling and then incubated on ice for 10 minutes. Cells were lysed by the addition of 7.5ml of ice cold Birnboim Doly Buffer III (Section 2.2.8.5), followed by

rocking to mix and incubation on ice for 10 minutes. Bacterial cell debris was removed by centrifugation (10,000 rpm, 15 minutes, Beckman JA 14, 6x250ml rotor). The supernatant was filtered through gauze to remove any floating cell debris. The nucleic acids were precipitated by the addition of 0.6 volumes of isopropanol. After incubation at room temperature for 15 minutes, the nucleic acid was recovered by centrifugation (Beckman JA 20, 12x50 ml rotor, 10,000 rpm, 10 minutes, 25°C). The pellet was washed with 70% (v/v) ethanol, air dried for 10-20 minutes and resuspended in 5ml TE. 0.5ml of 10mg ml^{-1} ethidium bromide was then added together with CsCl to a final concentration of 1g ml^{-1}. The gradient was then set up by centrifugation at 55,000 rpm for 16 hours at 25°C in a Beckman Ti70.1 rotor in polycarbonate tubes. The DNA bands were usually visible. If the bands were not visible, long wavelength U.V. light was used to visualise the plasmid band which migrates in front of any contaminating host chromosomal DNA. The plasmid band was removed by gently inserting a sterile 1ml hypodermic syringe fitted with a 21G11/2 gauge needle just below the plasmid band. Ethidium bromide was removed by extracting several times with propan-2-ol saturated with CsCl. The DNA was then diluted to double the volume with distilled water and dialysed three times against 5 litres of TE. The DNA concentration and purity (A260/A280) was then measured spectrophotometrically (Section 2.2.4).

2.5.3 Digestion of DNA with Restriction Enzymes.

The methods used were as described by Sambrook *et al.*, (1989). Restriction digests were carried out using the BRL React buffers which were provided with each batch of enzyme. There are ten different React buffers with a range of salt concentrations, each one suitable for a range of enzymes. Analytical digests were carried out in a volume of 20µl or 30µl at 37°C. Preparative digests were carried out in larger volumes. When DNA was digested with two restriction enzymes, the endonuclease requiring the lower salt buffer was used first. After the recommended duration of digestion the salt concentration was adjusted and the second enzyme added.

2.5.4 Phenol / Chloroform Extraction of Nucleic Acids.

Restriction digests and other solutions containing DNA were deproteinised by extracting with phenol and chloroform. The volume of the sample was measured and an equal volume of TE-saturated phenol:chloroform:isoamyl alcohol (25:24:1) was

added. This mixture was vortexed, allowed to stand until the two layers separated and then vortexed again. The two layers were then separated by centrifugation in a microfuge for 5 minutes. The aqueous phase was removed into a fresh tube and an equal volume of TE added to the bottom, phenol layer. The mixture was vortexed and the two layers separated as in the first extraction. The two aqueous, top, layers were pooled and extracted in an identical manner with an equal volume of TE-saturated chloroform / isoamyl alcohol (24:1) which removes any residual traces of phenol. The DNA was recovered from the aqueous layer by ethanol precipitation (Section 2.5.5).

2.5.5 Concentration of Nucleic Acids by Ethanol and Isopropanol Precipitation.

Sodium acetate was added to the DNA solution to a final concentration of 0.3M by adding 0.1 volumes of a 3.0M solution. 2.5 volumes of cold ethanol (-20°C) were added and mixed. The mixture was then incubated at -20°C overnight or for 1 hour and the precipitated DNA recovered by centrifugation in a microfuge at 4°C for 30 minutes or for larger volumes, centrifugation at 10,000 rpm for 30 minutes at 4°C (Beckman JA20 or JA17 rotor). The DNA pellet was washed in 70% (v/v) ethanol and resuspended in a suitable volume of TE. An alternative to the addition of sodium acetate was to add 0.5 volumes of 7.5M ammonium acetate, however, when the DNA was to be treated with T4 polynucleotide kinase, sodium acetate was used since ammonium ions inhibit this enzyme. Precipitation with isopropanol was achieved by adding an equal volume of isopropanol and then treating the mixture the same as for ethanol precipitations.

2.5.6 Ligations.

The insert DNA was digested with the appropriate restriction enzymes, isolated as described (Sections 2.5.3 and 2.5.10) and ethanol precipitated (Section 2.5.5). Plasmid vector (normally Bluescript KS+) was linearised by digestion with the appropriate restriction enzymes and treated with bacterial alkaline phosphatase to yield DNA ready to use. This treatment prevented self-ligation of the vector and facilitated a low-background of false positive clones. For ligation reactions where the insert size was similar to the vector size an equal concentration of foreign DNA to vector DNA was employed. For those reactions where the vector size was much greater than the insert size, at least a four-fold excess in concentration of foreign DNA to vector DNA was

typical. Depending on the size of insert, ligation mixtures contained 25ng vector and between 25 to 400ng insert in a final volume between 5 to 10µl. Vector plus insert were pre-incubated at 45°C for 5 minutes to melt any cohesive termini that had reannealed. The ligations were performed overnight in 66mM-Tris HCl;pH 7.6, 6.6mM MgCl2, 0.5mM ATP, 10mM DTT at 16°C using 0.5 Units of bacteriophage T4 DNA ligase.

2.5.7 Transformation of *E.coli* with Plasmid DNA.

2.5.7a Preparation of Competent Cells.

A single colony of *E.coli* DS941 or *E.coli* XL-1 blue from a stock minimal media plate was used to inoculate 10ml of 2 x YT which was incubated overnight at 37°C. 0.3ml of overnight culture was used to inoculate 30ml of 2 x YT which was grown at 37°C for around 2 hours to an OD at 600nm of 0.4 - 0.6. The cells were cooled on ice for 10 minutes and gently harvested (Jouan, 3,000 rpm, 10 minutes). When being prepared by the $CaCl_2$ method (Cohen *et al.*, 1972) the cells were resuspended in one half of the original culture volume of ice-cold, sterile 50mM $CaCl_2$, 10mM-Tris HCl;pH 8.0 and placed in an ice bath for 15 minutes. Following centrifugation (Jouan, 2,000 rpm, 15 minutes) the cells were resuspended in 1/15th original culture volume of 50mM $CaCl_2$, 10mM Tris HCl;pH 8 and mixed in the ratio 3:1 culture to glycerol. These cells were dispensed into chilled microfuge tubes in 200-400µl aliquots and stored at -70°C.

2.5.7b Transformation of Competent Cells.

Transformations were carried out in sterile 1.5ml microfuge tubes. An aliquot of ligation mix (any DNA being used in the transforming process) containing up to 25ng plasmid DNA was added to 100µl aliquots of competent cells and the mixture was incubated on ice for at least 30 minutes. The DNA / cell mix was then heat shocked at 42°C for 2 minutes. 1ml of L-broth was added to the tubes and they were incubated without shaking at 37°C for 1 hour. The cells were then plated onto appropriately prepared antibiotic / chromogenic containing LB plates and incubated overnight at 37°C.

2.5.7c Selection of pUC Derived Recombinant Clones.

The Bluescript and the pUC plasmids (from which the Bluescript vectors have been designed) were constructed as cloning vectors using β–galactosidase activity as the basis of selection (Messing, 1983 ; Norrander *et al.,* 1983; Yanisch-Perron *et al.,* 1985). The vector carries a segment of *E.coli* DNA that contains the regulatory sequences and the coding information for the first 146 amino-acids of the β–galactosidase gene (*lacZ*). Vectors of this type are used in host cells (*e.g. E.coli*, XL1-Blue) that code for the carboxy-terminal portion of β–galactosidase, therefore though neither the host-encoded nor the plasmid-encoded fragments are themselves active, they can associate to form an active enzyme. A polycloning site has been inserted, in-frame, within the coding region of the plasmid which does not affect the complementation, however insertion of additional DNA into the polycloning site generally destroys the complementation. Active β–galactosidase cleaves the chromogenic substrate X-Gal, to produce a blue chromophore, when transformed cells are grown in the presence of the non-metabolisable *lac* operon inducer, IPTG. However in recombinant plasmids the ability for complementation is lost, the enzyme is inactive and consequently the colonies appear white. False positive white colonies appear at low frequency, probably due to incorrect self-ligation of the vector (Yanisch-Perron *et al.,* 1985).

2.5.8 Electrophoresis through Agarose Gels.

2.5.8.1 Electrophoresis of DNA.

DNA was separated at room temperature on horizontal submerged agarose gels as described by Sambrook *et al.,* (1989). 0.5% to 1.5% agarose gels were used throughout this study and were used to size restriction fragments accurately in the range of 0.8 - 10kb. Nucleic acids were visualised by ultraviolet fluorescence after staining with EtBr on a long wave U.V. transilluminator (U.V. Products Inc.). For DNA work the EtBr was either added at a concentration of 0.2μg ml^{-1} to the agarose prior to the pouring of the gel or, after running the gel was soaked in 1x running buffer plus 0.6μg ml^{-1} EtBr. Agarose powder was dissolved at 100°C in 1x running buffer (1 x TBE or 1 x TAE) This was then cooled to 50°C before it was poured into a horizontal gel former using an appropriate Teflon well former, or comb. After setting the gel was removed to a gel running tank and submerged under running buffer. Samples for agarose gels were prepared by addition of 0.2 volumes of loading buffer (Section

2.2.8.1). After the loading of samples the gel was run at 25 to 150V with restriction fragment markers of known size run alongside the unknown fragments. This was usually the 1kb ladder (BRL Gibco Ltd., Paisley, U.K.), or phage DNA restricted with *Hin* dIII or with *EcoR* I + *Hin* dIII.

2.5.8.2 Formaldehyde / Agarose Gel Electrophoresis of RNA.

This was performed as described by Sambrooke *et al.,* (1989), with modifications as described by Rosen and Villia-Komaroff (1991). RNA stored at -20°C as an aqueous solution was added to 15.5µl of MMF solution (MMF=2µl 10x MOPS Buffer, (0.2M MOPS (3-[N-Morpholino]propanesulfonic acid), 0.05M sodium acetate, 0.01M EDTA(pH8.0), 10µl Formamide, 3.5µl 40% (w/v) Formaldehyde). 1µl of 400µg ml^{-1} EtBr was then added to the samples which were heated to 65°C for 10 minutes before being snap cooled on ice. After adding 3µl of gel loading buffer the samples were ready for loading. Formaldehyde gels were prepared by dissolving 5g of agarose in 385ml of water. To this was then added 25ml of 20X MOPS and after the solution had cooled to 50°C, 90ml of Formaldehyde solution. The solution was then poured into a horizontal gel former using an appropriate Teflon well former, or comb and allowed to set for 45 minutes. After setting the gel was covered in 1X Running Buffer (20mM MOPS, 5mM sodium acetate and 1mM EDTA;pH8.0, the comb removed and the samples loaded. The gel was run at 50V for 12 - 16 hours, the running buffer being circulated through the system by the means of a peristaltic pump.

2.5.8.3 Alkaline Gel Electrophoresis.

This was carried out as described by Sambrook *et al.,* 1989 and was used to analyse i) the size of radiolabelled first strand cDNA, ii) the size of random primed probes, iii) the results of the calibration of the reagents (especially *DNase* 1) for Nick Translation reactions (Section 2.5.13.1). A 1.4% agarose gel was prepared by dissolving 1.4g of agarose powder in 100ml boiling 50mM NaCl, 1mM EDTA;pH7.0. After cooling to 45-55°C the agarose was poured into a horizontal gel former using an appropriate Teflon well former, or comb and allowed to set. The gel was then placed in an electrophoresis tank and covered with alkaline running buffer to a depth of 5-10mm and allowed to soak for >30 minutes. The DNA samples were treated just prior to loading by heating to 75°C for 2 minutes with 50% Formamide added to stop reannealing and 2ml alkaline loading buffer (Section 2.2.8.1). Samples were

electrophoresed for 13-17 hours at 25V and sized against end labelled (Section 2.5.13.3) 1kb ladder (BRL Gibco Ltd., Paisley, U.K.). After the gel had run it was dried down by sandwiching it between 20-30 sheets of 3MM Whatman paper with a large weight on top and then sealing the dried down gel into a plastic hybridisation bag for visualisation of the labelled DNA by direct autoradiography at -70°C.

2.5.9 Photography of Electrophoresis Gels.

After gels had run they were photographed with UV transillumination (short wavelength 240nm, long wavelength 340nm) using a Polaroid camera loaded with Polaroid 4X5' Land Film (N° 57) fitted with a Kodak Wratten Filter N° 23A (red). For RNA work the EtBr was added to the samples prior to heat denaturation of the samples as described by Rosen and Villia-Komaroff, 1991.

2.5.10 Recovery of DNA from Agarose Gels.

The DNA band of interest was visualised by ethidium bromide staining and a long wavelength ultra-violet lamp. The desired DNA band was excised from the gel. A hole was pierced in the base of a sterile 0.5ml eppendorf tube using a 26G3/8 sterile syringe needle and then plugged with a small amount of sterile, siliconised glass wool. The agarose gel chip containing DNA was placed into the 0.5ml eppendorf tube which in turn was placed inside a 1.5ml eppendorf tube and microfuged for 10 minutes at 12,000 rpm. The 1.5ml eppendorf was then removed and the buffer, containing DNA, retained. The 0.5ml eppendorf tube was microfuged once more collecting any remaining DNA in a fresh 1.5ml eppendorf. The two DNA containing fractions were pooled and microfuged at 12,000 rpm, at 4°C for 30 minutes to remove any traces of agarose. The supernatant was decanted and DNA recovered by ethanol precipitation. DNA obtained in this way was sufficiently pure for cloning purposes.

For radiolabelling of DNA fragments the above method of recovery was found to be unsatisfactory presumably as residual agarose interfered with the Klenow enzyme, so electroelution into dialysis bags was employed. The DNA was excised from the agarose gel as described above. This was then put into a small piece of dialysis bag with 600µl of 1X TBE. The ends of the bag were sealed with dialysis clips and any air bubbles removed. The DNA was then allowed to migrate out of the chip by electrophoresis (typically at 100 volts for 3 hours) onto the side of the dialysis bag.

The DNA was recovered by reversing the current for 30 seconds at 100V to loosen it from the side of the bag. The agarose chip was then removed and the TBE placed in an eppendorf tube. The sides of the bag were washed clean of DNA with 200µl of TBE to maximise recovery. Residual agarose was removed by two centrifugations at 12,000g in a microcentrifuge at room temperature for 15 minutes each. The supernatant removed each time to a fresh eppendorf tube. The DNA was then recovered by ethanol or isopropanol precipitation (Section 2.5.5).

2.5.11 DNA Transfer to Nitrocellulose Membranes (Southern Blotting).

This method is based on that of Southern (1975). After electrophoresis the agarose gel was placed in 250ml of 0.25M HCl and soaked for 20 minutes. The gel was then placed in 250ml of denaturing solution (1.5M NaCl, 0.5M NaOH) for 30 minutes and then into neutralising solution (1M Tris, 1.5M NaCl) for 60 minutes. The DNA was blotted onto a nitrocellulose (Hybond-C EXTRA, Amersham) membrane using 20X SSC transfer buffer (Section 2.2.8.5) overnight. After blotting the membrane was washed in 2X SSC as described in detail in Sambrook *et al.,* (1989). The filters were air dried and then baked at 80°C for 2 hours to cross link the DNA to the membrane.

2.5.12 RNA Transfer to Nitrocellulose Membranes (Northern Blotting).

The method used was based on that described by Thomas (1980), with modifications as described by Sambrooke *et al.,* (1989). After electrophoresis and photography (Section 2.5.8.2 and 2.5.9), the formaldehyde gel was soaked for 20 minutes in 20X SSC before the gel was inverted and capillary blotted to nitrocellulose using 20X SSC as the transfer buffer as for Southern blotting. No pre-treatment of the gel was necessary as the formaldehyde denatured the RNA prior to transfer.

2.5.13 Radiolabelling of DNA And RNA.

2.5.13.1 ^{32}P Labelling of DNA by Nick Translation.

This was carried out using the procedure of Rigby *et al.,* (1977). The nick translation reaction was set up as 5µl 10x Nick translation buffer, (500mM Tris-HCl;pH7.2, 100mM $MgSO_4$, 1mM DTT, 500µg ml^{-1} BSA [ultra pure], 0.1-1.0µg of DNA, 1µl β–mercaptoethanol, 10µl nucleotide mix [500µM each of the dNTPs not being used as the radiolabel], 30-50µCi α^{32}P dCTP [650-800 Ci mM^{-1}], 2µl 100ng μl^{-1} *DNase* 1, 5 units of *DNA polymerase* 1 (*DNase* 1 was calibrated prior to use as described by

Rigby *et al.*, (1977) and Sambrook *et al.*, (1989). The mix was then incubated for 1-2 hours at 16°C and then heated to 65°C for 10 minutes to denature the enzymes. The unincorporated nucleotides were then removed by either G_{50} Sepharose column chromatography or by selective precipitation with spermine (Section 2.5.14).

2.5.13.2 ^{32}P Labelling of DNA by Random Priming.

DNA was labelled to very high specific activity using the random priming technique of Feinberg and Vogelstein (1983), with modifications as described by Hodgson and Fisk (1987). Random sequence hexanucleotides were used to prime DNA synthesis on a single stranded DNA template by the large fragment of *DNA polymerase* 1 (Klenow Fragment). Random priming reactions were typically carried out in a 50µl volume. 10-50ng of linearised DNA was denatured at 99°C for 7 minutes in the presence of random hexamers (Gibco, BRL Ltd., Paisley, U.K.) at a concentration of 750µg ml^{-1}. This mixture was snap cooled on ice and the following added sequentially: 5µl of 2M Hepes; pH7.6, 2µl 1mg ml^{-1} Acetylated BSA, 10X Klenow reaction buffer to which had been added a 1mM solution of the 3 unlabelled dNTPs and 1µl (5 units) of Klenow. To this mixture 30µCi ^{32}P dNTP 800Ci mM^{1} (typically dCTP) was added. The reaction was left for at least one hour at 37°C, or more normally overnight at room temperature. The unincorporated nucleotides were then removed by either G_{50} Sepharose column chromatography or by selective precipitation with spermine (Section 2.5.14). Typically 60-80% of the radiolabelled nucleotide was incorporated.

2.5.13.3 ^{32}P Labelling of the 5′ Terminus of DNA Fragments.

Bacteriophage T4 polynucleotide kinase can catalyse the transfer to a free hydroxyl group of the g-phosphate group from ATP to the 5′ terminus of DNA. In a total reaction volume of 10µl the mixture contained 1.5 pmoles of 'ends', kinase buffer (50mM Tris-HCl;pH8.0, 10mM $MgCl_2$, 5mM DTT, 1mM spermidine), 8 pmoles γ^{32}P-ATP and 10 units of T4 polynucleotide kinase. The reaction was carried out at 37°C for 45 minutes by which time it had gone to completion. The unincorporated nucleotides were then removed by either G_{50} Sepharose column chromatography or by selective precipitation with spermine (Section 2.5.14).

2.5.13.4 Production and ^{32}P Labelling of First Strand cDNA.

The method used for the generation of radiolabelled first strand cDNA for reverse Northern analysis was based on that described by Okayama & Berg (1982), later modified by Gubler & Hoffman (1983) and by Berger and Kimmel (1987). The aim of the experiment was to produce radiolabelled first strand cDNA to use in identifying transcribed regions within the chromosomal walks (so called reverse Northerns). The reaction was performed using 100µCi α^{32}P-dCTP (800Ci mmol^{-1}) dried down in a vacuum dryer. To this was then added 4µl 5x first strand buffer (250mM Tris-HCl;pH8.3, 250mM KCl, 50mM $MgCl_2$, 2.5mM spermidine), 1µl of oligo $d(T)_{12-18}$ (2mg ml^{-1}), 2µl of a 20mM mix of dATP, dGTP and dTTP, 1µl of Human Placental Ribonuclease Inhibitor (200u µl^{-1}), 2µl of actinomycin D (10mg ml^{-1}), which was added as an inhibitor of hair pin bend formation, 1µg of staged polyA$^+$ RNA had been previously denatured by heating to 65°C for 5 minutes, before being snap cooled and added to the mixture. dH_2O to 19µl was then added and finally 1ml (200 units) of MMLV reverse transcriptase. The reaction was allowed to proceed for 60 minutes at 37°C before 1µl of 10mM dCTP was added as a 'chase' along with another 1ml (200 units) of MMLV reverse transcriptase and the reaction allowed to proceed for a further 60 minutes. The RNA strand was hydrolysed by the addition of 21µl of hydrolysing solution (0.6M NaOH, 20mM EDTA) and left at 65°C for 30 minutes. The radiolabelled cDNA was then separated from unincorporated nucleotides by G_{50} Sepharose column chromatography (Section 2.5.14). The incorporation was calculated using TCA precipitation (Section 2.5.15) and the length of the first strand cDNA checked by NaOH denaturing gel electrophoresis versus radiolabelled 1kb ladder (Section 2.5.8.3).

2.5.14 Removal of Unincorporated Nucleotides.

Unincorporated nucleotides were removed in two ways, either by G_{50} Sepharose column chromatography or more normally by selective precipitation with spermine. The polyvalent cation spermine selectively precipitates double stranded DNA in excess of 100 base pairs in size when in a low to medium salt environment, but will leave nucleotides and primers in solution. To a 50µl probe reaction mix, spermine was added to a final concentration of 10mM (*i.e.* 2.5µl of a 0.2M solution added to the normal 50µl probe reaction mix). This was then incubated on ice for 20-30 minutes

and then centrifuged at 12,000g for 15 minutes at 4°C. The supernatant was then removed and the pellet resuspended in 490µl of probe mix (10mM EDTA, 0.5% SDS) and allowed to resuspend at room temperature for 30 minutes before use.

2.5.15 Determination of Percentage Radiolabel Incorporation.

Two methods were employed. Approximate percentage incorporations were determined by comparing the counts per second (cps) of the labelled DNA to the cps of the separated (Section 2.5.14) unincorporated nucleotides using a hand held Geiger Mueller Tube, held approximately 2.5 feet away from the eppendorf tube containing the radioactivity. Accurate percentage incorporations was calculated by performing a Trichloroacetic acid (TCA) precipitation. 1µl of the labelling reaction was diluted into 99µl of 0.2M EDTA. 3µl of this diluted sample was then dotted onto Whatman GF/C glass microfiber filters in duplicate. The filters were air dried and one of the filters was then washed three times in >10ml of ice cold TCA solution (10% TCA, 1% sodium pyrophosphate) before being rinsed in 95% ethanol and then allowed to air dry. After air drying cps were determined by scintillation counting. It was not necessary to use scintillation fluid for counting ^{32}P-labelled samples as the Cerenkov radiation emitted can be detected by the scintillation counter set to monitor the tritium window. Although the absolute number of cps will not be the same between the two methods (because Cerenkov counting is less than half as efficient), they will be in proportion from sample to sample. Percentage incorporation and thus specific activity was calculated by comparing the cps of the unwashed relative to the washed filter.

2.5.16 Hybridisation of Filter Bound Nucleic Acid.

The temperature of hybridisation and the salt concentration and temperature of washing solutions were dependent on the particular experiment being carried out and precise details are given in the text. In all cases double stranded probes were boiled at 100°C for 10 minutes, snap cooled and then added to the hybridisation solution in the bag.

2.5.17 Aqueous Hybridisations.

For DNA to DNA hybridisations Southern blots were pre-wetted and placed directly into a polythene bag containing pre-hybridisation solution (5x SSC, 0.05% (w/v) sodium pyrophosphate, 5X Denhardts, 0.1% (w/v) SDS, 100µg ml^{-1} sonicated salmon

sperm DNA). The volume of pre-hybridisation fluid was determined by filter surface area x 0.2. The bag was placed in an shaking waterbath and the filter pre-hybridised for at least 3 hours at 65°C. After pre-hybridisation the bag was opened, two thirds of the fluid was removed and denatured labelled DNA fragment (Section 2.5.16) added. The bag was resealed and the hybridisation carried out for >16hr in a shaking waterbath at the appropriate temperature.

2.5.18 50% Formamide Hybridisations.

DNA to RNA hybridisations were carried out at 42°C in 50% formamide hybridisation buffer (50% Formamide, 5xSSC, 0.05% (w/v) sodium pyrophosphate, 5X Denhardts, 0.1% (w/v) SDS). Pre-hybridisation of the filters was carried out for 3 hours and the hybridisation carried out for >16hr in a shaking waterbath at the appropriate temperature.

2.5.19 Washing of Filters.

The hybridised filters were washed in large volumes of buffer of appropriate ionic strength, depending on the experiment. The temperature of this buffer was 65°C. After washing the filters were autoradiographed damp (under Saran wrap) as described in Section 2.2.6.

2.6 Screening of Phage Clones by Colony Hybridisation.

Recombinant phage clones were screened as described by Sambrooke *et al.*, 1989. An appropriate dilution of phage clones was mixed with the host cell strain (typically XL1-Blue or LE392). and incubated at 37°C for 20 minutes to allow attachment of the phage. After this period 4-7ml of molten agar was added to the cell / phage mixture which was then spread out onto a suitable agar plate (*e.g.* the selective antibiotic tetracycline was added to the agar if the host cell was XL1-Blue). The plates were allowed to set for 15 minutes and were then inverted and incubated overnight at 37°C. The plates were then removed from the incubator and allowed to cool to 4°C for 30 minutes to ensure the top agar remained stuck to the bottom agar. Reinforced nitrocellulose filters (Hybond-C EXTRA, Amersham) were then placed in contact with the phage plaques and allowed to sit for 1-20 minutes. (If duplicate lifts were being taken then the first filter was allowed to sit for 30 seconds, the second for 2 minutes, the third for 5 minutes and the final filter lift was allowed to sit on the plate

for a minimum of 20 minutes). Alignment marks were made on the filters by piercing the nitrocellulose with a needle filled with ink and care was taken to ensure the duplicate lifts were pierced in the same place. The filters were removed using two pairs of blunt forceps and placed colony side down into denaturation solution (1.5M-NaCl, 0.5M-NaOH) and left for 1 minute. The filters were then transferred to neutralising solution (1.5M NaCl, 0.5M Tris HCl;pH 8.0 for 2-5 minutes. The filters were then transferred to 2X SSC for a minimum of 10 minutes before being allowed to air dry on 3MM filter paper, colony side up. Filters were baked at 80°C for 2 hours to cross-link the DNA to the nitrocellulose filters prior to being hybridised with a nucleic acid probe.

For screening of a genomic or cDNA library between 5-10000 pfu were plated out per 10cm square plate and when screening a phage plug it was typically the 10^{-4} or 10^{-5} dilutions that were plated out.

2.7 Bacteriophage Plate Lysate.

A single well isolated plaque was picked into 1ml of phage buffer' (Section 2.2.8) containing a drop of $CHCl_3$ and kept at 4°C for 4-6 hours to allow the phage particles to diffuse out of the top agar. 50-100µl of the phage suspension (approximately 10^5pfu) was mixed with 100µl of indicator bacteria and incubated for 20 minutes at 37°C. 3ml of 0.7% top agar was added to the tubes and the mix plated out onto a dry LB plate. The plates were then inverted and incubated overnight at 37°C.

The plates were overlaid with 6ml of phage buffer' and left on a rocker for an hour at room temperature. The phage buffer was removed and centrifuged at 15,000g for 5 minutes to pellet bulk matter. 4ml of the supernatant was removed and 10µl of 100mg ml^{-1} RNase / DNase added and left at room temperature for 30 minutes. To this was then added 4ml of precipitation solution (20% PEG, 2.5M NaCl) and the solution was left on ice for 30 to 60 minutes.

The solution was centrifuged again at 15,000g for 10 minutes, the supernatant poured off and then the tubes were rotated 180° and respun to ensure that all the PEG was removed. (As much PEG was removed as possible to ensure that it did not inhibit any subsequent reactions). The phage pellet was then resuspended in 500µl of phage

buffer and transferred to an eppendorf tube. To this 500µl of phage buffer plus pellet, 5µl of 10% SDS and 5µl of 0.5M EDTA was then added and the phage suspension incubated at 70°C for 15 minutes. A normal phenol / chloroform extraction was then performed (Section 2.5.4). Approximately 300µl of the aqueous phase was then removed and an isopropanol precipitation performed (Section 2.5.5). The solution was left at room temperature for 30 minutes and centrifuged for 30 minutes. The supernatant was removed and the pellet washed with 70% ethanol and once with 100% ethanol. It was then dried in a vacuum drier and resuspended in 100µl of TE. 5µl of this was used for restriction digests.

2.8 Large Scale Bacteriophage Preparation.

0.2ml of fresh host cells (typically XL1-Blue or LE392) were mixed with $5x10^6$ to $5x10^7$ phage particles in phage buffer (Section 2.2.8) and incubated for 20 minutes at 37°C. This cell / phage mixture was then used to inoculate three 250ml conical flasks containing 100ml of NZCYM broth which were grown overnight as described (Section 2.3.4). The flasks were shaken continuously until good lysis was seen to have occurred which typically was after about 15-16 hours.

After lysis the three flasks were combined and spun at 5,000 rpm for 20 minutes in the Beckman JA 14 (6x250ml rotor) to remove cell debris. The supernatant was then placed in fresh 250ml buckets and the phage particles pelleted by centrifugation at 9,000 rpm for four hours at 18°C. As much of the supernatant as possible was then removed before the pellet was air dried and resuspended in 4ml of phage buffer. After resuspension 3.5g of CsCl was dissolved in the solution which was then placed in Beckman 0.5X2" ultra clear tubes and centrifuged to equilibrium (16-18 hours at 35,000 rpm (20°C) in a SW50.1 rotor). The phage particles banded half way up the tubes and were collected by side puncture using a 21G 1 1/2 gauge needle inserted just below the phage band.

2.9 Extraction of Bacteriophage Phage DNA.

CsCl was removed from the purified phage preparation by dialysis at 4°C for a minimum of four hours against 5 litres of TE. After transferring the phage suspension to a centrifuge tube, 0.2 volumes of 0.5M EDTA;pH8.0 were added to a final concentration of 20mM. The tube was then incubated at 65°C for 10 minutes.

Proteinase K or Pronase E was then added to a final concentration of 50μg ml^{-1} as well as SDS to a final concentration of 0.5%. These were mixed by inversion and incubated at 37°C for one hour.

The mixture was then phenol extracted by adding an equal volume of equilibrated phenol and mixed by inversion very gently for 3-5 minutes. This was followed by an equal volume of chloroform and isoamyl alcohol (24:1) which was also mixed gently for 30 seconds. The phases were then separated by centrifugation at 1,600 rpm for 10 minutes at room temperature. The aqueous phase was removed and re-extracted with a 50:50 mixture of phenol and chloroform which was separated by centrifugation as before. A final extraction with chloroform and isoamyl alcohol (24:1) was performed before the sample was dialysed against three changes of 5 litres of TE at 4°C.

2.10 Preparation of *Drosophila* Genomic DNA.

Genomic DNA was prepared essentially as described by Rubin (1986). 0.5-1.5g of adult flies were cooled in liquid nitrogen and ground to a fine powder with a mortar and pestle. This powder was then removed using a precooled spatula to a Dounce homogeniser containing 9ml of homogenisation buffer (0.1M NaCl, 0.03M Tris-HCl;pH8.0, 0.01M EDTA;pH7.0 to which 0.5% Triton and 10mM β–mercaptoethanol had been added. The tissue was homogenised with 10 - 15 strokes of a loose pestle then with 7 - 10 strokes of a tight pestle (the tight pestle allows only nuclei through) and decanted into a 15ml Falcon tube which was centrifuged for 40 seconds in a Jouan CR312 centrifuge at 3,000g to sediment the insoluble material. The supernatant was decanted into a clean corex tube and the nuclei collected by centrifugation at 7,000 rpm for 7 minutes in a pre-cooled (4°C) Beckman JA20 or JA17 rotor. The pellet was resuspended in 1ml of homogenisation buffer without Triton or β–mercaptoethanol and drawn up and down in a Pasteur pipette to ensure thorough mixing. Nuclei were then lysed by the addition of 5ml Nuclear Lysis Buffer (0.1M Tris-HCl;8.0, 0.1M EDTA;pH7.0, 0.1M NaCl and 0.6ml of 10% sarcosyl. The solution was mixed gently by swirling and mixing with a Pasteur pipette. After mixing the solution was decanted into a pre-weighed 40ml oakridge tube, the weight of the supernatant calculated and 1.25g of CsCl added per gram of supernatant (final density of solution=1.7gml^{-1}). After the CsCl had dissolved the solution was loaded into

Beckman 5/8" x 3" rotor tubes, heat sealed and centrifuged to equilibrium at 40,000 rpm at 20°C for 60 hours in a Beckman Ti50 rotor.

After centrifugation the tubes were removed, punctured at their top and 0.5cm above the RNA pellets with 19G2 gauge needles. 0.5ml fractions were collected by dripping the centrifuged solution through the bottom needle and collecting it in eppendorf tubes. 5µl of each fraction was ran out on a 0.8% TBE agarose gel to determine which fractions contained DNA and its approximate size. Those fragments containing DNA were combined and the CsCl removed by dialysing against three changes of 5 litres of TE. The yield and purity of the DNA was estimated using spectrophotometry (Section 2.2.4). Typical yields were 150-170µg DNA per gramme of flies used.

2.11 Isolation of Total RNA.

A modification of the procedure described by Glisin *et al.*, 1974 and Ullrich *et al.*, 1977, was used as described by Berger and Kimmel, 1987 and Sambrook *et al.*, 1989. 0.5 to 10 grammes of developmentally staged tissue was homogenised using a kinemattica (Switzerland) polytron homogeniser, in an appropriate amount (normally 10ml of buffer for every 1g of tissue) of guanidinium thiocyanate homogenisation buffer (4.0M guanidinium thiocyanate, 0.1 M Tris HCl;pH7.5, 1% β–mercaptoethanol). The resulting homogenate was centrifuged at 5,000 rpm in a Beckman JA17 rotor to precipitate insoluble material. The resulting supernatant was layered onto a CsCl cushion (5.7M CsCl, 0.01M EDTA;pH7.5, using a hypodermic needle fitted with a 23G needle, in a RNase free polyallomer centrifuge tube. The CsCl pad was a third of the volume of the tube. The tubes were then filled to around 1mm from the lip, balanced and loaded into the relevant rotor (SW28.1 or SW50.1). They were then centrifuged at either 47,000 rpm for 8 hours (SW50.1) or at 27,000 for 26 hours (SW28.1) at 20°C.

After centrifugation the supernatant was aspirated off to the level of the CsCl cushion and the tube inverted quickly to displace the rest of the contents. The pellet was then allowed to dry for 30 minutes at room temperature. The RNA pellet was resuspended in 1ml of 5% phenol, 3% sarcosine and then transferred to a 15ml falcon tube where it was then allowed to dissolve completely. (Often it was necessary to alternatively heat and chill the pellet to get complete resuspension). Two phenol / chloroform

extractions were performed (Section 2.5.4) on the resulting solution, the solution being made 0.3M with sodium acetate(pH4.0) to ensure the pH of the phenol remained below 7. Two chloroform extractions were then performed to remove all traces of phenol and then 2.5 volumes of ice cold ethanol were added followed by incubation at -20°C to precipitate the RNA. The resulting pellet was washed twice in 70% ethanol at room temperature and then resuspended in a suitable volume of RNase free water. The concentration of the RNA (and any contaminating DNA or guanidinium) was measured spectrophotometrically (Section 2.2.4) The integrity of the RNA was checked visually, firstly by the presence of 'tight' ribosomal bands on a non-denaturing MOPS gel (Section 2.5.8) and secondly by the absence of smearing when a control probe was used to probe the RNA on a Northern blot (Sections 2.5.8.2, 2.5.12 and 2.5.18).

2.12 Isolation of Small Amounts of Total RNA.

The method used was as described by Chomczynski and Sacchi, 1987 with modifications as described by Meltezer *et al.,* 1990. 0.1 to 0.5g of tissue was homogenised at 4°C in 10ml of denaturing buffer (4.0M guanidinium thiocyanate, 25mM sodium citrate;pH7.0, 0.5% Sarkosyl, 0.1M β–mercaptoethanol) using a polytron homogenizer (Kinnemattica, Switzerland). 1ml of 2M sodium acetate (pH4.0), 10ml water saturated phenol and 2ml of chloroform were added successively to the homogenate. The mixture was vortexed after addition of each solution. (For smaller amounts of tissue the volumes of the solutions was halved). The tubes were centrifuged at 4,000g for 10 minutes, the aqueous phase saved and 10ml isopropanol (IPA) added to it to precipitate the RNA. The precipitate was collected by centrifugation at 3,000g for 10 minutes.

The pellet was then resuspended in 2ml of 4M LiCl to solubilise polysaccharides. The insoluble RNA was pelleted by centrifugation at 3,000g for 10 minutes, the resulting pellet then being dissolved in 2ml of water. 2ml of chloroform was added and mixed by vortexing. The samples were then centrifuged at 3,000g for 10 minutes and the upper aqueous phase collected and precipitated by the addition of 1 volume IPA or 2 volumes of ethanol (Section 2.5.5), depending upon the volume of the solution. The RNA integrity was then checked as described previously (Section 2.11)

2.13 Isolation of Poly A$^+$ RNA.

Batch affinity chromatography on oligo(dT) cellulose was utilised essentially as described by Sambrook *et al.*, 1989 and Berger and Kimmel, 1987. Commercially obtained oligo(dT) cellulose (Pharmacia) was swollen at 4°C for 2 hours by the addition of 20ml of 1x oligo(dT) binding buffer (20mM Tris HCl;pH7.6, 0.5M NaCl, 1mM EDTA;pH8.0, 0.1% sodium lauryl sarcosinate). 0.3g (dry weight) of oligo(dT) cellulose was used for each 0.5mg of RNA. The matrix was activated by the addition of 40 volumes of 0.1M NaOH followed by three washes in 40 volumes of 1X oligo(dT) binding buffer. The pH of the final water wash was checked to ensure the pH was 7 or less, (the pH was normally acidic due to the presence of carbonic acid formed when DEPC was broken down into H_20 and CO_2 during autoclaving).

The RNA was dissolved in water and then an equal volume of 2x oligo(dT) binding buffer was added (40mM Tris.HCl;pH 7.6, 1.0M NaCl, 2mM EDTA;pH8.0, 0.2% sodium lauryl sarcosinate). Ten volumes of fresh 1x oligo(dT) binding buffer was added to the oligo(dT) cellulose which was then heated to 65°C for 5 minutes before being added to the oligo(dT) cellulose and allowed to mix on a 'rock and roller' for 45 minutes at room temperature. The oligo(dT) cellulose was then precipitated by centrifugation at 4,000 rpm in a Jouan CR312 centrifuge and the supernatant removed to a fresh Falcon tube. (All supernatants and washes were kept until the end of the procedure to ensure the polyA$^+$ RNA was not lost mistakenly). The oligo(dT) cellulose was washed three times in 40 volumes of 1x oligo(dT) cellulose binding buffer to remove ribosomal RNA before the bound polyA$^+$ RNA was eluted by heating the oligo(dT) cellulose to 70°C in 4x5ml aliquots of water. All four aliquots were combined and 1µl dotted out onto an EtBr plate to show that RNA was still present and had not been lost during the procedure. The 1x oligo(dT) cellulose was then treated in 40 volumes of 0.1M NaOH to hydrolyse any residual bound RNA and to denature any RNases before being washed three times in DEPC treated water. The pH of the final water wash was again checked to ensure the pH was 7 or less.

The whole procedure was then repeated, so in total the RNA was washed six times and bound twice, before it was finally eluted in 4x2ml aliquots of water by heating to 70°C. The RNA was then ethanol precipitated (Section 2.5.5) and the resulting solution put at -20°C overnight to allow precipitation. The RNA was precipitated by

ultra-centrifugation using a pre-cooled SW28.1 rotor at 0°C at 28,000 rpm. The poly(A)$^+$ RNA was then visualised by electrophoresis on a non denaturing 1.0% agarose MOPS gel and compared with total RNA. Typically all the rRNA bands were missing except the 18S which runs at 1.4kb.

2.14 *Exonuclease* III Deletions.

This is an adaptation of the original protocol (Henikoff, 1984), essentially as described in the Promega Protocols and Applications Guide (1993). Around 20µg of super coiled DNA in 50µl of TE was cut with a restriction enzyme (Section 2.5.3) at the required end of the vector polylinker but not within the insert to leave a 3′ overhang (*e.g.* *Apa*I, *Kpn*I, *Pst*I, *Sst*I in pBS II). The DNA was checked to ensure linearisation by running out a small amount on a minigel. (200ng of the DNA was then frozen at -20°C for later controls). If complete digestion had occurred then the plasmid was digested with a restriction enzyme to generate a 5′ overhang either at the same end of the polylinker as above or near that end of the insert. This digestion was checked to ensure the enzyme had cut to completion by incubating 1µl (10-20ng) of the DNA in 2µl of *Exo* III buffer, 16µl of water and 50 units *Exo* III at 37°C for 15 minutes. The DNA frozen after the first digestion was treated in exactly the same way and then both samples were run out on a minigel with untreated samples of both DNAs. The DNA with the 5′ overhang was observed to be the only one degraded, if there was any full length band in this DNA it means the second digest did not go to completion. The DNA was then phenol extracted (Section 2.5.4) and ethanol precipitated (Section 2.5.5) and the pellet then resuspended in 60µl of TE. 6µl of this was then removed and 1µl ran on a mini-gel to check that the approximate DNA concentration expected (>150ng) was present. 6µl of 10X *Exo* III Buffer (0.66M Tris-HCl;pH8.0, 6.6mM $MgCl_2$) was then added to the tube which was then pre-heated to 37°C for 10 minutes.

30µl of 7.4X S1 buffer (0.3M KAc;pH4.6, 2.5M NaCl, 10mM $ZnSO_4$, 50% Glycerol) was diluted into 192µl of ice-cold water. To 9µl of this diluted buffer 1µl (1000 units) of S1 nuclease was added. The solution was mixed by pippetting and 0.7µl (70 units) was added to the S1 buffer to a total volume of 212µl. This was mixed and 7.5µl dispensed into pre-cooled eppendorf tubes on ice, with 2.5µl from the 5µl remaining above, in tube 1. 500-600 units of *Exo* III nuclease was then added to the DNA which

had been pre-heated to 37°C and the digestion allowed to proceed. At approximately 30 second intervals, 2.5µl of the DNA was removed to the S1 mix tubes. After all the samples had been taken, the tubes were incubated at room temperature for 30 minutes before 1µl of S1 stop mix (0.3M Tris base, 0.05M EDTA) was added and the tubes heated to 70°C for 10 minutes. The deletions were checked by running out a small amount of each sample on a minigel; typical rates were 250-350 bp digested per minute.

The DNA was then blunt ended by the addition of 1µl (5 units) of Klenow polymerase at 37°C followed by the addition of 1µl 10mM dNTP mix incubating at 37°C for a further 10 minutes. 40µl of *Exo* III ligase mix was then added to each tube (100µl 10X Ligase Buffer, 100µl 50% PEG, 10µl 0.1M DTT and 5 units of ligase made up to 1000µl with sterile water), which was then incubated at room temperature for 3 hours. The ligated plasmids were then used to transform *E.coli* as described (Section 2.5.7) and transformants obtained. Seven transformants were picked from each time point and plasmids isolated as described (Sections 2.5.1 and 2.5.2). Plasmids in the appropriate size range were then used for sequencing (Section 2.15)

2.15 Nucleotide Sequencing.

2.15.1 Sequencing by the Chain Termination Method.

Sequencing was carried out using the United States Biochemical Corporation (USB) sequencing kit and protocols described in 'Step-By-Step Protocols For Sequencing With Sequenase Version 2.0' were strictly adhered to. Sequenase Version 2.0 is a site directed mutant of bacteriophage T7 DNA polymerase which carries no 3'-5' exonuclease activity and gives a high rate of polymerisation of nucleotides (Tabor and Richardson, 1989).

a) Preparation of Sequencing Template.

Sequencing was carried out using double stranded DNA. In order to use double stranded DNA as a template for sequencing it must be denatured. The alkaline denaturation method described in 'Step-By-Step Protocols For Sequencing With Sequenase Version 2.0' was used for this purpose. The DNA was denatured in 0.2M NaOH, 0.2mM EDTA (30 minutes at 37°C) and the mixture neutralised by the addition of 0.1 volumes of 3M sodium acetate (pH5.5). The DNA was precipitated

with 2.5 volumes of ethanol and the pelleted DNA washed with 70% ethanol, air dried and resuspended in 7µl distilled water before use in sequencing.

b) Annealing Primer to Template:

7µl of prepared template was annealed for 1 hour at 37°C in a mixture also containing:

1µl primer, (pBluescript, -40, -20, T3 or T7, see below)

2µl 200mM Tris-HCl;pH7.5, 100mM $MgCl_2$, 250mM NaCl (5x Sequenase Buffer).

The primers used had the sequences:-

T3	3'-GAAATCACTCCCAATTA-5'.
T7	3'-AATACGACTCACTATAG-5'
M13 (-20)	3'-GTAAAACGACGGCCAGT-5'
M13 Reverse Primer	3'-GTACCAGTATCGACAA-5'

c) Sequencing Reactions:

To the annealed primer / template mix, the following were added; 1µl DTT (100mM), 2µl labelling mix (1.5mM dGTP, 1.5mM dCTP, 1.5mM dTTP), 0.5µl (5µCi) [α^{35}S]-dATP at >600Ci $mmol^{-1}$ (Amersham SJ.304) and 1.5 units of Sequenase Version 2.0. After mixing the labelling reaction was incubated at room temperature for 2-5 minutes. After this incubation, 3.5µl of the labelling reaction were transferred to each of four microfuge tubes labelled A, C, G and T, in which 2.5µl of the appropriate termination mix (80µM dATP, 80mM dCTP, 80mM dGTP, 80mM dTTP, 8mM ddNTP [appropriate analogue], 50mM NaCl) had been spotted at the bottom of the tube. The contents of the tubes were mixed well and incubated at 37°C. After a 5 minutes incubation at 42°C, 4µl of stop mix (95% (v/v) formamide, 20mM EDTA, 0.05% (w/v) Bromophenol Blue, 0.05% (w/v) Xylene Cyanol FF was added to each tube and mixed with the reaction by a brief spin in a microfuge. The sequencing reactions were then stored at -20°C.

For sequencing areas which were particularly G/C rich, the labelling mix contained 3.0µM dITP, 1.5mM dCTP, 1.5µM dTTP and the termination mixes were 80µM dATP, 80µM dCTP, 80µM dITP, 80µM dTTP, 8µM ddNTP (appropriate analogue), 50mM NaCl, except the ddGTP termination mix (80µM dATP, 80µM dCTP, 160µM dITP, 80µM dTTP, 8µM ddNTP [appropriate analogue], 50mM NaCl).

2.15.2 Polyacrylamide Gel Electrophoresis.

The nested set of primer extended fragments produced by the sequencing reactions described above were resolved by electrophoresis on thin polyacrylamide gels using the Bio-Rad SequiGen^{Tm} sequencing gel apparatus. Gels contained:

		Final Concs
40% Acrylamide (38% acrylamide, 2% bisacrylamide)	18ml.	6%
10X TBE	12ml	1X TBE
Urea	60g	8.33M
Distilled Water	36ml	--------
10% (v/v) ammonium persulphate	600µl	0.05%
TEMED (N, N, N', N', tetramethylethylenediamine)	15µl	--------

Gels were pre-run in TBE for at least 30 minutes at 50mA constant current before loading samples. Samples were prepared by heating to 75-80°C for 2 minutes before loading, during which the top of the gel was thoroughly cleared of any unpolymerised acrylamide or urea and the lanes formed using a sharks tooth comb. The heat-denatured samples were loaded immediately onto the gel and electrophoresis carried out at 60W (constant power), 50mA, 2,000V. The duration of electrophoresis depended on the length of sequence required. DNA extending 250bp away from the primer was obtained from gels run for 2-5 hours, or until the dye front reached the anode, while sequence from 200-400bp from the primer was obtained from gels run for 6 hours. Routinely, 3µl of each sample was run until the dye front reached the anode and 3µl run for 6 hours in order to obtain the maximum sequence from each reaction.

All gels were fixed in 10% (v/v) acetic acid, 10% (v/v) methanol for 30 minutes following electrophoresis and were then dried onto a sheet of Whatman 3MM paper and autoradiographed as described (Section 2.2.6).

2.16 Computer Programs used in the Analysis of DNA and Amino Acid Sequences.

A number of programs for the manipulation and analysis of DNA sequences of the UWGCG (University Of Wisconsin Genetics Computer Group) package (Devereux *et al.*, 1984) were run on a VAX computer. This package contains programs for the

comparison of DNA and amino-acid sequences with those in GenBank and EMBL (European Molecular Biology laboratory, Heidelberg, Germany) databases. The following UWGCG programs are some of those that were used:

SeqEd is an interactive editor which allows entering and modification of sequences.

Map displays both strands of a DNA sequence with a restriction map shown above the sequence and possible protein translations shown below.

BestFit makes an optimal alignment of the best segment of similarity between two sequences and inserts gaps if necessary (Smith and Waterman, 1981).

FastA searches for sequences similar to the query sequence in any group of sequences, using a Pearson and Lipman search (Pearson and Lipman, 1988).

TFastA Does a Pearson and Lipman search as for FastA, but first translates the query sequence in all six reading frames.

Strings finds sequences by searching sequence documentation for character patterns *e.g.* ribosomal.

LineUp is a screen editor for editing multiple sequence alignment.

CodonPreference is a frame specific gene finder which tries to recognise protein coding sequences by comparison to a codon frequency table, or by their composition in the third position of each codon (Staden *et al.*, 1982; Gribskov *et al.*, 1984).

TestCode identifies protein coding sequences by plotting a measure of the non-randomness of the composition at every third base (Sharp *et al.*, 1988).

Translate translates nucleotide sequence into peptide sequence.

2.17 *In situ* Hybridisation to Tissue Sections using Digoxigenin-Labelled Probes.

The method used is essentially as described in the Boehringer Mannheim non-radioactive DNA labelling and detection kit protocol.

2.17.1 Subbing of Slides:- Slides were subbed in a solution composed of 0.5% porcine gelatine and 0.05% chromium potassium sulphate. New slides were placed in a tray and placed into a square glass jar, 2-3ml of 10% SDS added and distilled water allowed to run into the jar until no more bubbles were formed (this is to remove any grease or dust). The slides were then covered in the chrome allum solution for 1-2 minutes before being removed to a piece of blotting paper and after air drying for 10 minutes, were placed in a 60°C oven where they were left for >30 minutes to dry completely.

2.17.2 Cutting:- Sections were cut on a Anglian Scientific 620 Compound Cryotome (cryostat). Tissue was first embedded in Tissue-Tek O.C.T. (Miles Inc., Elkhart, USA). After cutting, (typically 8-12µm sections) the sections were dried onto the slides for 1 minute at 45°C, then allowed to dry at room temperature for a further 30 minutes. They were then fixed in a 4% (w/v) paraformaldehyde solution in 1X PBS for 7 minutes and then washed three times in 1X PBS for 5 minutes each wash. Sections were then stored at room temperature until used.

2.17.3 Preparation of Dig-UTP Probes:- 50-1000ng of linearised DNA in 15µl of water was denatured at 100°C for 10 minutes and then snap cooled in an ice/NaCl mix. To this denatured DNA was then added 2µl hexanucleotide mixture (as supplied by Boehringer Mannheim), 2µl dNTP labelling mixture (1mM dATP, dCTP, dGTP, 0.65mM dTTP, 0.35mM Dig-dUTP;pH6.5, finally 1µl (5 units) of Klenow polymerase was added and the whole mixture incubated overnight (12-16 hours) at room temperature. After stopping the reaction with 2µl of 0.2M EDTA;pH8.0 labelled DNA was precipitated by adding 2.5µl 4M LiCl and 75µl pre-chilled 100% ethanol followed by centrifugation at 12000g for 30 minutes at 4°C. The pellet was washed in 500µl of 70% ethanol and then dried in a vacuum drier for 20 minutes before resuspending in 50µl water. To calculate incorporation 10mCi of dCTP was added as

a trace label and a TCA precipitation performed followed by scintillation counting (Section 2.5.15).

2.17.4 Hybridisation of Sections:- Sections were first incubated in 2X SSC for 10 minutes at room temperature, the SSC removed and 200-300µl of pre-hybridisation solution (50% deiononized formamide, 4X SSC, 1X Denhardts, 100µg ml^{-1} freshly denatured salmon sperm DNA, 125µg ml^{-1} yeast tRNA and 10% Dextran sulphate) was added to each slide. The slides were then incubated at room temperature for 60 minutes. All incubations were carried out in rigid plastic boxes lined with several sheets of flat, damp 3MM paper. After the pre-hybridisation step the slides were briefly submerged in 2X SSC to remove the pre-hybridisation solution and after removal of residual 2X SSC with lint-free tissue, 30-50µl of hybridisation solution was added and the whole slide covered with a Parafilm coverslip that was wrinkle free (Hybridisation solution = pre-hybridisation solution plus denatured labelled DNA at a concentration of 50ng labelled DNA ml^{-1}). The slides were then incubated for 16 hours (overnight) at 42°C.

2.17.5 Visualisation of Bound Antibody:- The slides were washed in 2X SSC for 60 minutes, 1X SSC for 60 minutes at room temperature, 0.5X SSC for 30 minutes at 37°C and finally in 1X SSC for 30 minutes again at room temperature. Detection was accomplished using the Boehringer Mannheim digoxigenin non-radioactive nucleic acid detection kit and the protocol used essentially as described by Boehringer Mannheim. All steps were carried out at room temperature.

Slides were washed in Buffer 1 (100mM Tris-HCl;pH8.0, 150mM NaCl;pH7.5), for 1 minute, then incubated for 30 minutes in Buffer 1 plus 2% normal sheep serum (obtained from Scottish Antibody Production Unit, Lanarkshire, Scotland) and 0.3% Triton X-100. The anti-digoxigenin antibody was diluted 1:500 with Buffer 1 plus 1% normal sheep serum and 0.3% Triton X-100 and 100µl applied to each slide which were then incubated for 4 hours in a humid chamber. The slides were then washed in Buffer 1 for 10 minutes with gentle shaking and then in Buffer 2 (100mM Tris-HCl;pH 9.5, 100mM NaCl;pH7.5 and MgCl) for a further 10 minutes with more gentle shaking.

The bound antibody was visualised by applying 400-500µl of colour solution to each slide (Colour solution= 45µl NBT, (70% v/v 75mg ml^{-1} nitroblue tetrazolium salt in dimethylformamide), 35µl X-Phosphate, (5-bromo-4-chloro-3-indolyl phosphate, tolidinium salt 50mg ml^{-1} in dimethylformamide) with 2.4mg Levamisole in 10ml of Buffer 2 to inhibit spurious colour reaction due to endogenous alkaline phosphatases). The slides were then incubated in a light proof, humid chamber for 2 - 24 hours, the colour reaction being monitored continuously and when judged to have stained to completion the reaction was stopped by immersing the slides briefly in Buffer 3 (10mM Tris-HCl;pH8.0, 1mM EDTA;pH8.0). Slides were then mounted (Section 2.2.8.6) using a gelatine / glycerol mix (Ashburner, 1989). Photographs were taken on a Leitz Vario-Orthomat microscope fitted with Nomarski optics.

Chapter Three

Chapter Three - Cloning of the *shak-B* Region and Localisation of Several Chromosomal Aberration Endpoints.

3.1 Introduction.

3.1.1 *Passover* Genetics.

The *Passover* mutation (*Pas*) had been shown to lie proximal to *forked* by recombination and within *Df(1)16-3-22* at the base of the X chromosome (Thomas & Wyman, 1984b). Baird *et al.*, (1990) placed *Pas* near the boundary of the *R-9-29* and *R-9-28* complementation groups by deficiency mapping. They performed a complementation analysis in which *Pas* failed to complement most of the previously identified alleles belonging to the *R-9-29* (now renamed *shaking-B*) locus.

One way to determine how the gene has its effects is by cloning the DNA containing the *shaking-B* (*shak-B*) locus. It would then be possible to isolate transcripts from the gene and so allow tissue expression patterns to be determined. As well as this, once specific transcripts have been identified, it is then easier to the locate the alterations in the genome at the molecular level associated with the different alleles. This then allows the correlation of their position(s) with their ability to stop, or alter the expression of the transcripts. This allows the assigning of specific transcripts or groups of transcripts to specific functions. In the longer term it should be possible to determine the structure and function of the protein products *i.e.* are they transmembrane proteins and where are the protein(s) located in the cell?

It is known that at the genetic level there is a complex interaction between *shak-B*^{Pas} and the other alleles of the *shak-B* locus (Section 1.2.5). This genetic complexity is further compounded by the interaction of the *shak-B* region with proximal deficiencies in the 19E region (as described in Section 1.2.6). These findings have led to the suggestions that *shak-B* may be split by an intervening gene (or genes) located in one or more of its introns. Another possibility is that the coding region is at 19E3 with a long range interaction due to an enhancer or other control region in 19E5-6 (Baird *et al.*, 1990). Alternatively, position effects (for example see Pirotta (1990), and Section 3.10.3) may be having an effect. Cloning the *shak-B* region would allow us to confirm which of these suggestions is responsible for the genetic complexity observed, or to determine if another explanation exists. Once cloned, it would be possible to tie in the

characteristics of the region at the molecular level with the genetics. The molecular endpoints of the various deficiencies and duplications that have been shown genetically to affect the locus when heterozygous with other deficiencies or with *shak-B* alleles, could be determined. This would allow the wealth of genetic information available about the locus and the region as a whole to be utilised. For example any transcribed regions near to key chromosomal abnormalities could be localised, thus giving indications of where the *shak-B* transcript (or transcripts) may lie.

3.1.2 Initial Work in the Cloning of the *shak-B* Locus.

Previous walks that started in the 19E-F region terminated prematurely as they ran into blocks of repetitive DNAs (Miklos *et al.*, 1984). To overcome this problem two microdissection experiments were performed (Pirotta *et al.* 1983; Miklos *et al.*, 1988) on bands 19E to 20, to generate microclones that could be used as entry points for chromosomal walks (Bender *et al.*, 1983). In the first experiment the whole 19EF-20 region was excised but this was later found to have been contaminated with chromosome 4 DNA sequences. In the second experiment four microclone libraries were generated that "began deep in subdivision 20 (mini-library 1) through to subdivision 19E (mini-library 4)". The four libraries together were estimated to contain *ca.* 1500kb of DNA from across the 19E- 20 region. The microclones from each library were then localised by deficiency mapping, using deficiencies and duplications covering or impinging upon the *shak-B* locus (J. Davies, this laboratory). Three walks were started using microclones. A fourth walk was initiated using a genomic phage clone isolated in the chromosomal walk that cloned *runt* (Gergen and Weischaus, 1986) and kindly donated to this study by Peter Gergen. This walk was then extended proximally towards *shak-B*. The three walks started with microclones were named after the number assigned to the microclone *i.e.* 798, 896 and 952. The walk started using the *runt* clone was termed the *runt* walk.

For the cloning of the *shak-B* locus it was decided to use genomic libraries constructed using the EMBL phage vectors (Frischauf *et al.*, 1983) in preference to cosmid vectors (Collins & Hohn, 1978). This was for several reasons. Firstly phage clones are easier to use with respect to analysis, for example restriction mapping. However, the principal reason was that phage vectors are less prone to recombination and / or rearrangements. This point was considered to be important as previous work

at the base of the X chromosome had shown single copy DNA to be interspersed with repetitive DNA at a relatively high frequency (Miklos *et al.*, 1988 and Chapter 4).

In this chapter I report my part in the cloning of the *shak-B* locus and the surrounding regions. I also report the results of studies which used Southern hybridisation to map genetically defined endpoints of chromosomal abnormalities onto the molecular map.

3.2 Results.

3.2.1 The 798 Walk.

The 798 walk was initially thought to lie in the region of interest (19E3). However, the distal endpoints of one deficiency, *Df(1)GA40*, which breaks distally between *runt* and *mell* and proximally deep in division 20 (Table 2.5 and Figure 3.06a,b and c) was shown to be located in the walk. A second deficiency, *Df(1)mal^{lo}*, which breaks distally between *mal* and *LB5* (at 18F4) and proximally between *runt* and *mell* (19E1) (Table 2.5 and Figure 1.03) may also lie within the 90kb encompassed by this walk (Section 3.4). These findings allowed the walk to be oriented and localised to band 19E1, *i.e.* distal to *runt* and near to, or possibly containing, the *melanised like* locus. Two clones were isolated from an EMBL 4 *OrR* genomic library (Section 2.3.1) using a 2.5kb *Eco* RI / *Hin* dIII fragment from λ798C10O4A as a probe. These extended the proximal end of the walk and were called λ7A1 and λ7H7 (Appendix 3 and 3.1 for locations and maps). Studies of this walk and of the *runt* walk were discontinued when it became apparent that the 952 and 896 walks contained DNA from the relevant regions of the chromosome.

3.2.2 The 896 and 952 Walks.

The 896 walk was oriented and localised prior to the start of this study by the localisation of the distal breakpoint of *Df(1)T2-14A* (J. Davies and C. Taylor personal communication). This removes the 19E5 to 19E7 region (Section 2.4.2 and Table 2.5). The distal end of the 896 walk was then extended towards the *shak-B* locus. The 952 walk was oriented and localised in a similar manner, firstly by crossing the distal breakpoint of *Df(1)LB6* (Section 3.7) which removes the 19E3 to 20A2 region and also by crossing the proximal breakpoint of *Df(1)16-3-35* (Section 3.6). The distal end of the walk was extended (towards *runt*) by the isolation of four phage clones. These were isolated using a 3.4kb *Hin* dIII / *Hin* dIII fragment from λC5BO3 as a probe.

Several clones were isolated from the EMBL 4 *Oregon R* (*OrR*) genomic library and the two extending the furthest were restriction mapped. These were named λAGO01 and λAGO02 (Appendix 1 and 1.1 for locations and maps).

When λAGO02 was hybridised to a Southern blot of the restricted *runt* walk phages, cross hybridisation occurred to the most proximal *runt* walk clone, λRC56. To determine which fragment(s) were cross hybridising, the proximal clones from the *runt* walk and the most distal clones from the 952 walk were restricted with the appropriate enzymes, electrophoresed through an agarose gel, and then bidirectionally blotted (Sections 2.5.11) to nitrocellulose. One of the blots was probed with the most terminal fragment of λAGO02 (Figure 3.01a), a 1.0kb *Eco* RI fragment that contains an *opa* repeat (Section 4.3.6 for details). The duplicate blot was probed with the single copy 1.9kb *Eco* RI to *Hin* dIII fragment that lies just proximal to the 1.0kb fragment (Figure 3.01b). The *opa* repeat containing fragment hybridised to the two most proximal fragments of λRC56, both of which are known to be repetitive and thus could contain *opa* repeats themselves. The 1.9kb region hybridised only to the expected regions in the 952 walk, suggesting that the homology observed between the *runt* and 952 walks was due solely to the repetitive sequences.

For the next step in the walk, several clones were isolated from a library prepared using *Oregon-R* DNA in the phage lambda Gem 11 vector (S. Tomlinson, personal communication, Section 2.3.1). These were isolated using a 1.9kb *Eco* RI / *Hin* dIII fragment from λAG0O2 as a probe. The two that appeared to extend the walk the furthest were restriction mapped These were termed λAGO07 and λAGO09 (Appendix 1 and 1.1 for locations and maps). No hybridisation of the most distal restriction fragment to the *runt* walk was found. The distal endpoint of $Dp(1)Ymal^{171}$, which covers the region 19E3 to deep within division 20 (Section 2.4.2, Table 2.6) was shown to lie within λAGO01 (Section 3.5). This duplication covers the *shak-B* locus but not *runt* (Gergen and Wieschaus, 1986 and personal observations) so walking in this direction was discontinued as the distal extent of *shak-B* had thus been reached.

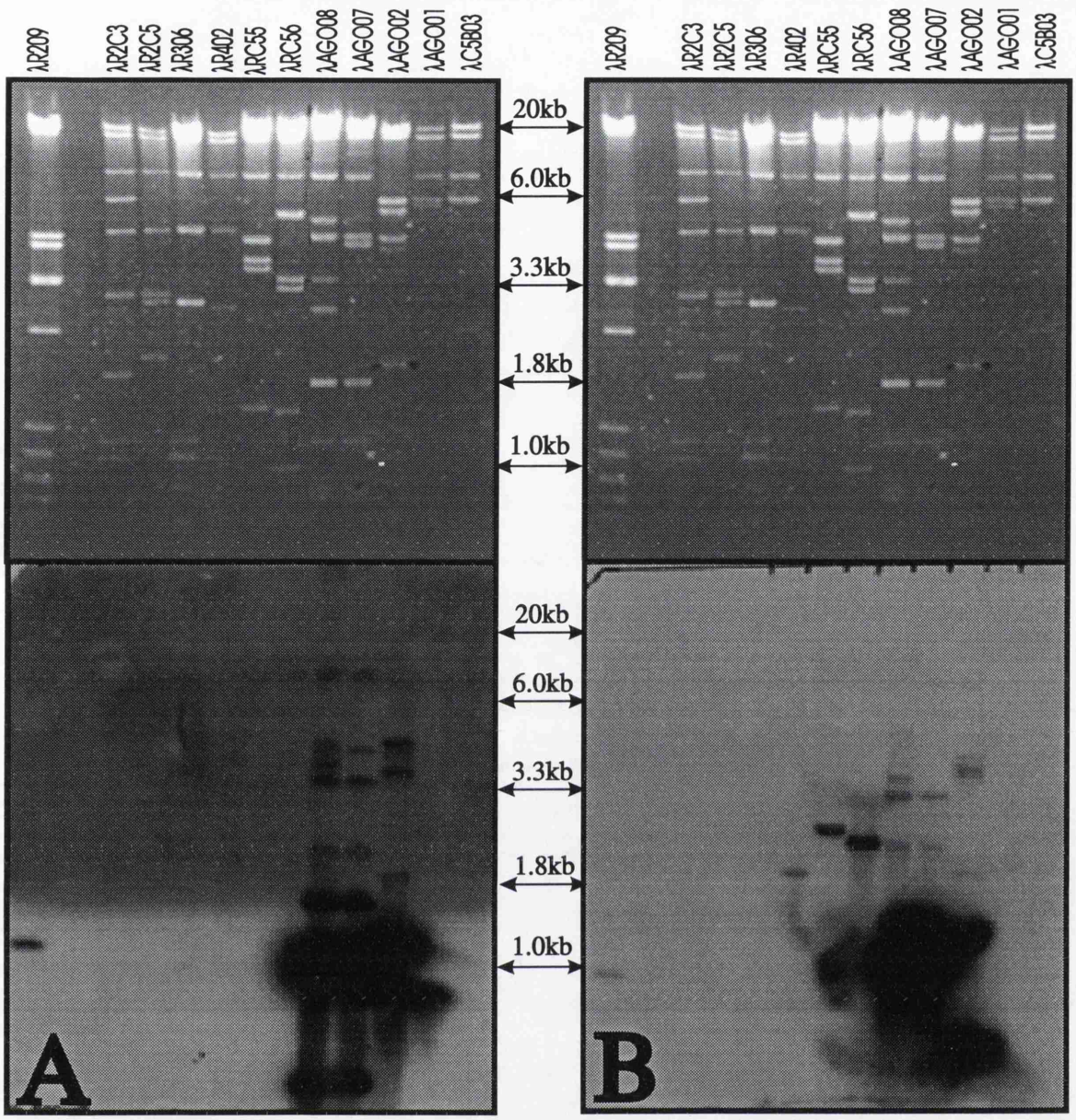


Figure 3.01a and b-Ethidium bromide stained gel and duplicate Southern blots of phage clones from the distal end of the 952 walk and the proximal end of the *runt* walk hybridised with A)1.9kb *Eco* RI *Hin* dIII fragment from λAGOO2; B)1.0kb *Eco* RI *Hin* dIII fragment from λAGOO2. No hybridisation is observed to *runt* walk sequences with the 1.9kb fragment whereas the 1.0kb probe hybridised only to repetitive regions. See text (3.2.2) and Figure 3.01c for details. The genomic clones were restricted with the enzymes detailed in Section 5.3.

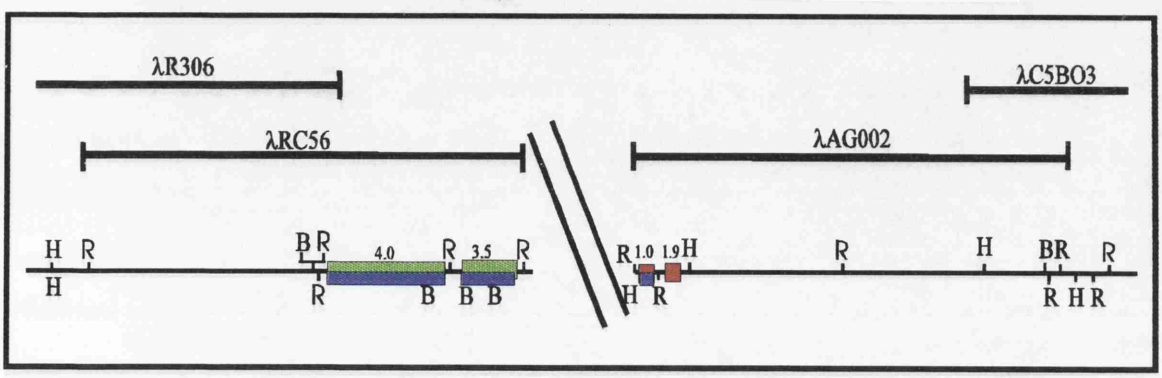


Figure 3.01c-Schematic diagram of the results from Figures 3.01A and B. See text (3.2.2) for explanation. *opa* repeat containing regions are shown in blue, whereas the fragments used as probes in the study are indicated red. The regions that hybridised (in the *runt* walk) with the 1.0kb fragment are indicated green.

3.2.3 Extension of the 952 Walk Proximally.

This study isolated one phage clone from the proximal end of the 952 walk from an EMBL 3 *CS* genomic library (Section 2.3.1). This clone was termed λ94C11 (Appendix 1 for location and see Figure 1.04 for map). The 896 and 952 walks were extended towards each other until both walks ran into regions that seemed to be unrepresented in the phage genomic libraries. The reasons for this are unknown. However, the distal end of the 896 walk (λH683) is slightly repetitive (see Chapter 4) and 30% of the DNA within the phage clones eventually isolated from between the walks (see below) is repetitive DNA. It was thus thought possible that, as in previous walks in this region (Miklos *et al.,* 1984; 1988), blocks of repeats were interfering with cloning in some manner. In an attempt to traverse this region, cloned single copy fragments, from the relevant ends of the 896 and 952 walks were isolated and used to isolate cosmid clones. The fragments were a 4.3 *Eco* RI / *Eco* RI fragment from λ9915 at the proximal end of the 952 walk and a 2.7kb *Hin* dIII / *Sal* I fragment from λH486 at almost the distal end of the 896 walk. Fragments from the most distal clone of the 896 walk (λH486) were not used as they were shown to be slightly repetitive (see Chapter 4 and Appendix 2). The rational behind this approach was that cosmid clones, being two to three times the size of phage clones, would extend over this area. Two cosmids were isolated using the 952 walk fragment, six using the 896 walk fragment, and six that seemed to cross-hybridise weakly with both fragments (Jorg Hoheisel personal communication).

DNA was prepared in duplicate from each cosmid by minipreps (Section 2.5.2) and the DNA digested with various restriction enzymes and Southern blotted. These Southern blots were then hybridised with the original fragments used to isolate them from the cosmid library. This was done to verify that the cosmids originated from the region of interest. One of the 952 cosmids and the six clones that seemed to hybridise with both the 896 and 952 fragments, did not hybridise back to the probe used to select them, and so were not characterised further. A single 952 cosmid and the 896 cosmids were found to hybridise to the fragments used to isolate them (Figure 3.02) but not to hybridise to each other (other than vector sequences). The extent to which the cosmids overlapped with their original parent walks was then determined by sequentially hybridising whole phage clones, from the ends of the walks, to restriction digested and Southern blotted cosmid DNA (Figure 3.03a and b). The 952 cosmid

(c9A1) was found to be approximately 55kb in size of which approximately 11kb was not previously cloned in the 952 walk. Of the 896 cosmids the largest (c8C1) was approximately 45kb of which 17kb was not previously cloned in the 896 walk.

The restriction digested cosmids were then hybridised with nick-translated (Section 2.5.13.1) genomic DNA to identify regions containing repetitive DNAs. Bands unique to the cosmid and not containing repetitive DNA were then isolated, radiolabelled and used to screen the *CS*, EMBL 3 genomic library (S. Ji, personal communication). Three clones were isolated using fragments from the c9A1 cosmid and mapped (see Figure 3.04 and Appendix 1 for maps and locations). A single phage clone was isolated using a fragment from the cosmid c8C1 (S. Ji, 1995). All the phage clones failed to hybridise to their original parent walks. The new 952 and 896 phage clones did not cross hybridise to each other. Although all of the phage clones isolated hybridised to the cosmids used to isolate them, and the three isolated from the end of the 952 walk hybridised to each other, the patterns of hybridisation to the cosmid were not consistent with the restriction maps (Figure 3.04b). It was shown (Figure 3.04) that the three clones isolated with the c9A1 fragments have internal regions that hybridised to the cosmid, but terminal regions that did not. Several possible explanations could account for this result. It was possible that as the region of homology between the phage clones and the cosmid is repetitive all three clones may have originated from elsewhere in the genome, the detected homology being a result of repetitive DNA. Alternatively, it was thought possible that the cosmid c9A1 may have had an internal deletion which brought together non- contiguous fragments. Further analysis using the reverse Southern technique (Section 4.2.1) showed that the regions of homology (other than the fragment used as a probe to isolate the phage clones originally) were indeed partially repetitive (Figure 4.06). However, a genomic clone isolated after this study was finished (S.Ji and J Davies, personal communication) 'filled the gap' between the genomic clones and the end of the 952 walk. The most distal regions of λS4 and λS5 however (highlighted in green in Figure 3.04B) do not hybridise to the new phage clone (λ9FD3) or to the cosmid. This result, together with the fact that the S series phages and the cosmid were isolated from libraries made from different strains, suggests that that the 'extra' DNA may be a transposable element which has inserted into the region, *i.e.* it is present in this region in the genome of the strain from which the S series phages were isolated but not in

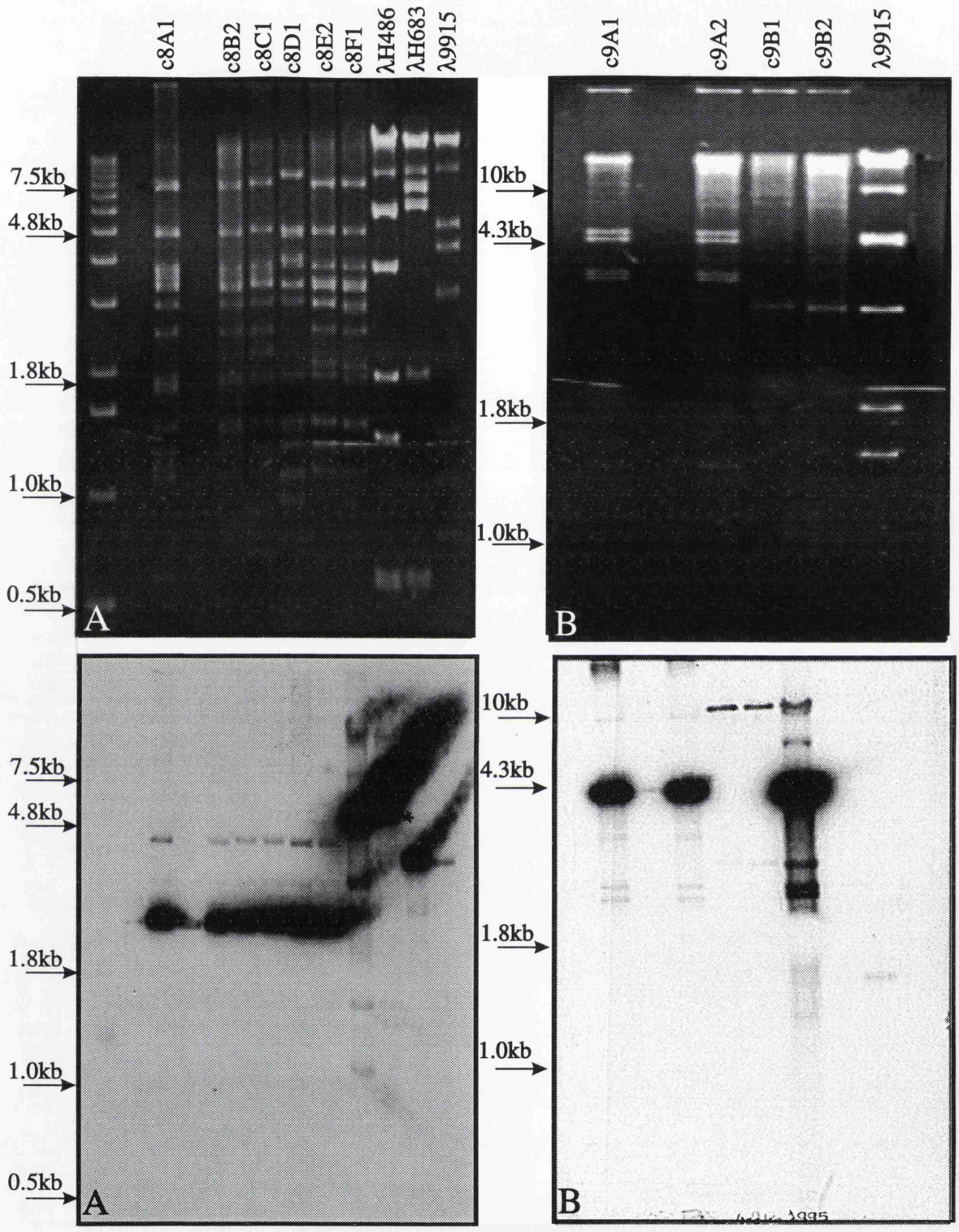


Figure 3.02- Ethidium bromide stained gel and Southern blots of restricted cosmid and phage DNA (cosmids tracks are prefixed with c), hybridised with the fragments used to isolate them from the cosmid library. The 896 walk cosmids were challenged with a 2.7kb *Eco* RI - *Sal* I fragment from λH683 (A). The 952 walk cosmids were challenged with a 4.3kb *Eco* RI fragment from λ9915 (B). See text (3.2.2) for further details. The band that is hybridising in the λ9915 track in A comes from a previous hybridisation in which the cosmids were checked for cross homology to the other walks. As only this band in the λ9915 positive control was detected, the filter was not stripped. The hybridising band in the λH683 track is partially obscured by the smear from the λH486 track so it has been asterixed for clarity.

The cosmids were digested with *Eco* RI and *Sal* I, the phages with the enzymes listed in Section 5.3.

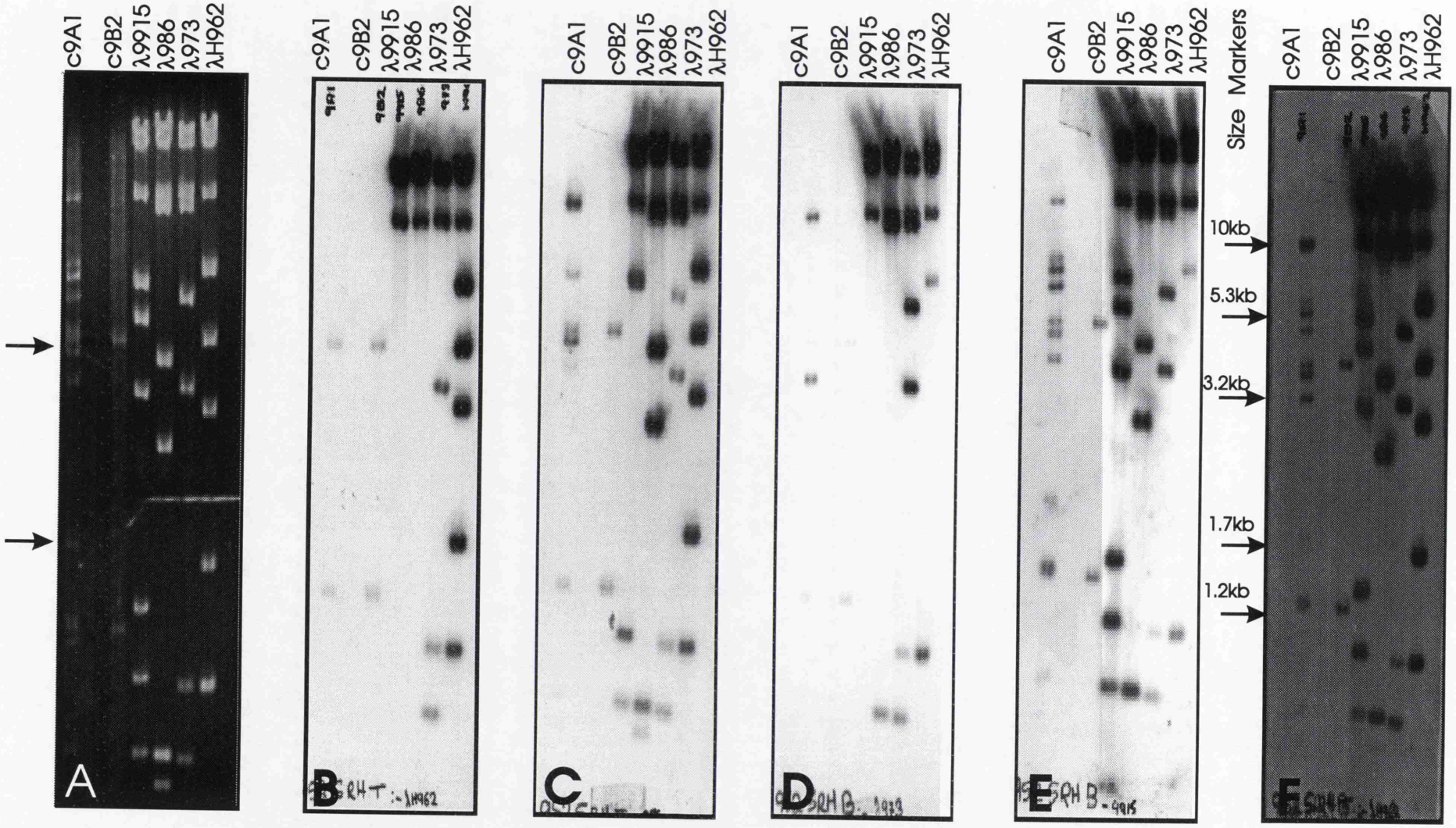


Figure 3.03a-Sequential hybridisation of phage clones from the proximal end of the 952 walk to the 952 walk cosmids and to 952 walk phages. A= Ethidium bromide stained gel showing the restricted phage clones, B=λH962 as probe, C=λH973 as probe, D=λH968 as probe, E=λ9915 as probe, F=Composite of four results, made by overlaying the four autorads. The arrows on the ethidium bromide stained gel (A) indicate comid restriction fragments that fail to hybridise to the 952 walk phages. (Cosmids are labelled c, phages are labelled λ).

The cosmids were digested with *Eco* RI and *Sal* I, the phages with the enzymes listed in Section 5.3.

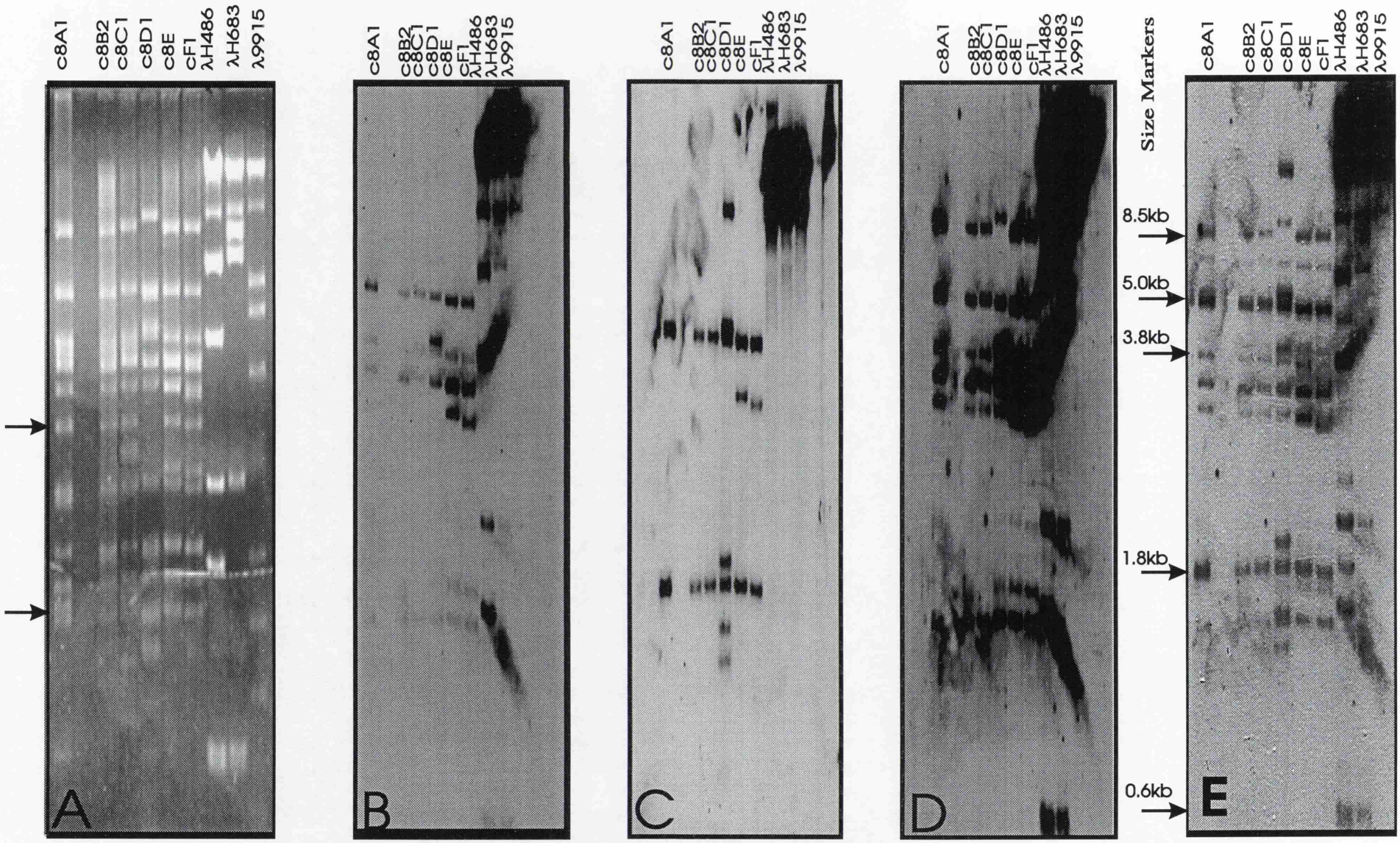


Figure 3.03b-Hybridisation of phage clones from the distal end of the 896 walk to the 896 walk cosmids. A=Ethidium bromide stained gel showing the restricted phage clones, B=λH382 as probe, C=λM283 as probe D=λH683 as probe, E=composite image of all three results made by overlaying the three autorads. The arrows on the gel photograph (A) indicate bands that fail to hybridise (*i.e.* do not overlap with) the 896 walk. (Cosmids are labelled c, phages are labelled λ).

The cosmids were digested with *Eco* RI and *Sal* I, the phages with the enzymes listed in Section 5.3.

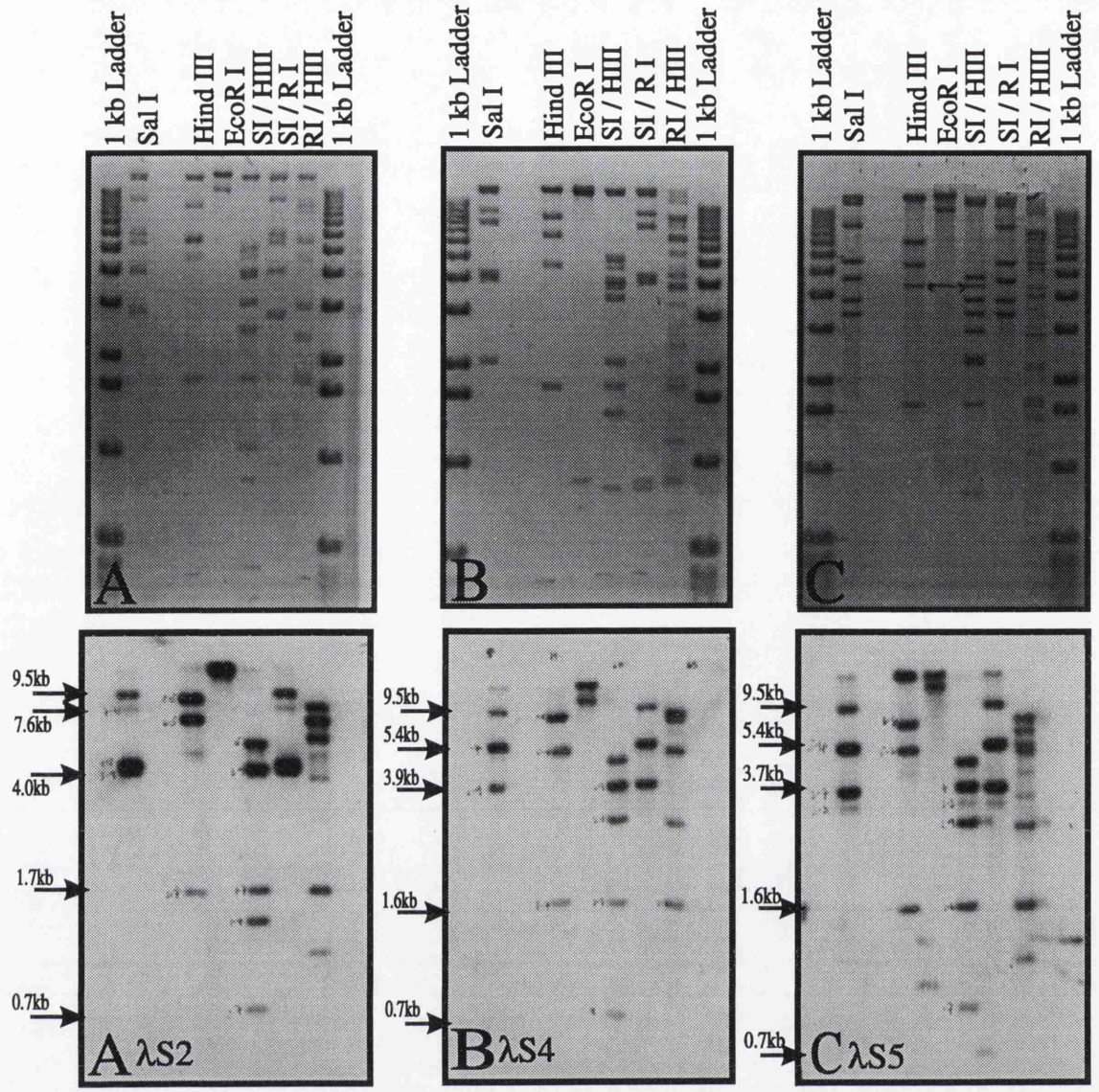


Figure 3.04a-Ethidium bromide stained gels and Southern blots of the λS genomic clones (isolated using fragments from the 952 walk cosmid, c9A1) hybridised with the c9A1cosmid. The molecular weights of some restriction fragments are indicated.

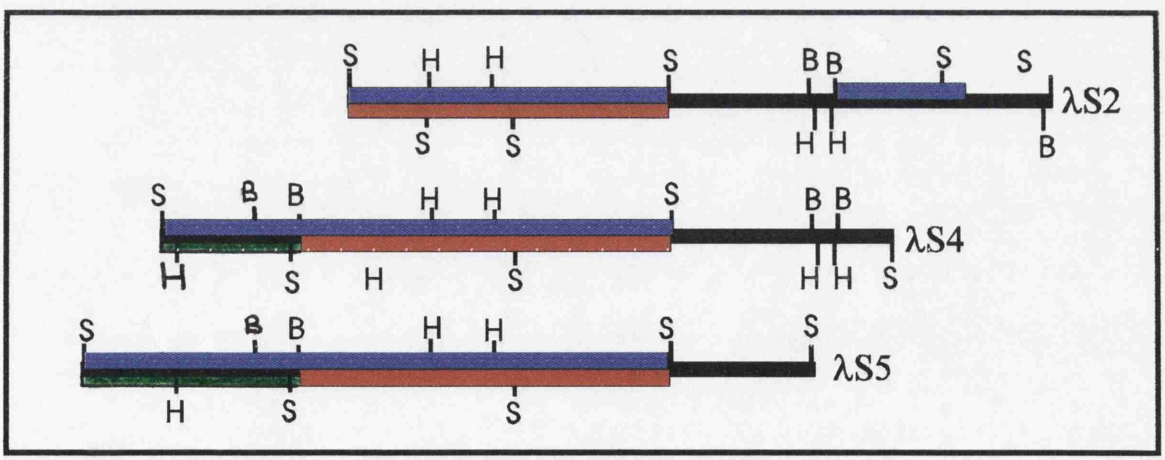


Figure 3.04b-Schematic diagram of the results from Figure 3.04a. The regions of overlap between λS2, 4 and 5 and the cosmid (c9A1) are shown in red. Repetitive regions are shown in blue. No homology exists at the termini of the inserts of the genomic clones to the parent cosmid.

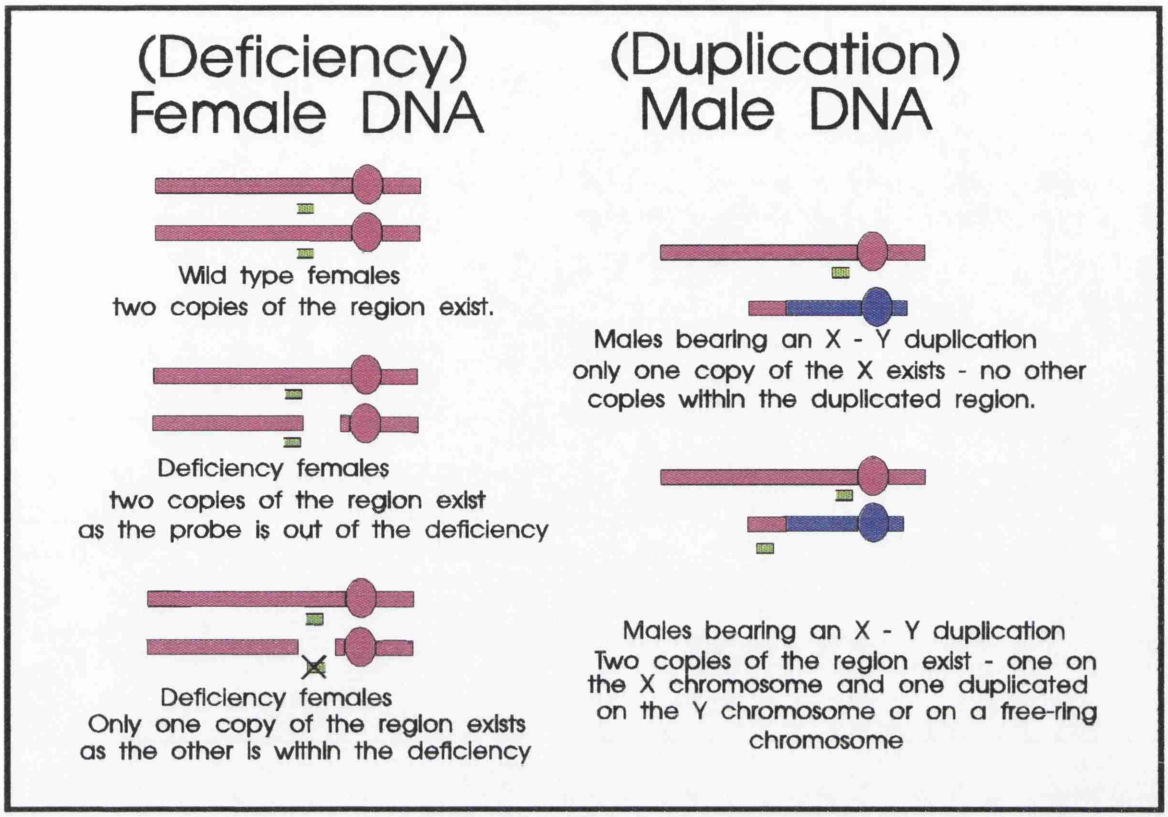
those from the strains which the cosmid clone c9A1 and λ9FD3 were isolated (See Appendix 1.2). The region surrounding the insertion is repetitive and sequence analysis (Section 4.2.4) suggests it may be the remnants of a retrotransposon. The re-entry of transposable elements into the genome by inserting into other transposable elements is a well documented phenomenon (Dawid *et al.,* 1981; Scherer *et al.,* 1982; Di Nocera and Dawid, 1983; Young *et al.,* 1983). The genomic clones extending the 952 walk proximally (the S series phages) do not hybridise to the 896 walk, and as yet the two walks have not been joined.

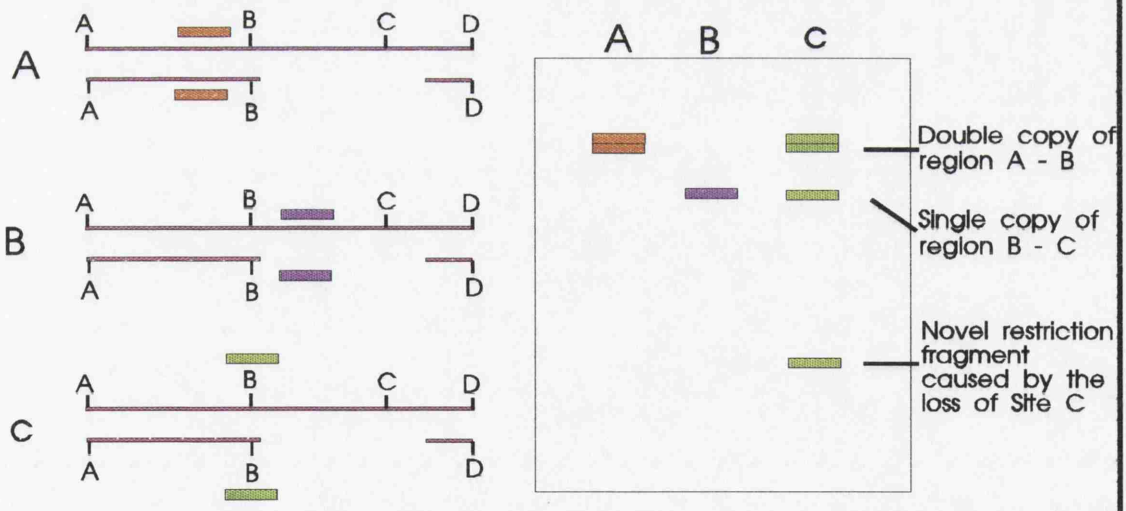
3.3 Mapping Chromosomal Aberration Endpoints by Southern Hybridisation.

3.3.1 Approximate Location of Deficiency and Duplication Endpoints Within the Walks.

Genomic DNA was prepared from fly strains bearing the relevant deficiencies and duplications (Tables 2.5 and 2.6). Heterozygous females were collected to prepare the DNA from strains bearing chromosomal deficiencies. For strains bearing duplications, males were collected. The rationale behind this is summarised in Figure 3.05A. On hybridisation to Southern blots, a band of half the intensity is expected if the probe falls into a region uncovered by a deficiency, relative to a probe that is outwith the deficiency (Figure 3.05B - probe B). Females were not used for the analysis of duplications as there would be two copies of the X chromosome outside the duplication compared to three within the duplication (two copies of the X and one copy of the duplicated DNA). This intensity ratio of 2 : 3 or 2 : 2 is more difficult to distinguish than the 1 : 2 or 1 : 1 intensities obtained using male flies (one X chromosome and one copy of the duplicated DNA). When the endpoint of a deficiency or a duplication was localised to a small enough region (>10kb), this region was then hybridised to a genomic blot of wild-type and deficiency DNAs. If the endpoint lies within the hybridising region then novel restriction fragments caused by the loss of the original restriction fragment should be visible (summarised in Figure 3.05B - probe C).

Three tracks of genomic DNA, comprising two unknowns and a strain which had been previously characterised (*i.e.* the endpoints of the deficiency / duplication had been previously determined or which had been shown genetically to break some distance from the *shak-B* region), were electrophoresed and prepared as described (Sections





In A the probe hybridises to a region which has two copies and thus a signal intensity of two will be seen. In B an intensity of one is observed as the second area (B - C) is within the deficiency and so only the non-deficient chromosome will be detected. In C the restriction fragment originating from region A - B will give an intensity of two as both copies of the chromosome exist, whilst the region B - C will give an intensity of one as it does not exist on the deficiency chromosome. A novel restriction fragment is formed by the joining of region B to D which will have an intensity of one as it only occurs in the deficiency chromsome.

Figure 3.05- Rationale behind the approach used in mapping the Deficiency and Duplication breakpoints and the use of 'multiblots' to determine the actual molecular endpoints. See text (3.3.1) for further explanations.

2.5.3; 2.5.8.1; 2.5.11; 2.5.16 and 2.5.17). As equal amounts of restricted genomic DNA were loaded, it was possible to make a direct comparison between the intensity of the signals in different tracks. The 'correctness' of estimation by the naked eye was confirmed using laser densitometry of the autoradiographic results under conditions recommended by Swillens *et al.,* (1988). The intensity of the signal was observed to be either half, equal to or double that of the control DNA track depending upon its source. Using this procedure 7 of the 10 deficiency / duplication breakpoints investigated were either localised to specific restriction fragments from the walks, or shown to be outwith the regions presently cloned.

A further two breakpoints (*Dp(1)mm2* and *Dp(1)mmRing*) were partially characterised in that they were shown not to break in the 952 walk. Results using fragments as probes from the proximal and distal ends of the 896 walk were confusing however as they seemed to suggest that the 896 walk was not covered by these duplications. The results were not consistent with the original cytological location of the breakpoints (Green *et al.,* 1987). Other studies (Prud'homme *et al.,* 1995) suggest that in at least one case (*Dp(1) mm2*) the organisation of the duplication has altered from that originally described. They place the distal endpoint of *Dp(1) mm2* at 19F which would agree with the results this study obtained with the 896 walk probes.

Problems were encountered with the chromosomal abnormality stocks that were ordered to complete the analysis of *Df(1)LB7*. *Df(1)LB7* (Section 2.4.2, Table 2.5) stocks were obtained from two different centres; Glasgow and Umea, Sweden. On crossing to flies bearing the lethal *shak-B* alleles, *shak-B*$^{R\text{-}9\text{-}29}$ and *shak-B*^{L41} (Section 2.4.2, Table 2.5) no complementation was seen. Flies were then obtained from the Department of Cell Sciences, University of Leiden, Leiden, Netherlands, where the stock was originally produced. This stock also failed to produce the expected F1 progeny, suggesting contamination or a recombination / rearrangement event may have occurred prior to the distribution of the stock to the centres mentioned.

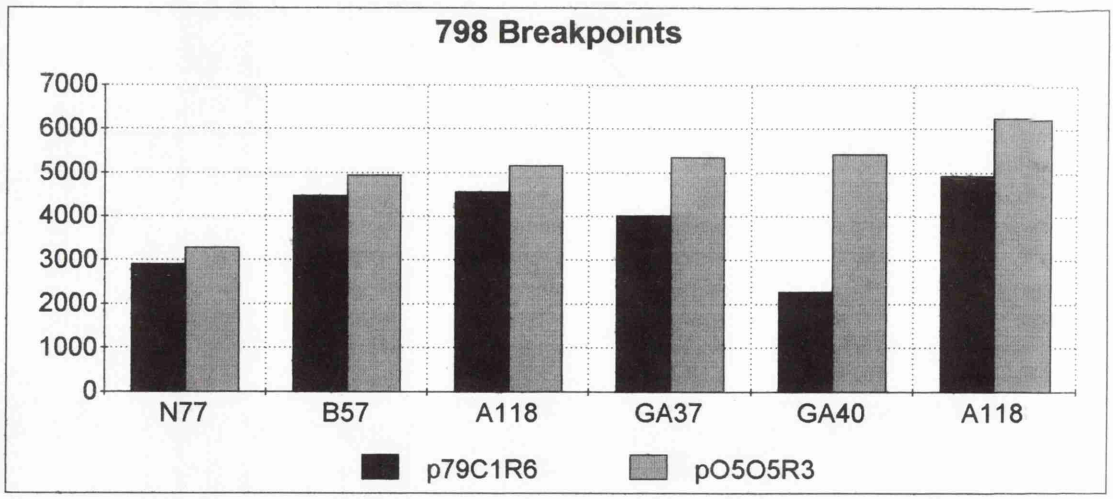
Results with *Dp(1)mm2* and *Dp(1)mmRing* DNA, collected in the early stages of this study, suggested the endpoints of these duplications occurred outwith the 952 walk although as mentioned above the results with 896 walk probes were confusing. After this initial analysis but prior to a more detailed analysis of these breakpoints it was

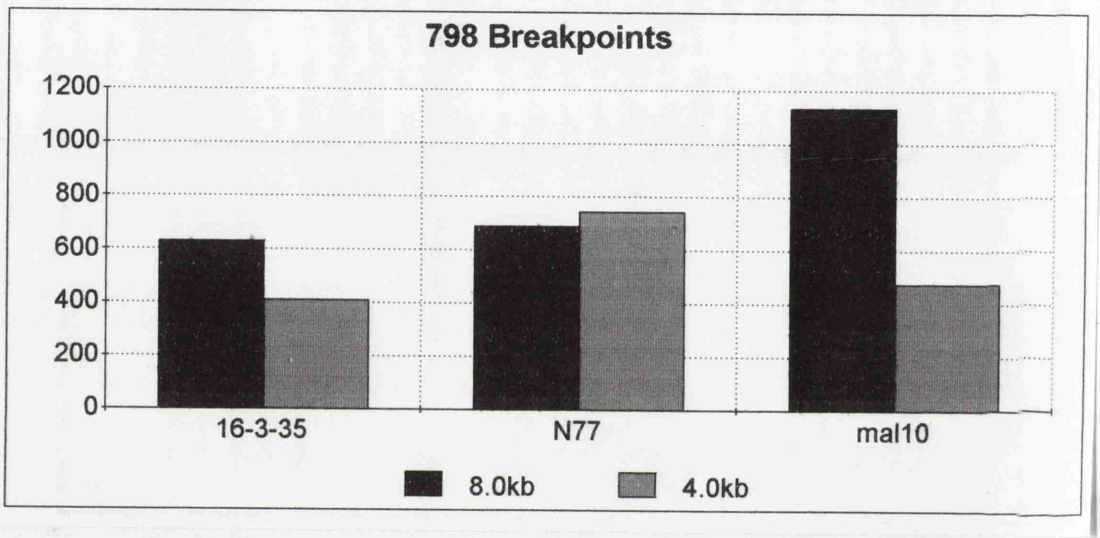
noted that both the *Dp(1)mmRing* and *Dp(1)mm2* stocks maintained at Glasgow, had in the intervening time, lost the *yellow* phenotype indicative of the loss of the mini-ring chromosome. Other stocks were ordered from three different stock centres, but the flies died in transit or shortly after arrival. Time constraints meant the reordering and retesting of these three strains was unfeasible. As no more DNA was available for testing it was impossible to localise the distal endpoints of *Dp(1)mmRing* and *Dp(1)mm2*, any further than in the early parts of this study *i.e.* the endpoints are located outwith the 952 walk (Section 3.9).

3.4 Location of Deficiency Endpoints Within the 798 Walk.

As previously mentioned, the 798 walk was originally thought to be in the region of interest (19E3); however, preliminary analysis showed that the distal end of *Df(1)GA40* (and probably the distal end of *Df(1)N77* and the proximal end of *Df(1)mal^{l0}*) were situated within this walk. Five deficiency DNAs, *Df(1)N77*, *Df(1)B57*, *Df(1)GA37*, *Df(1)GA40* and *Df(1)A118* (see Section 2.4.2, Table 2.5, Figure 1.03 for details), were restricted with *Bam* HI. The fragments produced were electrophoresed on a 0.8% agarose gel, blotted, and probed using two single copy *Eco* RI fragments from each end of the 798 walk. These were p0505R3, which hybridises to a 5.0kb *Bam*HI fragment, and p79C1R6, which hybridises to a 6.5kb *Bam*HI fragment (Figure 3.06 and Appendix 3 for locations). The results obtained are shown in Figure 3.06a. Comparison of signal intensities between the control DNA, *Df(1)A118* (known not to uncover this region thus giving an intensity of two), and the experimental DNA showed that the distal breakpoint of *Df(1)GA40* was situated between these two fragments. The other deficiencies tested did not break in the walk, except *Df(1)N77* which breaks distal to both fragments.

Two fragments (respectively more distal and proximal than those above) p7C10O4AR1 and a 3.2kb *Eco* RI fragment from the phage clone λO5R3O8B (see Figure 3.06 for locations), were hybridised to *Eco* RI digested *Df(1)16-3-35*, *Df(1)N77* and *Df(1)mal^{l0}* DNAs. The intensities in the *Df(1)mal^{l0}* tracks were consistent with the 3.2kb *Eco* RI fragment being in the deficiency [see C, Figure 3.06a], with p79C10O4AR1 from the other end of the walk being out of the deficiency [see D, Figure 3.06a]). This suggests the proximal end of *Df(1)mal^{l0}* breaks in the 798 walk. In the *Df(1) N77* track the 3.2kb λ05R308B fragment hybridises with an intensity





Df DNA in track / Fragment used as probe	NUMBER OF PIXELS DETECTED PER BAND					
	N77	**B57**	**A118**	**GA37**	**GA40**	**A118**
p79C1R6	2906	4464	4551	4009	2283	4921
p0505R3	3258	4932	5156	5341	5423	6240

Df DNA in track / Fragment used as probe	NUMBER OF PIXELS DETECTED PER BAND		
	16-3-35	**N77**	**mal[10]**
8.0kb	630.4	688.6	1133
4.0kb	409.1	744.9	476.6

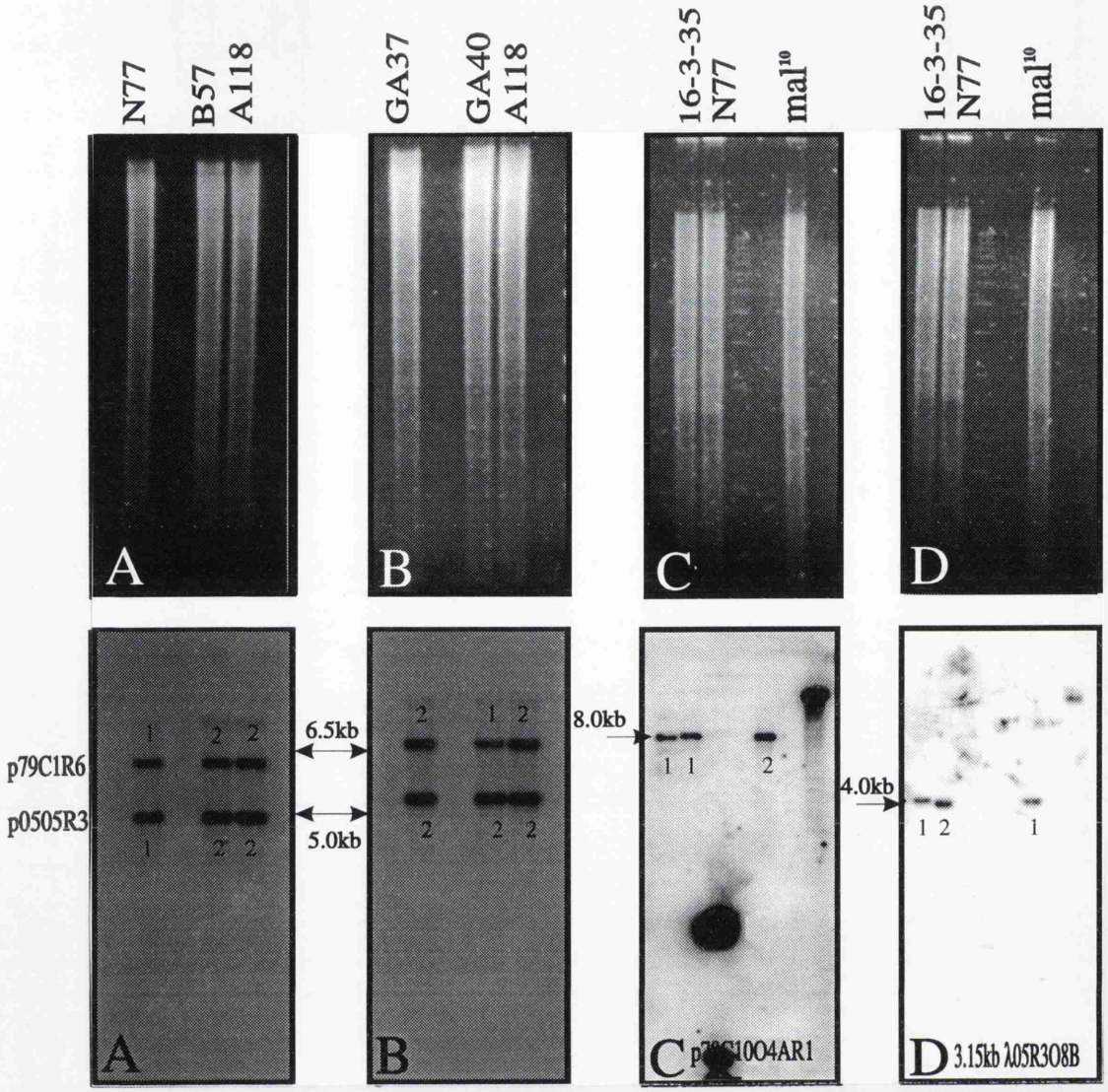


Figure 3.06a-Ethidium bromide stained gel and Southern blots of various deficiency DNAs that have endpoints in the 19E1 region. Df(1) *GA37, GA40, N77, B57* and *mal*^{10} are the uncharacterised deficiency DNAs. The Df(1)*16-3-35* and *A118* DNAs were included as intensity controls. A/B were restricied with *Bam* HI whereas C/D were restricted with *Eco* RI. The regions used as probes are indicated on each image and on the diagram below. The distal fragments (p0505R3 and 3.15kb λ05R3O8b) hybridise to 5kb and 4kb fragments respectivly, whereas the proximal fragments (p79C1R6 and p79C10O4AR1) hybridise to 6.5kb and 8.0kb fragments respectivly

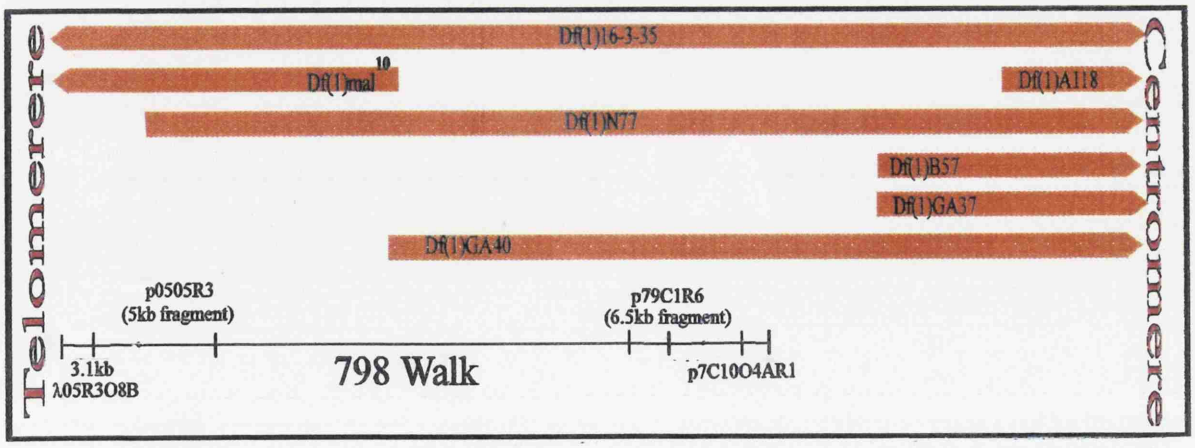


Figure 3.06b-Schematic represention of the deficiencies in the area in which the 798 walk is situated. *Df(1)mal*^{10}, *Df(1)N77* and *Df(1)GA40* all break in the walk itself, whereas the other deficiency breakpoints are situated outwith the region presently cloned.

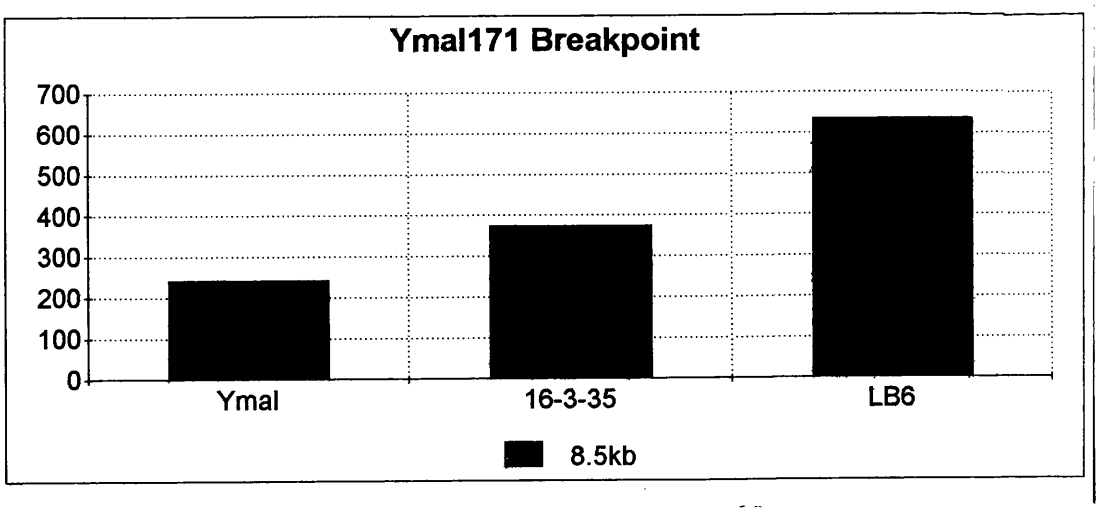
double that of the p7C1004AR1 fragment. As the result in the *Df(1) N77* track in Figure 3.06A suggests that the distal breakpoint of the deficiency uncovers p79C1R6 but fails to cover p7C1004AR1 this strongly suggests that that the distal endpoint of *Df(1) N77* is located between p79C1R6 and P7C1004AR1 *i.e.* within in a 4kb region.

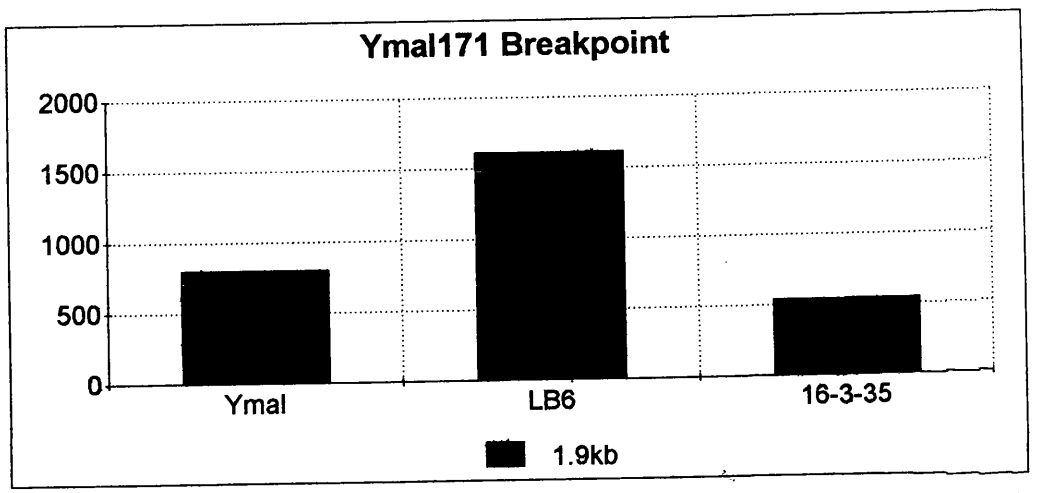
These results allow the orientation and localisation of the 798 walk, with p79C1R6 and p7C1004AR1 proximal and the 3.2kb λ05R308B and p0505R3 fragments being distal. The walk was therefore localised to an area distal to *runt* but proximal to *maroonlike i.e.* containing the *mellanised like* region (Figure 3.06b for summary of above). These and other results obtained at the same time (see below) suggested that the 952 and 896 walks, and not the 798 walk, were the walks containing or nearest to the *shak-B* locus. A more detailed analysis of the regions encompassed by the 952 and 896 walks was therefore undertaken.

3.5 Location of the Distal End of *Dp(1)Ymal*^{171}.

The *Ymal*^{171} duplication has been shown genetically to cover *shak-B* but not to cover *runt*, and so it defines the distal limit of the *shak-B* locus. A 1.9kb *Eco* RI / *Hin* dIII fragment, cloned and isolated from λAGO02, was shown to be outwith the duplication. This was demonstrated by comparing the signal intensity of the fragment when hybridised to two blots of *Dp(1)Ymal*^{171} DNA and equivalent amounts of two control DNAs (*Df(1)LB6* and *Df(1)16-3-35*). One blot was restricted with *Eco* RI (Figure 3.07a), and the other with *Eco* RI and *Hin* dIII (Figure 3.07b). *Df(1)LB6* removes DNA proximal to this region but does not uncover this fragment (Section 3.7) and therefore gives a signal intensity of two, as both copies of the X chromosome exist in this region. *Df(1)16-3-35* removes DNA distal to this fragment and has been shown to uncover this region (Section 3.6) and therefore gives an intensity of one as the deficiency chromosome has this region missing. The *Dp(1)Ymal*^{171} track gave an intensity of one on both blots using this 1.9kb region, and so this fragment lies outwith the duplication.

A 3.4kb *Hin* dIII fragment from the phage clone λC5B03 was hybridised to a genomic Southern blot of deficiency and duplication DNAs digested with *Hin* dIII. The distal *Hin* dIII site, bracketed and marked **X** in Figure 3.07, is polymorphic *i.e.* it occurs on the balancer chromosome FM6 (Lindsley and Zimm, 1990) but not on the other X or





Df DNA in track / Fragment used as probe	NUMBER OF PIXELS DETECTED PER BAND		
	Ymal[171]	**16-3-35**	**LB6**
8.5kb	242.8	374.7	633.5
1.9kb	801.8	1611	538

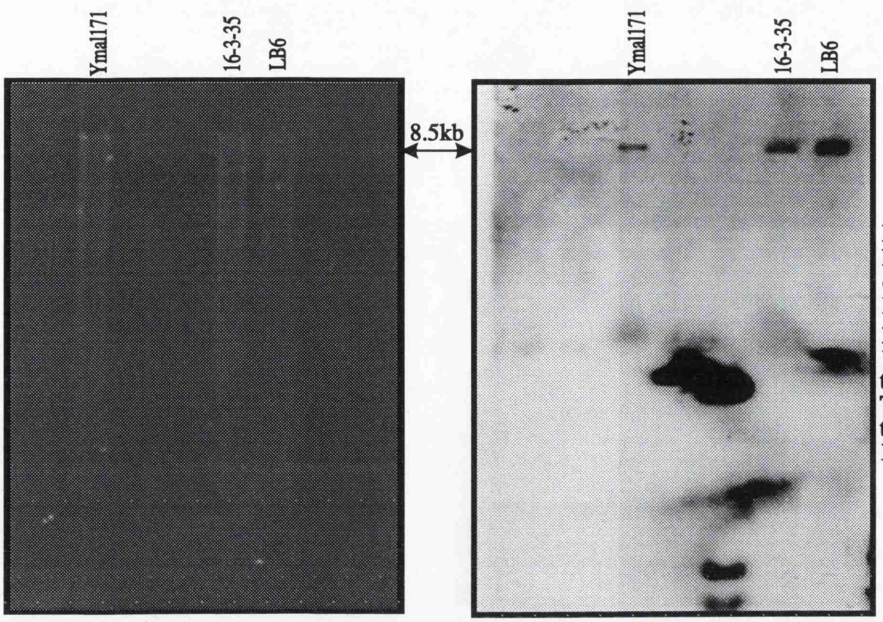


Figure 3.07a.
Ethidium bromide stained gel of *Eco* RI restricted genomic DNAs, and the autorad after probing the Southern blot with a 1.9kb fragment from the distal end of λAGO02. The intensities of 1-1-2 show this fragment is outwith the *Ymal*^{171} duplication.

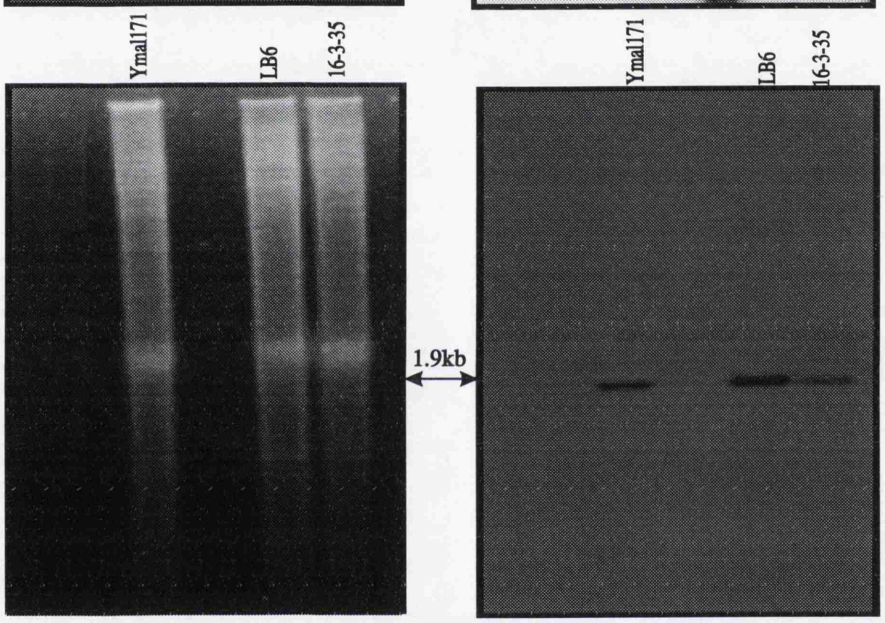


Figure 3.07b.
Ethidium bromide stained gel of *Eco* RI / *Hind* III restricted genomic DNAs, and the autorad after probing the Southern blot with a 1.9kb fragment from the distal end of λAGO02. The intensities of 1-2-1 show this fragment is outwith the *Ymal*^{171} duplication.

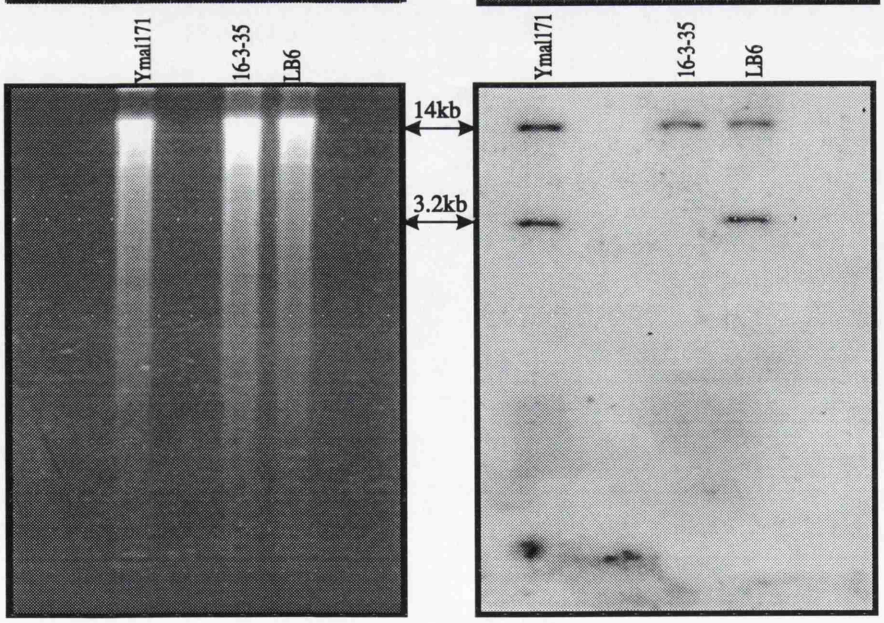


Figure 3.07c.
Ethidium bromide stained gel of *Eco* RI restricted genomic DNAs, and the autorad after probing the Southern blot with a 3.4kb fragment from λC5B03. The polymorphism shows this fragment is within the *Ymal*^{171} duplication.

attached X^Y chromosomes. Thus, as two bands occurred in the *Dp(1)Ymal*^{171} and *Df(1)A118* tracks, (one from the balancer and one from the normal X or attached X^Y chromosomes) and only one occurs in the *Df(1)16-3-35* track (from the balancer chromosome), then this showed this fragment must be within the duplication (Figure 3.07f for summary).

These fragments define an area of 10.8kb within which the *Dp(1)Ymal*^{171} endpoint must lie. This region can be divided into two fragments, one of 6.6kb which contains repetitive DNA and one of 4.0kb which apparently contains single copy DNA (Figure 3.07f for map). A genomic blot of *Dp(1)Ymal*^{171} and wild type *Canton-S* (*CS*) genomic DNAs, restricted with seven different restriction enzymes, was prepared and probed with the 7.4kb fragment from λAGO02 (Figure 3.07d). Several alterations in the *Dp(1)Ymal*^{171} DNA were visible, relative to the wild type DNA. These could be due to the juxtaposition of novel DNA being brought in by the *Ymal*^{171} duplication, however, it is also possible that as more than one band occurs in the altered *Dp(1)Ymal*^{171} tracks this may not be the duplication endpoint but simply a polymorphic region. It is impossible therefore, to state categorically that the *Dp(1)Ymal*^{171} breakpoint lies in this fragment and not within the repetitive fragment lying next to it. This repetitive fragment was also hybridised to the genomic blot described above. The number of bands (presumably due to hybridisation with the repetitive region), made it impossible to determine if the breakpoint was localised here (Figure 3.07e).

3.6 Location of the Proximal End of *Df(1)16-3-35.*

The approximate position of the region containing the *Df(1)16-3-35* proximal breakpoint was located as described (Section 3.3.1), to the genomic clone λ94C11 (Appendix 1 for maps and location). The 3.9kb, 4.1kb and 4.9kb *Eco* RI fragments from this clone were isolated, radiolabelled and hybridised (Sections 2.5.10, 2.5.17 and 2.5.13.2) to genomic blots of *Eco* RI restricted *Df(1)16-3-35, Df(1)LB6* and *Df(1)A118* DNAs. The 4.9kb fragment yielded bands with equal intensities in the *16-3-35* and A118 tracks whereas the intensity of the band in the *Df(1)LB6* track was half that of the *Df(1)A118* signal (Figure 3.08a). The 4.1kb fragment yielded bands with intensities which suggested that this fragment was uncovered by both the *LB6* and *16-3-35* deficiencies, as both had intensities half that of the *Df(1)A118* control DNA (Figure 3.08b). Thus the *Df(1)16-3-35* break-point must lie between these two

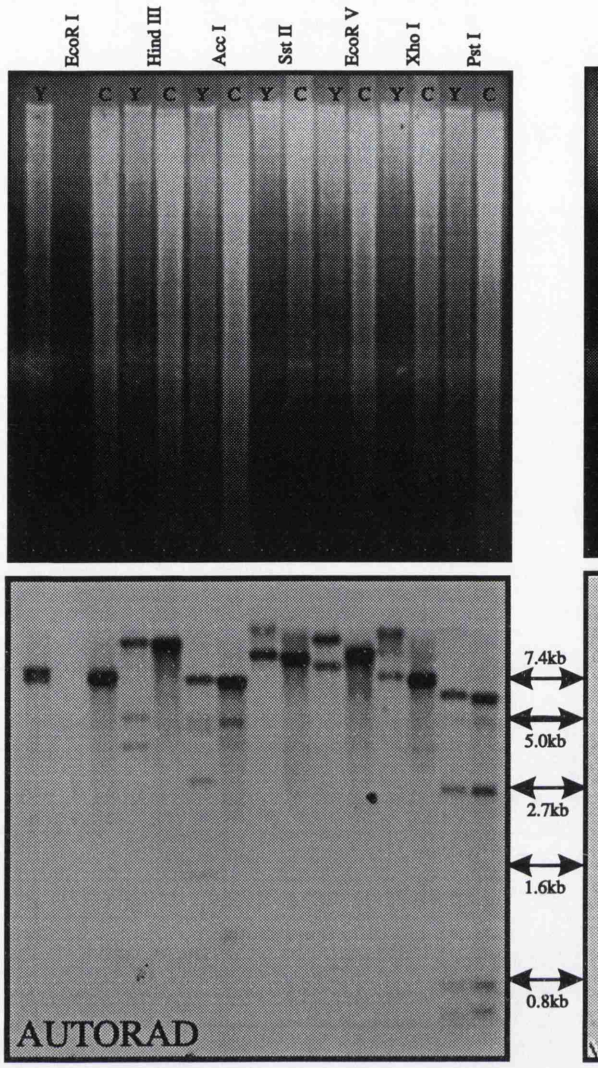


Figure 3.07d-Ethidium bromide stained gel and autorad of a genomic Southern blot of *Cs* (C) and Dp(1)*Ymal*^{171} (Y) DNAs hybridised with a 7.4kb *Eco* RI fragment from λAGO02. See below and text (3.5) for an interpretation of the results.

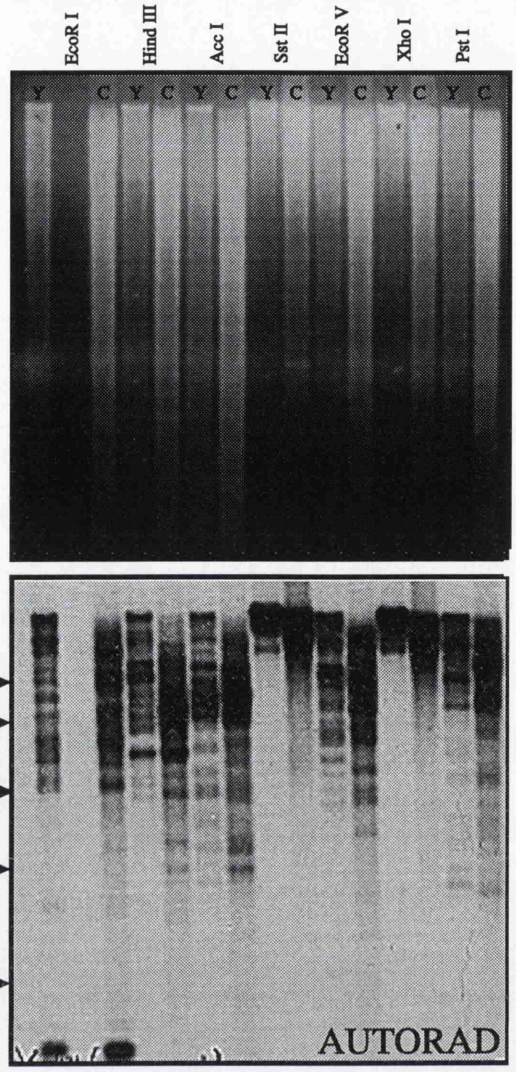


Figure 3.07e-Ethidium bromide stained gel and autorad of a genomic Southern blot of *Cs* (C) and Dp(1)*Ymal*^{171} (Y) DNAs hybridised with a 6.6kb *Eco* RI fragment from λAGO02. See below and text (3.5) for an interpretation of the results.

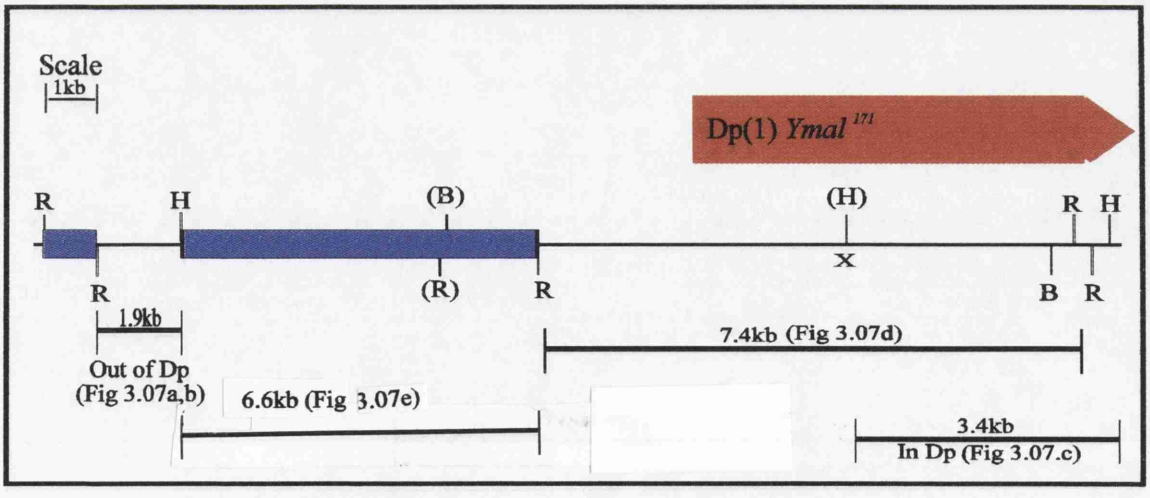
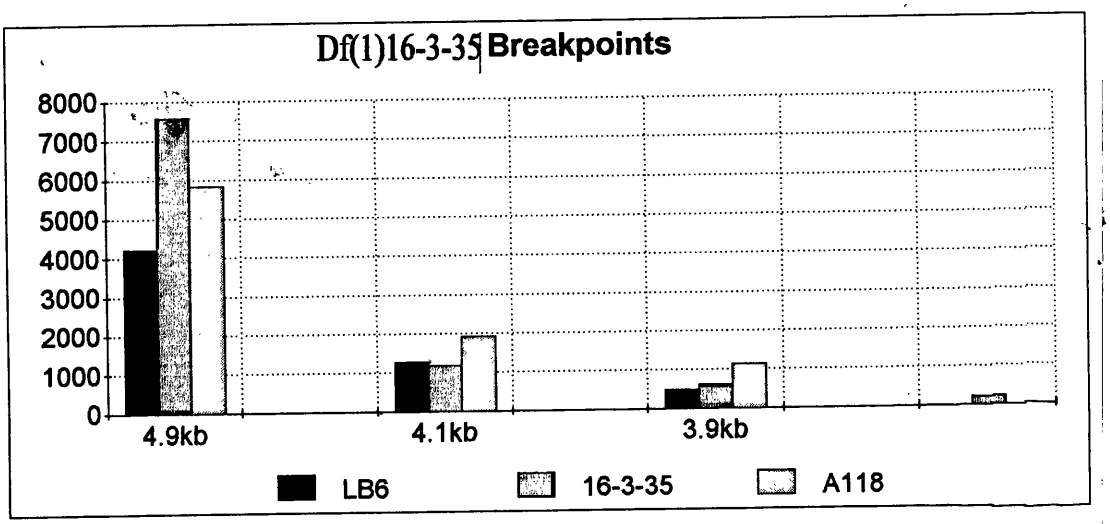


Figure 3.07f-Schematic representation of the region in which the distal endpoint of the *Ymal* ^{171} duplication lies. The bracketed restriction sites are polymorphic. The regions in blue are known to contain opa repeats and so are repetitive. See text (3.5) for full explanation. (Key to enzymes: R= *Eco* RI, B= *Bam* HI, H= *Hin* dIII).

fragments, in an area composed of two *Eco* RI fragments of 1.4kb and 3.9kb (Figure 3.08f).

The 3.9kb fragment was hybridised to a genomic blot of *Eco* RI restricted *Df(1)16-3-35* and *Df(1)A118* DNAs (Figure 3.08c). An extra band of 2.0kb was detected in the *Df(1)16-3-35* track. To confirm the presence of the breakpoint within this region the fragment was hybridised to a genomic blot of wild type (*OrR*) and *Df(1)16-3-35* DNAs restricted with five enzymes. Extra bands were observed in four of the five tracks containing *Df(1)16-3-35* DNA; however, extra high molecular weight bands were also observed in all of the *OrR* tracks (Figure 3.08d). To ensure this was not due to partial digestion of the DNA the experiment was repeated, again using five restriction enzymes, but with *Sal* I replacing *Kpn* I. Extra bands were again observed in all of the *Df(1)16-3-35* tracks but as before extra bands were observed in the *OrR* tracks (Figure 3.08e). When the sizes of the 'extra' bands from the first and second results were compared they were found to be in agreement. It is therefore unlikely that these bands originate from partial digestion of the DNA. Restriction map data for the *Eco* RI and *Hin* dIII sites supported the idea that these bands were due to polymorphic restriction sites in the *OrR* DNA (Figure 3.08f). If the *Hin* dIII and *Eco* RI sites at X are polymorphic then the size of the bands expected, matches the sizes of the bands observed *i.e.* 5.3kb (3.9kb+1.4kb) and 5.6kb (1.4kb+4.6kb) in the *Eco* RI and *Hin* dIII tracks respectively. No restriction sites are present in the cloned region to explain the sizes of the extra bands in the *Df(1)16-3-35* tracks and so these are due to the presence of novel restriction sites bought in by the *16-3-35* deficiency from outwith the region. The 'extra' bands in the *Df(1)16-3-35* DNA are fainter than those in the wild type DNA. This is presumably because the deficiency removes most of the 3.9kb fragment from the deficiency chromosome. The extent of the DNA in the novel restriction fragment homologous to the probe (compared to that from the balancer chromosome which overlap completely) is much less and so a weaker signal is observed. This location has since been confirmed using *in situ* hybridisation to *Df(1)16-3-35* chromosomes (Krishnan *et al.*, 1993).



Df DNA in track / Fragment used as probe	NUMBER OF PIXELS DETECTED PER BAND		
	LB6	**16-3-35**	**A118**
4.9kb	4213	7576	5835
4.1kb	1262	1183	1917
3.9kb	481.8	615.3	1142
		212.6	

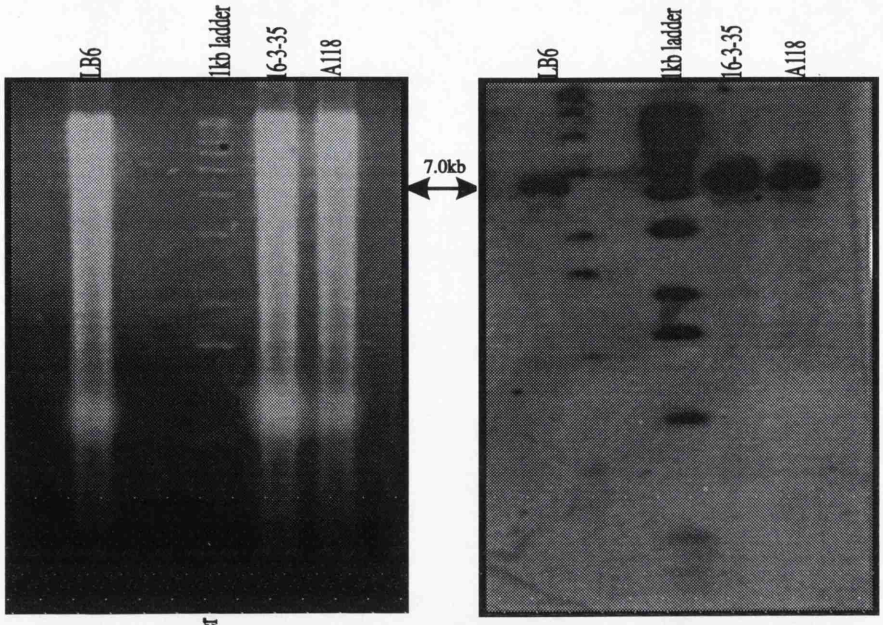


Figure 3.08a -
Ethidium bromide stained gel of *Eco* RI restricted genomic DNAs and the corresponding autoradiograph of the Southern blot, hybridised with the 4.9kb *Eco* RI region from λ94C11. The signal intensities of 1-2-2 show that this area is out of the 16-3-35 deficiency.

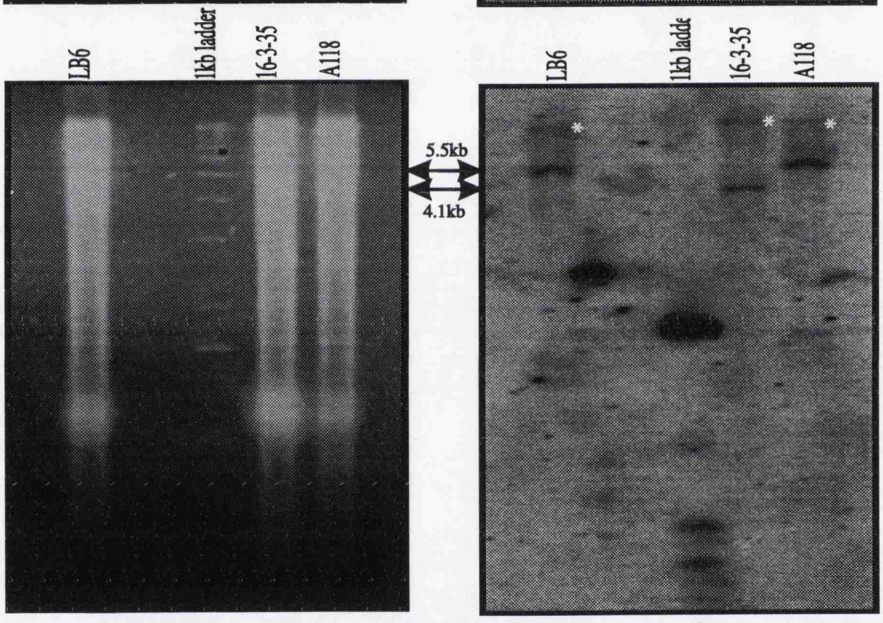


Figure 3.08b -
Ethidium bromide stained gel of *Eco* RI restricted genomic DNAs and the corresponding autoradiograph of the Southern blot, hybridised with the 4.1kb *Eco* RI region from λ94C11. The signal intensities of 1-1-2 show that this area is within the 16-3-35 deficiency. The bands marked by * are undigested material.

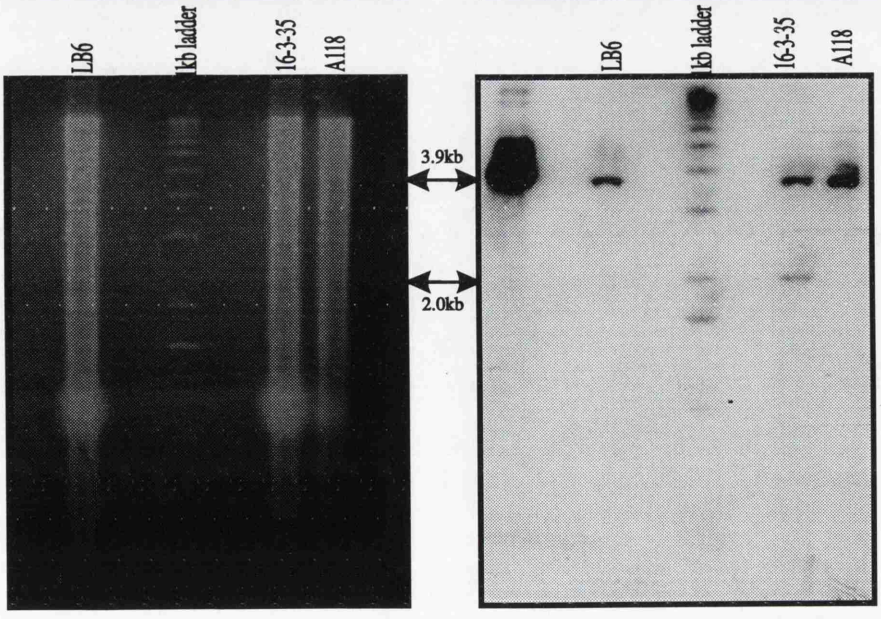


Figure 3.08c -
Ethidium bromide stained gel of *Eco* RI restricted genomic DNAs and the corresponding autoradiograph of the Southern blot, hybridised with the 3.9kb *Eco* RI region from λ94C11. The extra band at 2.0kb is caused by the 16-3-35 deficiency bringing in a novel restriction site.

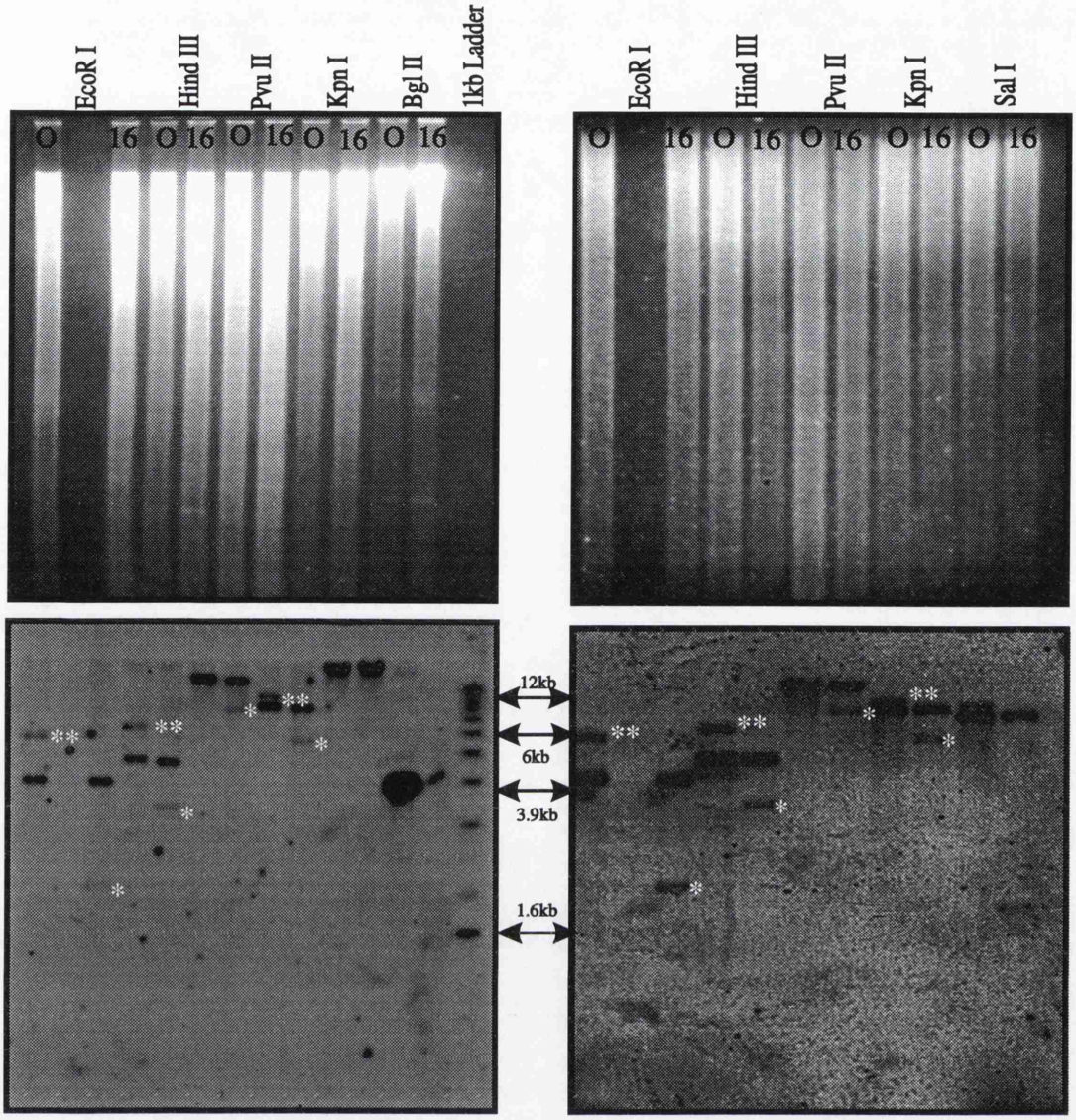


Figure 3.08d-Ethidium bromide stained gel (top) and the autoradiograph (bottom) arising from the hybridising of a 3.9kb *Eco* RI fragment from λ94C11 to a Southern blot of *OrR* (O) and Df(1)*16-3-35* DNAs. See below and text (Section 3.6) for an interpretation of the results.

Figure 3.08e-Ethidium bromide stained gel (top) of a similar Southern blot to that shown in Figure 3.08d. Extra bands of the same molecular weights are visible in the *OrR* tracks of both autoradiographs (bottom images), showing that they originate from polymorphisms in the *OrR* DNA.

(The bands labelled ** are polymorphic restriction fragments whilst those labelled * are caused by the Df(1)*16-3-35* breakpoint).

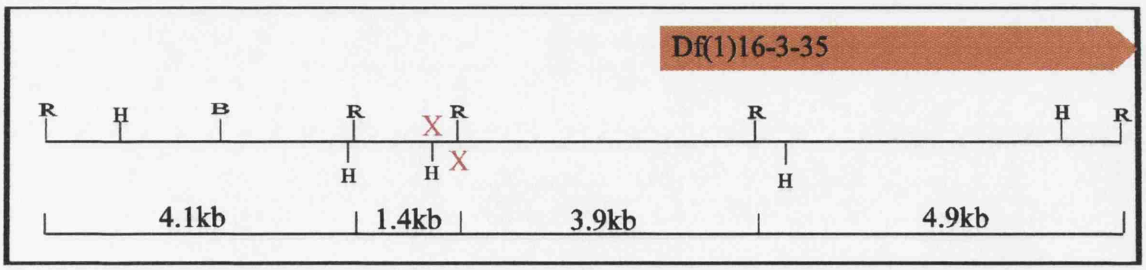


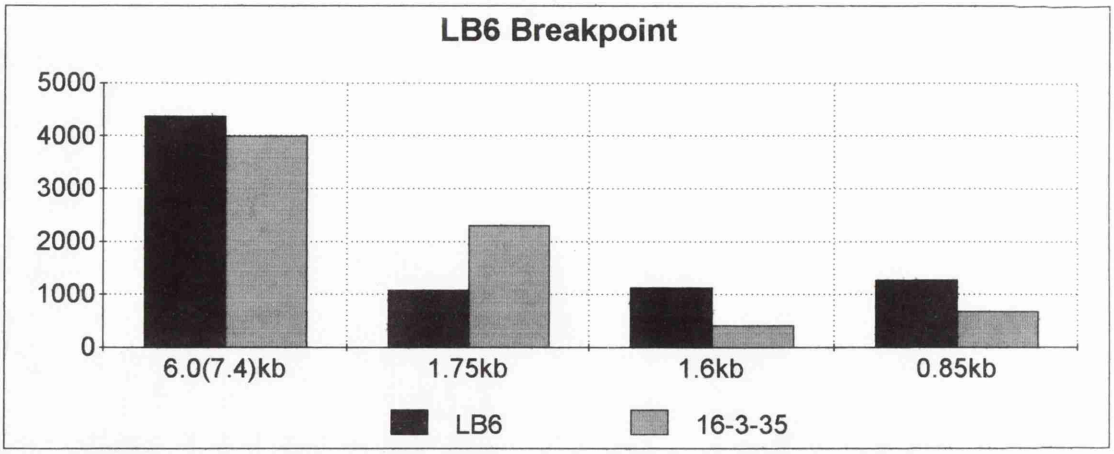
Figure 3.08f- Schematic diagram of the restriction sites and fragment sizes around the *16-3-35* endpoint. X marks two restriction sites that are known to be polymorphic (see Figs 3.08d and e and text [3.06]). This breakpoint is situated in the phage clone λ94C11. (Key to restriction enzymes: R= *Eco* RI, H= *Hin* dIII, B= *Bam* HI).

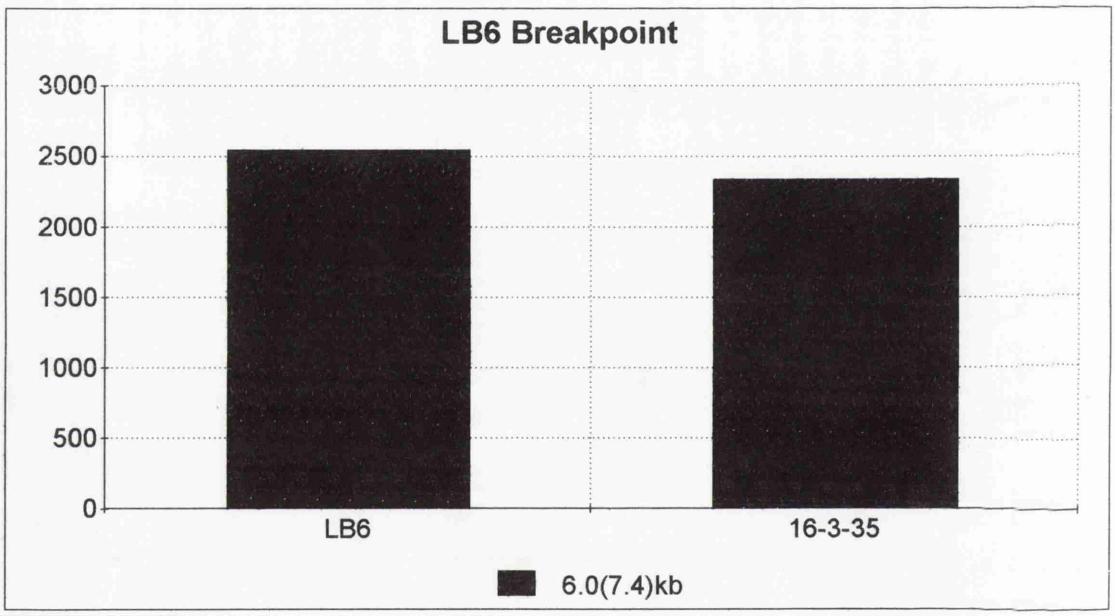
3.7 Location of the Distal End of *Df(1)LB6.*

The region containing the *Df(1)LB6* breakpoint was located as described (Section 3.3.1). It was shown that the distal end of *Df(1)LB6* was situated within the region of the genomic clone λ9405, (Appendix 1 for map and location). Single copy *Eco* R1 fragments from across λ9405 (see Figure 3.09d) were hybridised to a Southern blot of *Eco* RI restricted *Df(1)LB6* and *Df(1)16-3-35* DNAs. The *16-3-35* deficiency was known to uncover this area (Section 3.6), therefore all fragments from this region would give signal intensities half that of wild type DNA in the *Df(1)16-3-35* track, as only one copy of the X-chromosome was present.

Four subclones from across the λ9405 region (see Figure 3.09d for locations) were hybridised to a blot of *Eco* RI restricted *Df(1)LB6* and *Df(1)16-3-35* DNA (Figure 3.09a). The intensities observed indicated that the more distal fragments used as probes (p94.R3, p94.R5 and p94.R8) are all outside of the deficiency, all three giving signal intensities double that of the *Df(1)16-3-35* track. The p94.R2 subclone, from the proximal end of λ9405, was shown to lie within the deficiency as the signal intensity was equal to that of the *Df(1)16-3-35* track. The fragment observed to hybridise to the p94.R2 probe in the *Df(1)LB6* track is 7.4kb and not 6.0kb. This is because the proximal *Eco* RI site of the 6.0kb fragment is polymorphic and is not present in *Df(1)LB6* DNA. Therefore it was concluded that the deficiency breakpoint was situated in the DNA separating the p94.R3 and p94.R2 fragments *i.e.* within a 7.2kb region composed of p94.R1 and p94.R6 (Summarised in Figure 3.11d).

The 6.0kb p94.R1 subclone contained a region of repetitive DNA which precluded the use of the whole fragment as a probe. It was mapped further and shown to contain sites for the restriction enzymes *Pvu*II, *Bgl*II and *Eco*RV. The extent of the repetitive region was then mapped by reverse Southern analysis (see Chapter 4.2.1 for details and results of this). Further mapping (D.Crompton, personal communication) showed the presence of an *Acc*I site 1.5kb into the p94.R1 subclone. This 1.5kb *Eco* R1 - *Acc*I fragment is single copy, and so the fragment was used to probe a genomic blot of *Df(1)LB6* and *Df(1)16-3-35* DNAs. An extra band of >20kb was detected (Figure 3.11b). The high molecular weight band was of lower intensity than expected (it should be equal to the 6.0kb band). This can be the result of several reasons. It is possible that most of the 1.5kb fragment could be within the deficiency *i.e. Df(1)LB6*





Df DNA in track / Fragment used as probe	NUMBER OF PIXELS DETECTED PER BAND	
	LB6	**16-3-35**
6.0(7.4)kb	4366	3994
1.75kb	1076	2293
1.6kb	1118	404.4
0.85kb	1271	669.4

Df DNA in track / Fragment used as probe	NUMBER OF PIXELS DETECTED PER BAND	
	LB6	**16-3-35**
6.0(7.4)kb	2547	2338

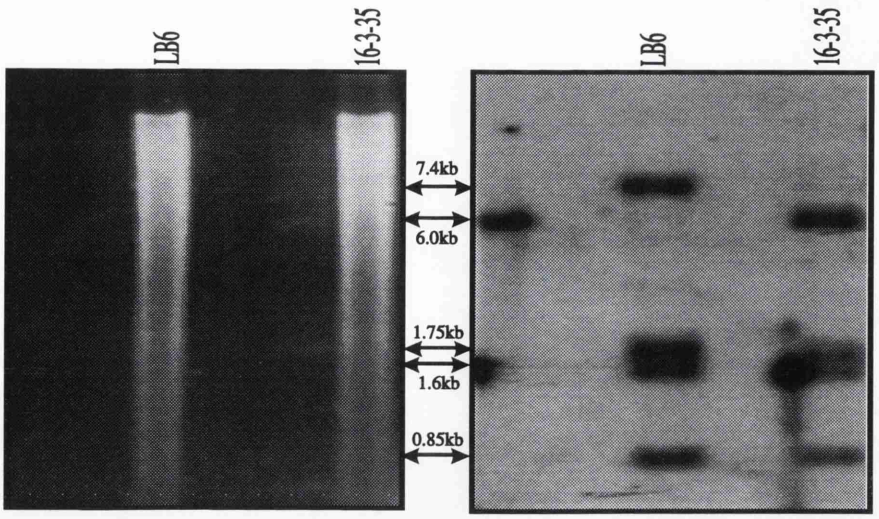


Figure 3.09a-Ethidium bromide stained gel image and Southern blot of *EcoR* I restricted LB6 and 16-3-35 DNAs challenged with various fragments from across λ9405. The more distal fragments are all double the intensity of the 16-3-35 track and thus are out of the deficiency. p94.R2 is in the deficiency and so is equally as intense as the 16-3-35 track.

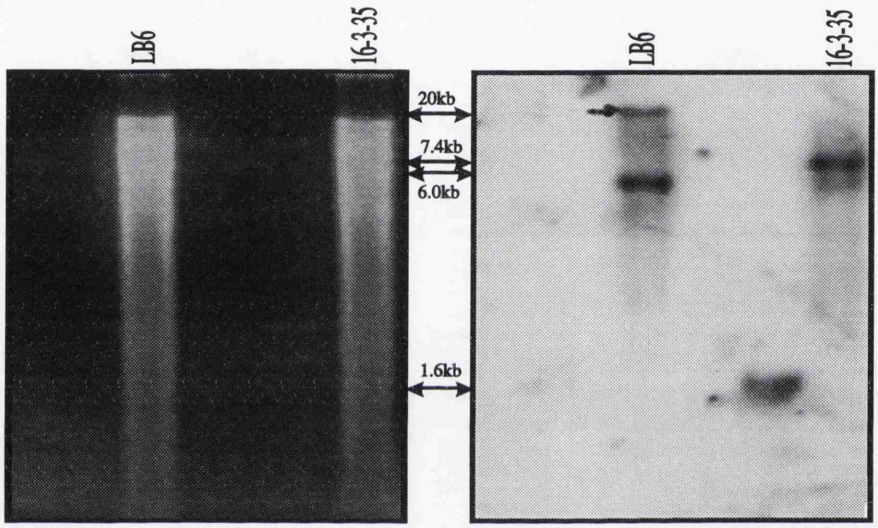


Figure 3.09b-Ethidium bromide stained gel image and Southern blot of *EcoR* I restricted LB6 and 16-3-35 DNAs challenged with a 1.5kb *EcoR* I/*Acc* I site from λ9405. An extra band is visible at ca.20kb. This is due to the precense of the LB6 breakpoint which is bringing in novel restriction sites.

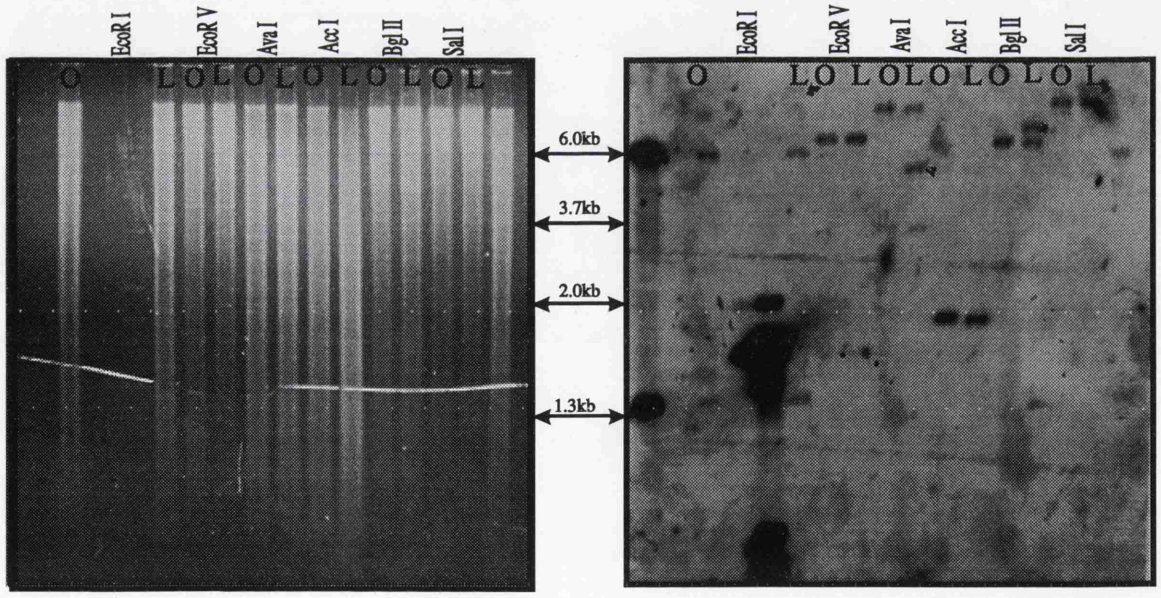


Figure3.09c-Genomic southern blot of LB6 and wild-type (OrR) DNAs restricted with various enzymes and challenged with the 1.5kb EcoR I/Acc I fragment from p94.RI. Extra bands caused by the juxtoposition of novel DNA are visible in 5 out of the 6 tracks. O=OrR DNA, L=LB6 DNA.

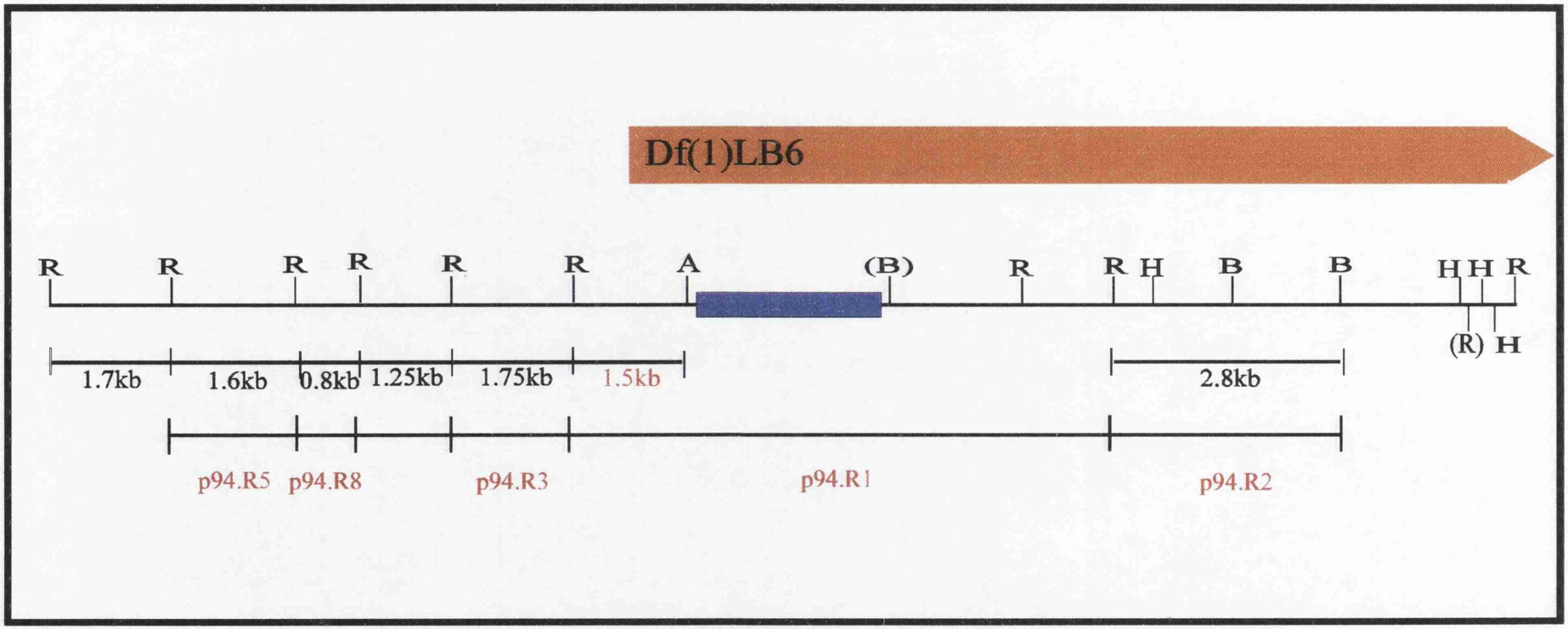


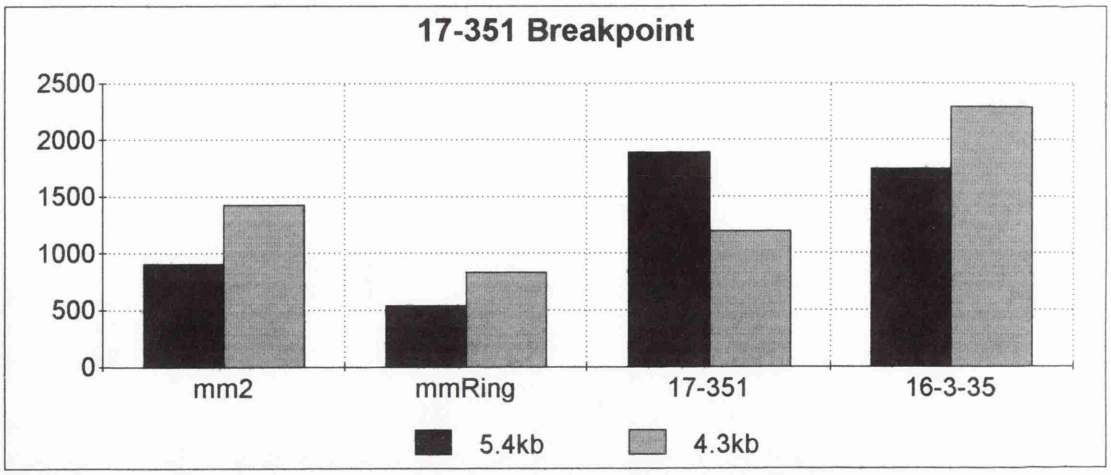
Figure 309-Restriction map of the genomic clone λ9405 in which the distal endpoint of Df(1)LB6 is situated. The regions used for the breakpoint analysis are labelled in red. Repetitive regions are labelled in blue. See text (3.8) for further details. (Key to restriction enzymes: R= *Eco* RI, H= *Hin* dIII, A= *Acc* I, B= *Bam* HI).

breaks very near to the distal *Eco* RI site of p94.R1 so leaving very little DNA to hybridise to the novel fragment. Alternatively the lower intensity may be explained by the lower efficiency with which high molecular weight DNA transfers during Southern blotting. The presence of the distal end of the *LB6* deficiency in this fragment was confirmed by hybridising the 1.5kb fragment to a genomic blot of wild type (*OrR*) and *Df(1)LB6* DNAs restricted with six different restriction enzymes. Extra bands caused by the juxtaposition of novel DNA from the proximal end of the *LB6* deficiency, were detected in five of the six tracks relative to the *OrR* tracks (labelled with asterixes in Figure 3.09c).

3.8 Location of the Distal End of *Df(1)17-3-51.*

The approximate position of the region containing the *Df(1)17-351* breakpoint was located as described (Section 3.3.1), to the genomic clone λ9915 (Appendix 1 for map and location). Two regions, a 4.3kb *Eco* RI fragment, from λ9915, and a 5.4kb *Eco* RI fragment from λH962, (Appendix 1 and 1.2 for locations) were hybridised to a blot of *Eco* RI and *Sal* I restricted genomic DNAs isolated from *Dp(1)mm2*, *Dp(1)mmRing*, *Df(1)17-351* and *Df(1)16-3-35* flies (Figure 3.10a). The *16-3-35* deficiency does not uncover any of this area, and so two copies of this region are expected *i.e.* fragments will give intensities of two in the *Df(1)16-3-35* track.

The 4.3kb *Eco* RI fragment yielded a signal half the intensity of the *Df(1)16-3-35* control DNA showing that this fragment lies within the *17-3-51* deficiency. The 5.4kb *Eco* RI fragment, however, was equally intense as its corresponding signal in the *Df(1)16-3-35* track indicating that it lies outside the deficiency. Therefore, the distal endpoint of *17-3-51* must lie in the 35kb of DNA separating these two fragments. The 1.8kb *Sal* I fragment, from the distal end of the genomic phage clone λ9915, is approximately half way between these two fragments and was therefore hybridised to *Df(1)17-351*, *Df(1)LB6* and *Df(1)16-3-35* DNA restricted with both *Eco* RI and *Sal* I. Instead of yielding the two *Eco* RI - *Sal* I fragments of 0.5kb and 1.3kb expected, the results obtained showed polymorphisms (probably arising because the mutations were isolated in different genetic backgrounds) to exist between all three strains within this region (Figure 3.10b and c for result and interpretation). The *LB6* deficiency which



Df DNA in track / Fragment used as probe	NUMBER OF PIXELS DETECTED PER BAND			
	mm2	**mmRing**	**17-351**	**16-3-35**
5.4kb	908.7	541.1	1892	1740
4.3kb	1427	832.6	1194	2291

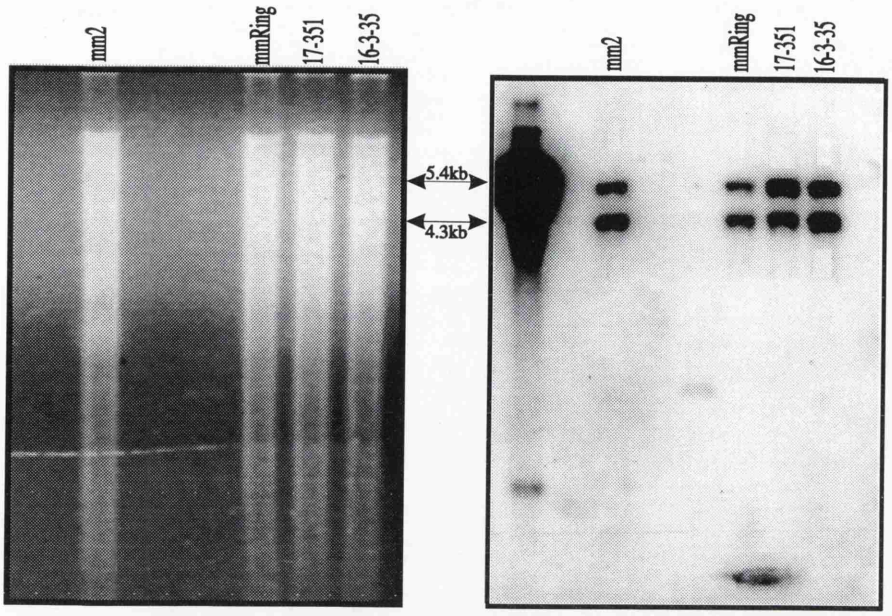


Figure 3.10 a-
Ethidium bromide stained gel and resulting genomic Southern blot of *Eco* RI restricted DNAs challenged a 5.4kb and 4.3kb *Eco* RI fragment from λ9915. See text for an explanation of the results.

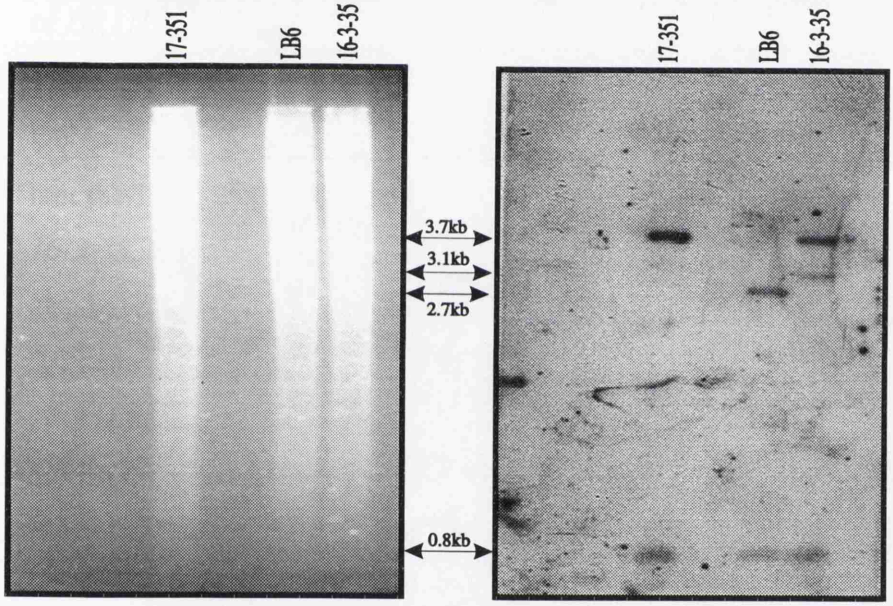


Figure 3.10 b-
Ethidium bromide stained gel and resulting genomic Southern blot of *Eco* RI / *Sal* I restricted DNAs challenged with a 1.8kb *Sal* I fragment from λ9915. See text and Fig 3.10C for further explanations.

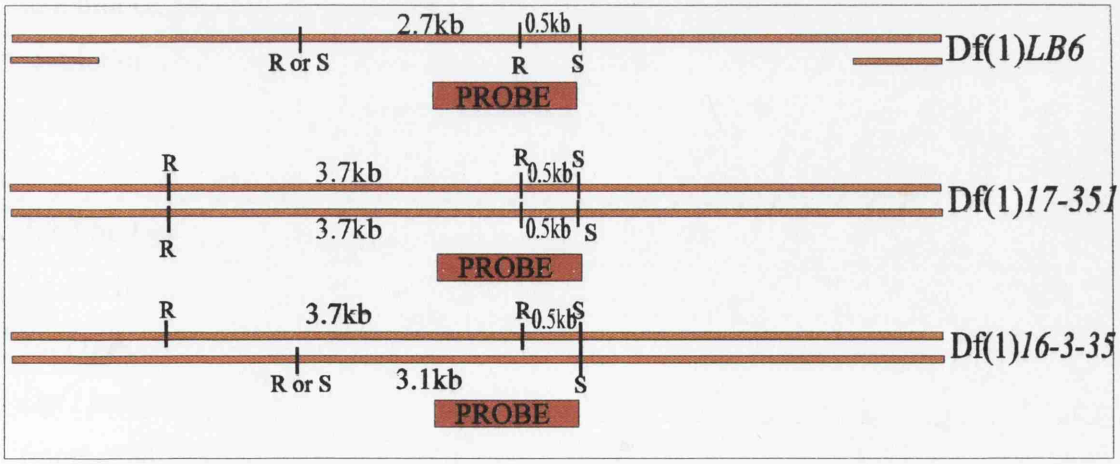
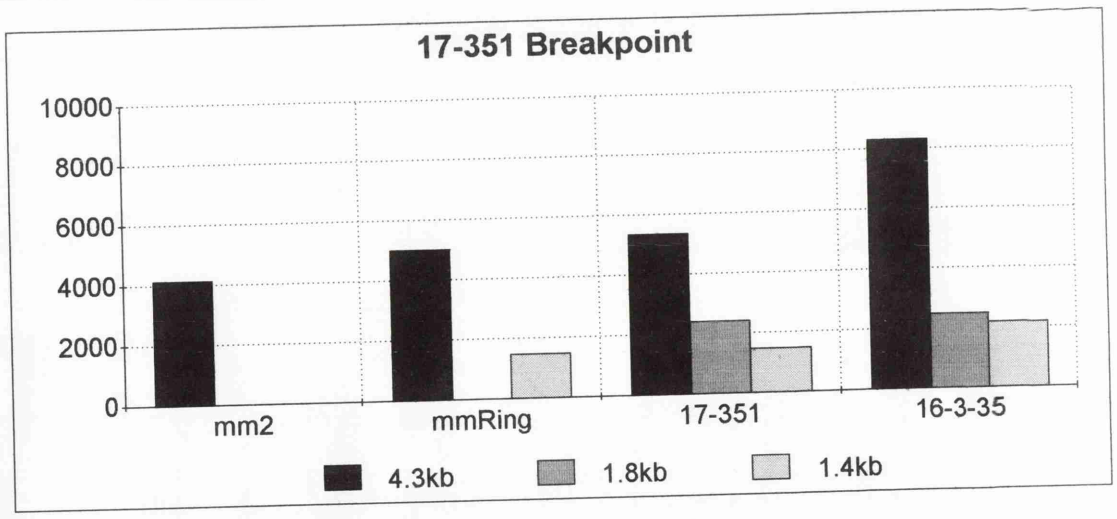


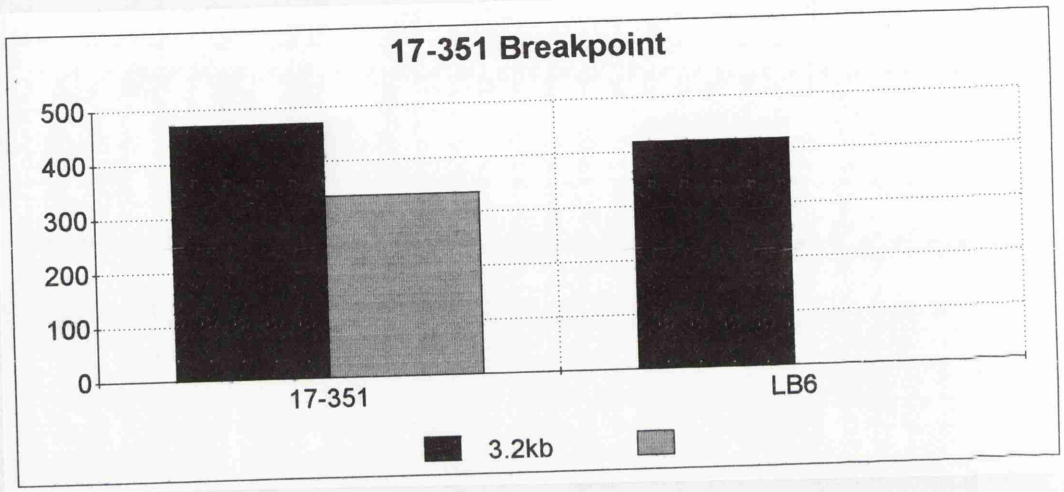
Figure 3.10c-Diagram of the region around the 1.8kb *Sal* I fragment used as a probe in Fig3.10c and an interpretation of the genomic blot showing the differences that exist between the three strains used in the analysis. The results are explained in the text (3.8). (Key to restriction enzymes: R= *Eco* RI and S= *Sal* I).

uncovers this region possesses the 0.5kb *Eco* RI to *Sal* I band with an *Eco* RI or *Sal* I restriction site *ca.* 2.7kb distal to this fragment. The *Df(1)17-351* strain had the 0.5kb *Eco* RI to *Sal* I band but the nearest distal restriction site to this fragment is an *Eco* RI site 3.7kb away. The *Df(1)16-3-35* DNA which does not uncover this region had the 3.7kb and 0.5kb fragments on one chromosome but has a single fragment of *ca.* 3.1kb on the other, presumably made up of the 2.7kb plus the 0.5kb band *i.e.* the *Eco* RI site from the 0.5kb fragment is missing (Figure 3.10c for summary of above). The signals in the *Df(1)17-351* DNA are twice as intense as any of the bands in either of the two control tracks. The *Df(1)16-3-35* signal looks half but is in fact equal in intensity to the signal in the *Df(1)17-351* DNA. The phenomenon is observed because the signal is split due to the polymorphisms on the balancer and deficiency chromosomes. This shows this fragment is again outwith the *Df(1)17-351* deficiency. The distal endpoint of *Df(1)17-351* must therefore be located within the 10kb of DNA separating the 1.8kb and 4.3kb fragments in λ9915. *Dp(1)mmRing*, *Df(1)17-351* and *Df(1)16-3-35* DNAs. A polymorphism present in both the *16-3-35* and the *17-351* DNAs showed that the 1.4kb fragment was outwith the *17-3-51* deficiency (Figure 3.10d). Both the *16-3-35* and *17-351* strains have one chromosome with the 1.4kb *Sal* I fragment, both also have a second copy of the chromosome present but on this chromosome the distal *Sal* I site is absent the next *Eco* RI or *Sal* I site being some 0.5kb away. This meant the *17-351* breakpoint must lie in either the 0.4kb *Eco* RI to *Sal* I fragment or the 3.2kb *Sal* I fragment. The 3.2kb fragment was hybridised to a genomic blot of *Sal* I restricted *Df(1)17-351*, *Df(1)LB6* and *Df(1)16-3-35* DNAs and an extra band of *ca.* 1.3kb was observed in the *17-3-51* DNA (Figure 3.10d). This fragment was hybridised to a blot of wild-type (*OrR*) and *Df(1)17-351* DNAs restricted with seven different restriction enzymes. With all seven different enzymes extra bands were observed (Figure 3.10f); thus the *Df(1)17-351* breakpoint lies within this fragment.

3.9 The *Dp(1)mm2* and *Dp(1)mmRing* Breakpoints.

As can be seen in Figures 3.11a and Figures 3.11d, the 952 walk fragments used in the *Df(1)17-351* localisation suggested that they were not covered by the *Dp(1)mm2* and *Dp(1)mmRing* duplications. When the genomic blots were rehybridised with fragments from the 896 walk (Figure 3.12) the hybridisation signals were very weak (presumably due their frequent use in hybridisations for localising the *Df(1)17-351*





Df DNA in track / Fragment used as probe	NUMBER OF PIXELS DETECTED PER BAND			
	mm2	**mmRing**	**17-351**	**16-3-35**
4.3kb	4123	4994	5341	8287
1.8kb			2397	2472
1.4kb		1439	1459	2163

Df DNA in track / Fragment used as probe	NUMBER OF PIXELS DETECTED PER BAND	
	17-351	**LB6**
3.2kb	470.4	419.3
1.0kb	334.8	------

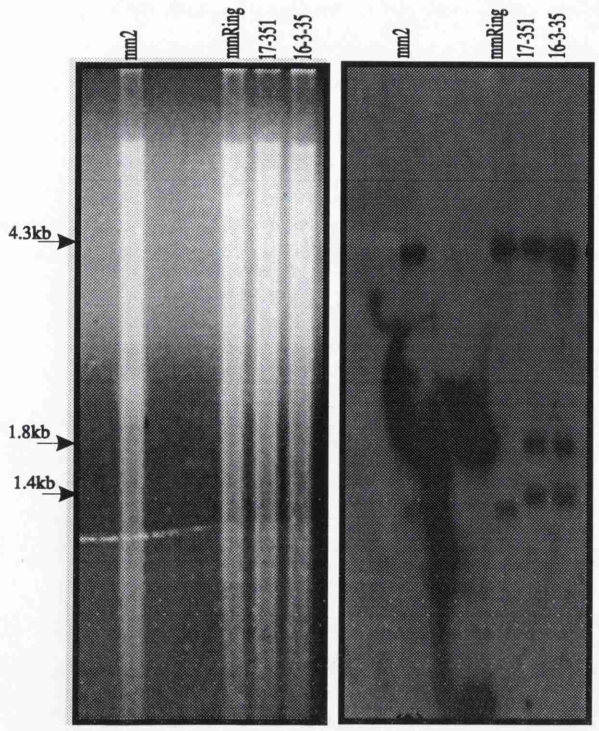


Figure 3.10d-Ethidium bromide stained gel and corresponding autoradiograph of *Eco* RI / *Sal* I restricted genomic DNAs, Southern blotted and challenged with the 1.8kb *Sal* I and 4.3kb *Eco* RI fragments from λ9915. See text (3.8) for explanation.

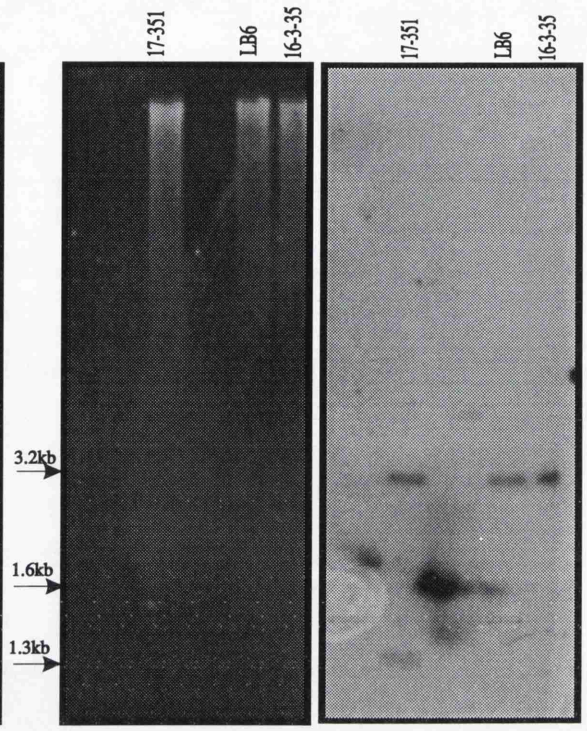


Figure 3.10e-Ethidium bromide stained gel and corresponding autoradiograph of DNAs restricted with *Sal* I and challenged with the 3.2kb *Sal* I fragment from λ9915. An extra band is visible at *ca.*1kb.

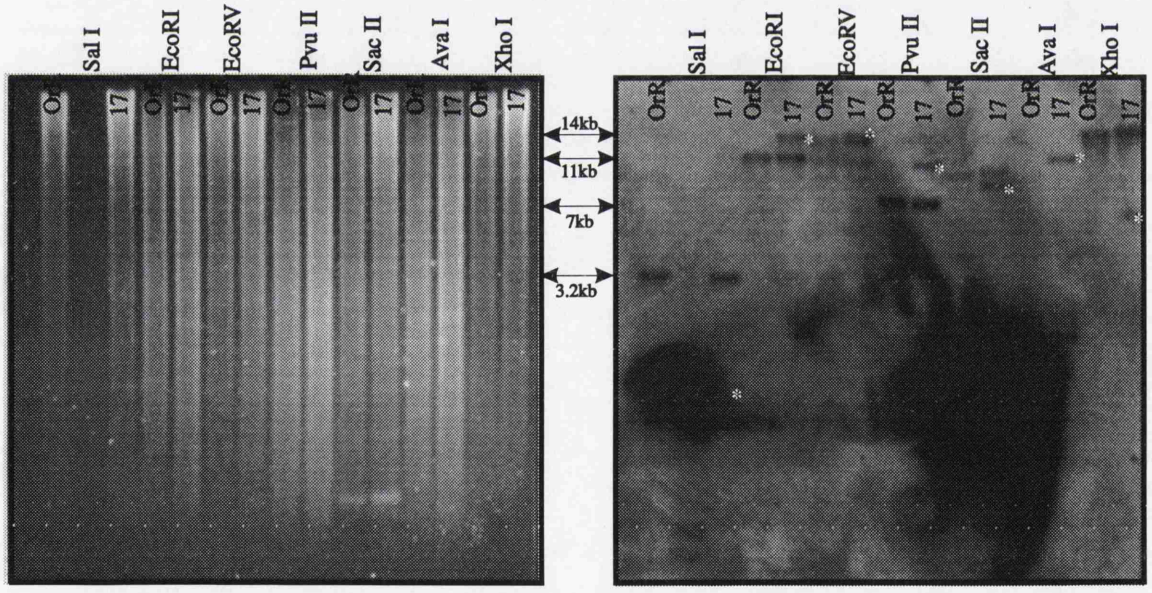


Figure 3.10f-Ethidium bromide stained gel and corresponding autoradiograph of *OrR* (O) and Df(1)*17-351* (17) DNAs restricted with several different restriction enzymes, Southern blotted and then hybridised with a 3.2kb *Sal* I fragment from λ9915. Extra bands are visible in all seven tracks showing the presence of novel DNA brought into the region by the endpoint of the *17-351* deficiency (indicated by asterixes). Background hybridisation to the filter is visible but does not obscure the restriction fragments caused by the deficiency.

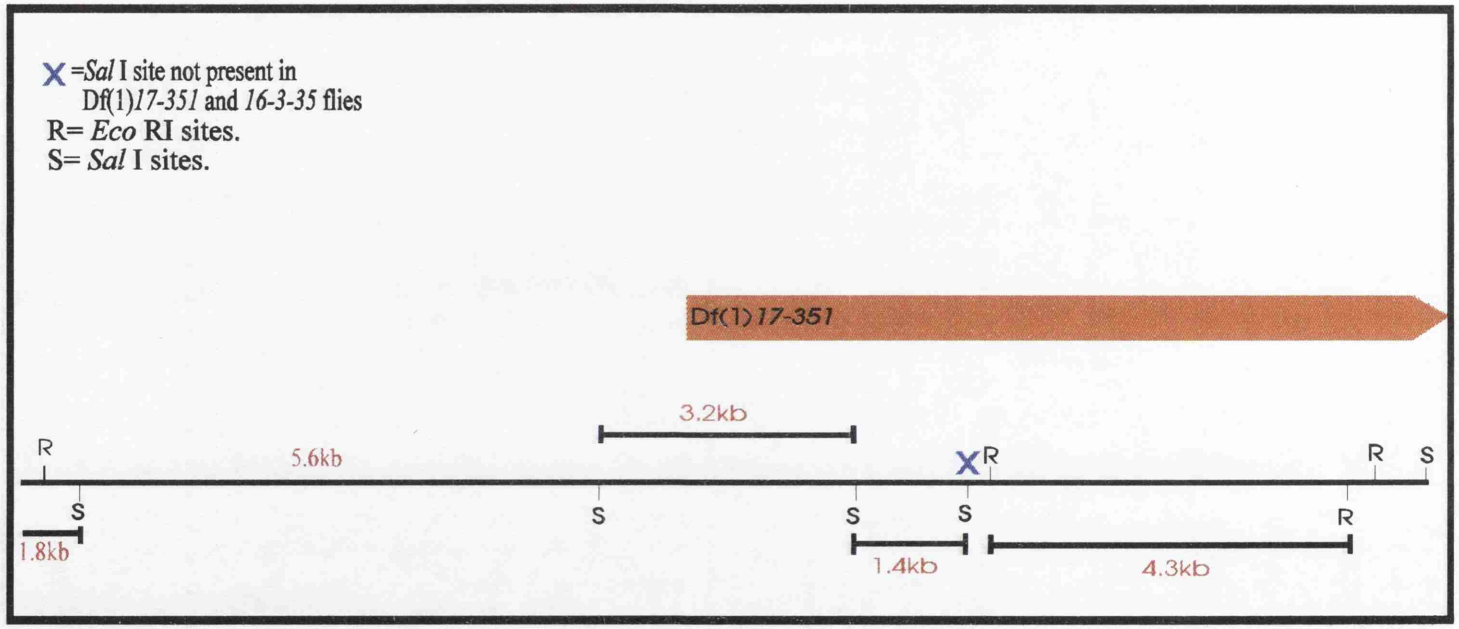


Figure 3.10g-Schematic representation of the region in which the Df(1)*17-351* breakpoint lies. Fragments used in the breakpoint analysis are highlighted red. See text (3.8) and figures for further explanation.

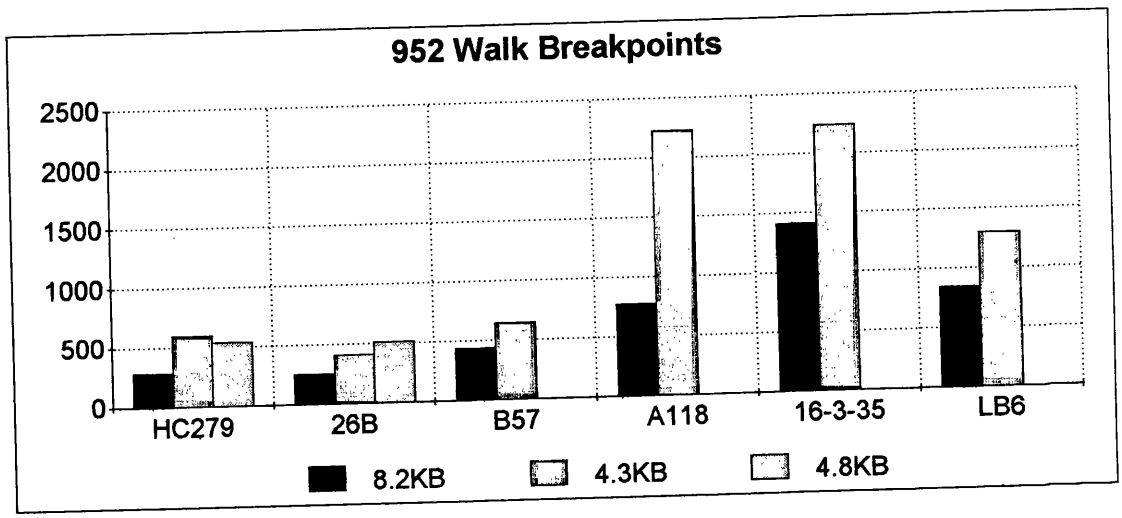
breakpoint causing the DNA to become dislodged). The results however suggested that the 896 walk was not covered by the duplications. As mentioned previously (Section 3.3.2) in the intervening time between these initial experiments and when the duplications were looked at in more detail, the fly stocks maintained at Glasgow had lost their mini-ring chromosome. As no more DNA was available for repeating the experiments and as problems were encounterd with new stocks, time constraints made it impossible to characterise these breakpoints further.

3.10 Location of Breakpoints Between Walks.

A blot of equally loaded *Eco* RI restricted genomic DNAs, prepared from *Df(1)HC279*, *Df(1)26B*, *Df(1)B57*, *Df(1)A118*, *Df(1)16-3-35* and *Df(1)LB6* flies, was probed with a 4.3kb *Eco* RI fragment from λ9915 (at the most proximal end of the 952 walk) and a 3.5kb *Ava* I fragment from λH683, the most distal fragment from the genomic clone at the distal end of the 896 walk (Appendices 1 and 2 for locations of phages).

The result obtained with the 4.3kb *Eco* RI fragment (Figure 3.11a) demonstrated the presence of a polymorphic *Eco* RI site on both the *Df(1)HC279* and *Df(1)26B* chromosomes. This is probably the distal *Eco* RI site of the fragment (see Figure 3.11b). The other chromosome in both strains had the 4.3kb band observed in the original genomic clones. Thus there are two copies of the chromosome at this region *i.e.* this region is not uncovered by these two deficiencies. The intensity of the signal in the *Df(1)A118* track indicated that this deficiency does not uncover this region either. This was because the signal intensity was equal to control deficiencies known not to uncover the region *i.e. Df(1)16-3-35* and double the intensity of deficiencies known to uncover this region *i.e. Df(1)LB6* and *Df(1)B57*.

The signal observed with the 3.5kb *Ava* I fragment showed all three deficiencies discussed above to uncover the fragment, as the signal intensity was half that of the *Df(1)16-3-35* signal and equal to the other deficiency DNAs which do uncover this region. Thus the distal endpoints of the three deficiencies under investigation were shown to lie outside regions at present cloned (Section 3.2.2), between the 896 and 952 walks (see Figure 3.11b for a summary of the above).



Df DNA in track / Fragment used as probe	NUMBER OF PIXELS DETECTED PER BAND					
	HC279	**26B**	**B57**	**A118**	**16-3-35**	**LB6**
8.2kb	285.6	251.1	432.9	772.1	1410	846.8
4.3kb	597.0	408.8	645.3	2224	2233	1299
4.8kb	539.5	511.7				

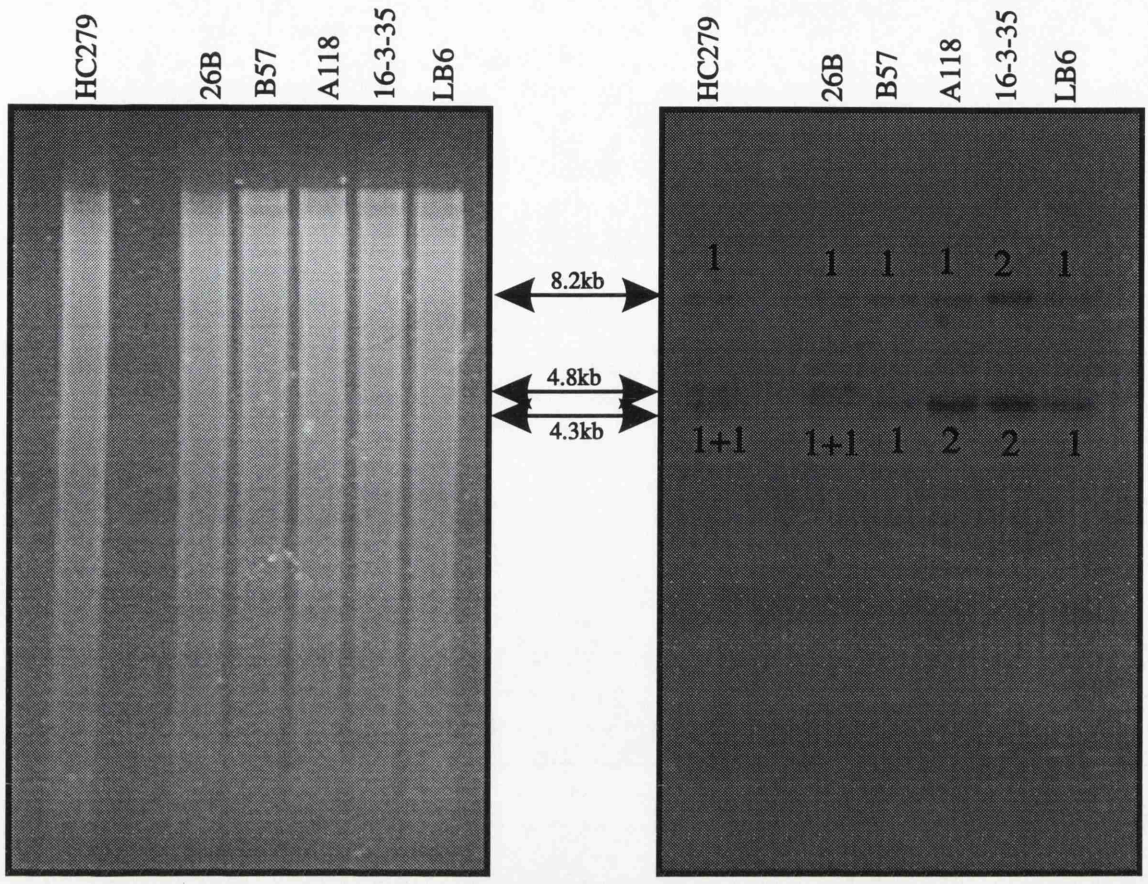


Figure 3.11a-Ethidium bromide stained gel and corresponding autoradiograph of *EcoR* I restricted deficiency DNAs, Southern blotted and then challenged with a 4.3 *Eco* RI fragment from λ9915 and a 3.5kb *Ava* I fragment from λH683. See text (3.10) and Figure 3.11b for an interpretation of the results. The bands are labelled with their relative intensities.

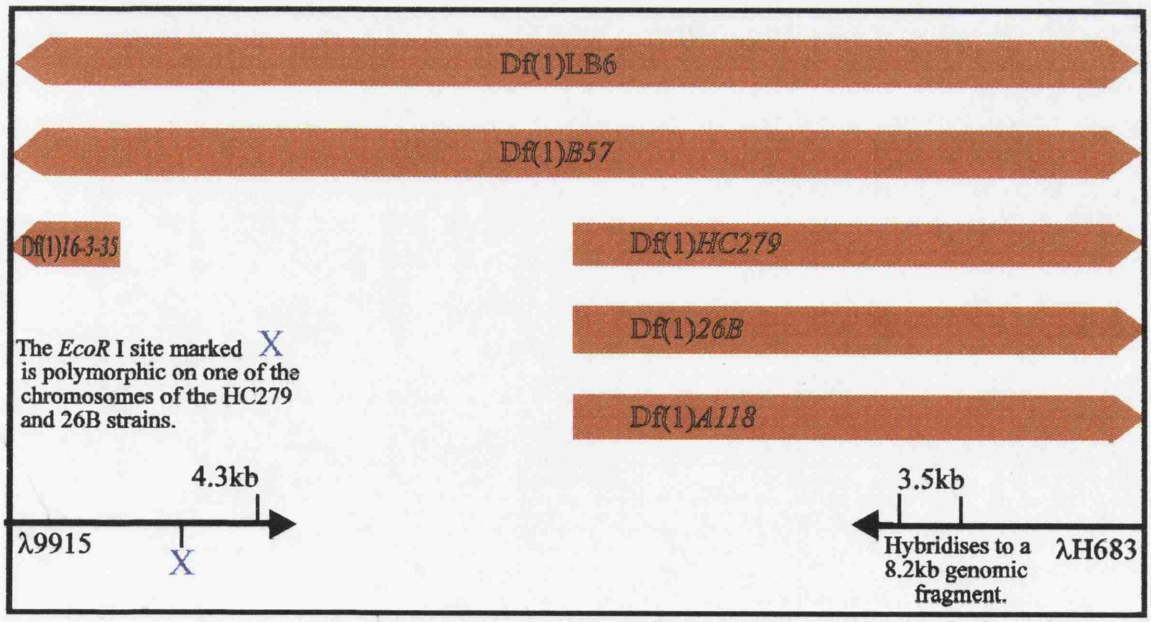


Figure 3.11b-Schematic representation of the results obtained in figure 3.11a. The *A118, HC279* and *26B* deficiencies all have their distal breakpoints between the two fragments used for the analysis and so lie between the 896 and 952 walks as they stood at this time.

3.11 Discussion.

3.11.1 Chromosomal Walking.

The nature of the problem which prevented the isolation of phage clones from the proximal end of the 952 walk is unknown. Several other studies however, have also reported difficulties in the cloning of DNA whilst chromosomal walking in this region at the base of the X chromosome. For example in the *su(f)* locus (Langley *et al.,* 1993; Mitchelson *et al.,* 1993); in the *unc* locus (HHealy *et al.,* 1988), and in the *B214* locus (Russell *et al.,* 1992). In *l(1)B124* in the 19F1-F2 region walking stopped 'despite extensive screening of three different libraries' (Russell *et al.,* 1992). To date, six phage clones have been isolated using fragments from the walks themselves or from fragments isolated from cosmids that extended the walks (S. Ji and this study). Although all the phage and cosmid clones shared regions of homology, as shown by hybridisation studies (S. Ji, personal communication and this study, Section 3.2.2), the restriction maps of the hybridising regions differed. The explanation for this was thought to be an insertion of a transposable element into the DNA from which the λS series phages were isolated. This does not explain however why two genomic libraries were screened with the distal 4.3kb *Eco* RI fragment from λ9915 (which is not repetitive) without success. One possible cause may be the repetitive nature of the DNA into which the insertion occurred as it is known that in *E. coli* host cells, homologous recombination can occur between repetitive DNA. It may have been that recombination between tandemly repeated region(s), leading to elimination, and therefore under-representaion of the sequences including the 4.3kb λ9915 sequence occurred in the two genomic libraries screened. This type of effect has been observed previously. For example when Lohe and Brutlag (1986) attempted to isolate and study repetitive DNAs from *Drosophila* they found that when isolating large plasmid clones containing satellite DNA through several rounds of purification in *E. coli*, 95% of the genomic inserts were deleted. Upon studying this effect they described the creation of inversions, duplications, deletions and a gradual shrinkage of the extent of the satellite DNA region with a concomitant alteration in restriction pattern.

Recombination events between homologous sequences are also known to occur in *in-vivo* within *Drosophila,*. These have been observed in internally repetitive genes, as unequal crossing over and / or slippage events during DNA replication, *e.g.* the *period, salivary gland secretion,* and *dec-1* genes (Costa *et al.,* 1991; Muskavitch and

Hogness, 1982; Andersson and Lambertsson, 1993). Recombination events have also been described in non-transcribed regions, for example the *Antp*^{37b} mutation has *Doc* elements at both ends of the inversion (Schneuwly *et al.,* 1987). Recent work also suggests that recombination between homologous sequences may lead indirectly to the elimination of chromosomal DNA (Spradling *et al.,* 1993). It has been suggested that such events may be responsible for such chromosomal features as under-representation of heterochromatic sequences in *Drosophila* polytene chromosomes (Karpen and Spradling, 1990) and in addition may be responsible for the creation of the chromocenter and other features such as ectopic fibres (Spradling *et al.,* 1993). It is suggested that this mechanism acts primarily on little studied repetitive DNA species. Heterogeneity (*i.e.* creation of chromosomal abnormalities) may be exaggerated near junctions, between centromeric heterochromatin and euchromatin (such as the 19D - 20F region), since *cis*-acting sequences that are presumed to control elimination may no longer be able to function properly (Spradling *et al.,* 1993). A mechanism of this type could cause differences between strains and although it would not interfere with cloning directly, if the region used as a probe was deleted in the strain used to construct the genomic libray from, it would be impossible to isolate genomic clones from this library.

Although these *in vivo* mechanisms are known to exist in *Drosophila* it is much more likely that the problems observed in the cloning of the DNA, at the proximal end of the 952 walk, is a result of artefacts created by the *E. coli* hosts. There are several options available for overcoming these difficulties. An unamplified (*i.e.* primary) genomic library may be constructed and used to isolate clones from the region using *rec*$^-$ strains of bacteria (*e.g.* (See Stratagene Catalogue). limiting the number of rounds of purification to the minimum possible to stop or limit recombination between repetitive regions. However, probably the easiest solution would be to use non-repetitive fragments from clones in the region to isolate sequences from outwith the area of the DNA that is under-represented in genomic phage libraries. This approach has since been used to isolate a genomic clone, λFD3, that contains sequences from λ9915 and which overlaps with the more proximal regions of the λS series phages and with the c9A1 cosmid (S. Ji and J Davies personal communication). It does not contain the region inserted in two of the S series clones (see Appendix 1.2 for summary).

3.11.2 Restriction Site Polymorphisms.

A large number of polymorphic restriction sites were encountered across all four walks between the different strains used in this study. Previous work suggests that a polymorphism *i.e.* a change in any one nucleotide, will occur at a probability of around 1 base pair every 200bp of DNA, even between closely related strains of *Drosophila melanogaster* (Miyashita and Langley, 1988). As this study was not concerned with the identification of the polymorphisms present in the walks, no definite comparisons could be drawn between the results obtained and other, more detailed studies carried out in *Drosophila.* However, several interesting observations can be made. Using the 952 walk as an example, in a region of 180kb of DNA a total of 113 sites were identified for four, hexanucleotide recognising restriction enzymes. Of these, 15% were polymorphic, *i.e.* present in one phage clone but not in another overlapping clone isolated from the same or a different library. Previous studies showed 34% of the hexanucleotide restriction enzyme recognition sites within 45kb of DNA from the *white* locus were polymorphic (Miyashita and Langley, 1988), whereas only 11% of similar sites were polymorphic in 13kb of DNA from the *Zw* locus (Miyashita, 1990). In the 6.4kb of DNA from the *suppresser of forked* (*su(f)*) locus (Langley *et al.,* 1993) none at all were polymorphic.

This figure of 17% is probably an underestimate of the actual number of polymorphic restriction sites present in the 19E3 region, as only three (sometimes four) restriction endonucleases were used to map and characterise most of the walk phages in the 952 walk, whereas the studies mentioned used seven restriction endonucleases. Also a large percentage of the fragments used as probes in the breakpoint localisation exhibited polymorphisms; indeed the presence of these polymorphisms often aided the analysis (*e.g.* Figure 3.11a). It is impossible to compare the occurrence of polymorphisms detected in this study with those mentioned above as the 64 different X chromosome lines investigated by Miyashita and Langley, (1988); Miyashita, (1990) and Langley *et al.,* (1993), were isogenic; and unfortunately the *CS* and *OrR* libraries used in this study were not (Pirotta *et al.,* 1983). When studying the *white* locus and 3B1 - 3C2 region of the X chromosome, Pirotta *et al.,* describe restriction site polymorphisms between homologous genomic clones from the *CS* and *OrR* strains. They also describe the isolation of several insertions in genomic clones from the same *OrR Heidleberg* library, showing that the flies used to construct the library

were not a pure strain (t has been shown that transposable element number is relatively stable within isochromosomal strains, Eggleston *et al.,* 1988). It is impossible to estimate just how many sub-strains existed in the fly population used to construct the libraries and therefore calculate just how many X chromosomes were available for detection in this study

It was observed that the majority of the fragments used as probes in the deficiency and duplication breakpoint analysis, had polymorphic restriction sites near to them; however, it is probable that this reflects the fact that more fly strains (*i.e.* different deficiency and duplication chromosomes) were analysed in these regions. Very few of

Very few polymorphisms were identified in overlapping phage clones that occurred in regions containing the deficiency or duplication endpoints. There seems to be no obvious clustering of the polymorphisms in any particular region of the walk including non-transcribed regions compared to transcribed regions.

Only two insertions were identified in the 180kb of DNA that makes up the 952 walk. These are a 1.2kb insertion in a clone containing DNA also found in λ9405, and the insertion in the λS series phages. The 1.2kb insertion is not repetitive (*i.e.* it gives single bands on a genomic blot) and is not transcribed at any stage of development as shown by the reverse Northern analysis. The λS insertion is repetitive (Section 4.2.1) but it is not known whether it is transcribed. It is possible that other insertions / deletions *etc.* occur in the walks. However, unless the size difference between two restriction fragments is greater than 400bp it is unlikely to be observed by the comparison of restriction fragments on a gel and would not be detected by this study.

A more detailed study of the occurrence of polymorphisms and other gross DNA changes (such as insertions and deletions) across the 19E3 region (952 walk), could prove to be very interesting. To date there has been work on only a handful of regions *e.g. white* (Miyashita and Langley, 1988); *Adh* (*e.g.* Kreitman and Hudson, 1991;

Laurie *et al.*, 1991); the male accessory gland gene *Mst26A* (Aguade *et al.*, 1992) and *suppresser of forked* locus (Langley *et al.*, 1993). Thus relative to other areas of evolutionary biology very little information is currently available on the number and type of nucleotide changes that occur in coding and non-coding DNA.

The complex nature of the *shak-B* locus, with its large number of splice variants and long distance interactions (*e.g.* the interactions that have been shown to occur at the genetic level between 19E3 and 19E5 / 6, makes it an ideal model system for determining how polymorphisms have occurred and altered the region in different strains. In addition it would be possible to see what effects these alterations have had on the organism *e.g.* there is a large difference between the *CS* and *OrR* strains in the extent if their GF responses and morphologies even though at the cytological level the 19E3 region is identical between the two strains. How the polymorhphisms between the two strains affect the *shak-B* locus would aid the long term understanding of the effects of polymorphic changes on gene structure and function, specifically on the *shak-B* locus, and more generally in *Drosophila melanogaster. Drosophila* has proved itself an ideal model system for undertaking studies of this nature. This was demonstrated by studies on the 'hitchhiking' effect (Kaplan *et al.*, 1989), a theory developed to explain the low incidence of polymorphisms in regions with low recombination frequencies. Work in *Drosophila* has demonstrated the validity of the theory and it is now widely accepted as being a plausible model (*e.g.* Stephan and Mitchell, 1992; Kliman and Hey, 1993a+b; Ohta, 1993; Eanes *et al.*, 1993). A study of this nature may also yield valuable information as to how these mechanisms function in other organisms, including humans.

3.11.3 Breakpoint Analysis.

The finding of the distal end of *Df(1) GA40* and *Df(1) N77* and the proximal end of *Df(1) mal*^{10} was important as this allowed the localisation of the 798 walk to 19E1, and suggests that it contains the *mell* locus (Schalet and Lefevre, 1976; Perrimon *et al.,* 1989), but that it was away from the region of interest to this study at 19E3.

mell has only been defined in females, *trans* heterozygous for the two overlapping deficiencies *Df(1)mal*^{10} and *Df(1)N77.* has only been where they have a darker body colour than normal (especially the thorax). Additionally, the abdominal tergites have slightly transverse wrinkles (Schalet and Lefevre, 1976). It was not possible using genetic techniques to tell whether *mell* is a separate locus or whether the phenotype is caused by a the chromosomal breakpoints encroaching upon existing gene landscapes (Perrimon *et al.,* 1989b). Now that this study has determined the exact molecular location of the *mell* region it should be possible to confirm or deny its status as a *bone fide* gene.

Of the breakpoints found in the 952 walk the location of the distal end of *Dp(1)Ymal*^{171} was one of the most important as it has been shown genetically (Section 1.2.5), to define the distal limits of the *shak-B* locus. The chromosomal walk could thus be stopped in this direction and DNA proximal to the *Ymal*^{171} breakpoint studied in more detail, either for the presence of *shak-B* transcripts and / or control regions or for other deficiencies and duplications that affect the locus.

The proximal end of *Df(1)16-3-35* and the distal end of *Df(1)LB6* define genetically the lethal domain of the *shak-B* locus (Baird *et al.,* 1990). The 17kb of DNA between these two endpoints must contain either structural or regulatory DNA that is necessary for a vital function in the developing embryo (Section 1.2.5). When DNA proximal to the *Df(1)16-3-35* proximal breakpoint was radiolabelled and used as a probe to screen a 12-24 hour embryonic cDNA library a 1.8kb clone was isolated (Crompton *et al.,* 1992). A variety of techniques have since been used to isolate a further four clones from this region, three of which have been isolated by workers from this lab. (M. Wilkin, S. Ji, D. Crompton, personal communications) and a fourth has been isolated by another group (Krishnan *et al.,* 1993). To date, at least five different transcripts have been isolated from the region defined by these two endpoints.

Little can be said about the breakpoints of *Df(1)A118*, *Df(1)HC279* and *Df(1)26B* as they break between the walks, and so the positions of the endpoints relative to the walks and to any transcripts and / or control regions is unknown. One breakpoint that may help explain the effects these deficiencies have on the *shak-B* locus is *Df(1)17-351*, the distal end of which was found in the proximal end of the 952 walk (Section 3.8). *Df(1)A118* and *Df(1)HC279* when heterozygous for *shak-B* alleles give severe neural effects *i.e.* they cause severe shaking under ether anaesthesia (G.L.G. Miklos, unpublished observations) whereas *Df(1)17-351* gives only a mild effect (shakes mildly under ether). The position of the *Df(1)17-351* breakpoint is intriguing as the molecular location does not fit the location expected from the genetic analysis of these deficiencies. It has less of an effect than *Df(1)A118* or *Df(1)HC279*, even though it deletes the same regions of DNA plus other more distal regions (Figure 3.10g). Several possible explanations could account for this. It is possible that an antimorphic effect is occurring due to a structural region of the *shak-B* locus or even a second locus the protein products of which interacts with *shak-B*, being disturbed by *Df(1)A118* and *Df(1)HC279*. If only a part of the coding region was deleted then the remaining portion of the protein could have a deleterious effect on the fly as the altered protein may compete for a target with the wild type protein. *Df(1)17-351* will delete all of the protein coding region and so will have a less severe effect as there is no interference with wild-type function. Another possibility is a position effect. *Df(1)A118* and *Df(1)HC279* may have their severe effects by bringing in novel sequences that interfere with the production of the *shak-B* transcripts or proteins produced either from the 19E3 or 19E5-19E6 regions. This is made more likely by the fact that both *Df(1)A118* and *Df(1)HC279* break genetically in the same place (at *unc,* a locus which is known to be rich in repetitive DNAs [Healy *et al.,* 1988]). This effect could also be due to the introduction of a new regulatory region, for example an enhancer that causes aberrant expression of the *shak-B* products either temporally or spatially. This effect is unlikely to be causing over expression of the protein as genetically this type of mutation would be dominant *i.e. Df(1)A118 / +* would have a mutant phenotype, which is not observed. Alternatively, a region that is regulating the 19E3 region could be replaced by a 'silencer' region which switches off production of *shak-B* products. A similar effect to this would be caused by the juxtaposition of heterochromatic DNA which would cause a condensation effect on the chromatin (Pirrotta, 1990). This effect can often stretch for several hundred kilobases and

effectively switches off the whole region as RNA polymerases cannot penetrate the condensed structure. This effect is observed when genes are placed next to α-heterochromatin (Section 1.4.1). However it is unlikely that this effect would also be observed when juxtaposed to the β–heterochromatin, present in and around the *unc* locus (Miklos *et al.,* 1984; Healy *et al,* 1988), which is where both the *A118* and *HC279* deficiencies have their proximal endpoints. Another possible explanation could be the deletion of one enhancer by *Df(1)A118* and *Df(1)HC279* which is also deleted by *Df(1)17-351* but which is then replaced by a novel enhancer almost as effective from the *tuh / eo* region at 20A1 / 2 in which the proximal end of *Df(1)17-351* lies.

One other factor that may be responsible for the effect of *Df(1)17-351* is the unusual structure of the deficiency. Prud'homme *et al.,* (1995) suggest that it may have multiple small deletions more distal than the 'major' deletion and so it is possible that this study has detected one of these smaller deletions, more distal than the one that actually has the effect on the *shak-B* locus. These speculations will not be resolved until the cloning of the 19E3 to 19E5 region, along with the associated localisations and characterisations of the *A118* and *HC279* deficiencies, is completed.

Chapter Four

Chapter Four - Characterisation of Repetitive DNA From the 19E Region.

4.1 Introduction.

Repetitive DNA is a characteristic feature of eukaryotes and can account for 10 - 60% of the genome (for example Goldberg *et al.,* 1975). General information about the organisation of this repetitive DNA has been obtained from an extensive analysis of the reassociation kinetics of DNA from many species (see Davidson and Britten, 1973 for review). This procedure makes use of the fact that in reassociation experiments repeated DNA sequences reanneal at low to moderate C_0t values and single copy sequences at high C_0t values. Using hydroxyapatite columns which holds back partly and completely double stranded molecules, it is possible to measure the percentage of DNA held back by the column as a function of the length of the DNA fragments, (a so called interspersion analysis). Using this approach studies have revealed two major patterns. The first is exemplified by human DNA in which low copy number sequences of about 2.2kb separate repetitive regions of around 0.3kb, in approximately 60% of the genome (Schmid and Deininger, 1975). In contrast to this pattern, Manning *et al.,* (1975) showed that "The average distance between repeated sequences in *Drosophila melanogaster* is at least 12kb, the largest fragments used in the experiment." Electron micrographs of the reannealed molecules also showed that on average the length of the repeats in the *Drosophila melanogaster* genome was 5 to 6kb (Manning *et al.,* 1975).

The characterisation and occurrence of this repetitive DNA in the *Drosophila* genome has since been investigated using a variety of different approaches. One approach involved the identification and cloning of DNA solely on its repetitiveness (*e.g.* Wensink *et al.,* 1974; 1979 ; Lohe and Brutlag, 1986). Others have involved locating and characterising transposable sequences (see Section 1.4.2.1) in the genome using *in situ* hybridisation based techniques (Ananiev *et al.,* 1984; Hey and Eanes personal communication). Still others have involved characterising repetitive DNA near to or within transcripts of genes under investigation (*e.g.* Magoulas and Hickey, 1992; Wharton *et al.,* 1985) or the characterisation of repetitive fragments from genomic regions that have been microdissected from polytene chromosomes and then cloned (Pirrotta *et al.,* 1983, Miklos *et al.,* 1984). Probably the most informative approach however, has involved the characterisation of repetitive DNA isolated in chromosome

walks (Bender *et al.,* 1983a). There are several examples of long 'walks' whose repetitive DNA content has been characterised to at least some degree. These include the 195kb of the *Bithorax*-complex walk (Bender *et al.,* 1983b), the 200kb walk that cloned the *white* locus (Pirotta *et al.,* 1983), the 115kb of DNA cloned from in and around the *Antenapedia*-complex (Scott *et al.,* 1983) and the 315kb of DNA that was isolated (Bender *et al.,* 1983a) and characterised (Hall *et al.,* 1983) in the cloning of the *ace-rosy* region.

4.1.1 The 19E Region.

Four chromosomal walks have been analysed in this study. These are located in a transition zone between the euchromatic and heterochromatic parts of the chromosome. The euchromatic half consists of approximately a thousand vital complementation groups (Lefevre and Watkins, 1986) and contains very few repetitive sequences compared to the heterochromatic half of the chromosome. The heterochromatic region is composed of two main types of repetitive DNA interspersed with single-copy gene sequences. α-heterochromatin is highly compacted chromatin which is located principally at the centromere. It is largely transcriptionally silent and only 30 or so vital genes can be assigned to all of the X and autosomal α-heterochromatic regions. Many of the genes in the α-heterochromatin seem to rely on the peculiar α-heterochromatic environment for their functioning (Section 1.4.1) *e.g.* the *rolled* locus (Eberl *et al.,* 1993). Other transcribed regions such as the 18S and 28S rRNA genes and some type I insertion sequences are also found in the a-heterochromatin of the X chromosome (Kidd and Glover 1980; Hilliker *et al.,* 1980a; Hilliker and Appels, 1982). It is thought that α-heterochromatin is made up of highly repetitive sequences such as the satellite DNAs (Brutlag, 1980; Miklos and Cotsell, 1990; Lohe and Brutlag, 1986; John and Miklos 1979; 1988). The α-heterochromatin is bounded by diffuse blocks of the second type of heterochromatic DNA, β–heterochromatin. This is transcriptionally active and has a gene frequency little different from that of the euchromatic portion of the chromosome (Miklos and Cotsell, 1990). β–heterochromatin does however, contain much more repetitive DNA than (most) euchromatic regions. The majority of this repetitive DNA is thought to be predominantly composed of nomadic elements (Ananiev *et al.,* 1984) and has a large proportion of clustered scrambled repeats (Wensink *et al.,* 1974; 1979) and other, as yet unidentified repetitive non-satellite repeats (see Yamamoto *et al.,* 1990 for more

details). Thus, from the point of view of studying repetitive DNAs, the area in which the walks are situated is a fascinating region, being as it is on the border between the largely non-repetitive euchromatin and the highly repetitive, heterochromatic parts of the X chromosome.

In this chapter, I report the use of three approaches to characterise repetitive DNA contained within the four chromosomal walks used in this study (Chapter 3). Firstly, the number of times a fragment is repeated in the genome was determined by hybridising them to blots of genomic DNA. Secondly small fragments (1kb or less), or the ends of larger fragments known to be located in repetitive regions, were sequenced. The sequences obtained were used to search the GENBANK and EMBL databases (Devereux *et al.*, 1984). The last approach involved using previously identified and characterised transposable elements to probe blots of the genomic phage clones from the walks.

4.2 Results.

4.2.1 Location of Repetitive Regions in the Four Walks.

460kb of DNA had previously been isolated in four chromosomal walks across the 19E1 - 19E4 region. This DNA was examined for repetitive sequences using the reverse Southern technique. Southern blots were prepared of the genomic phage clones from the walks. These Southern blots were hybridised with *OrR* genomic DNA labelled to low specific activity (5×10^6 cpm μg^{-1}) using nick-translation (Section 2.5.13.1). The results obtained are shown in Figures 4.01. The λS genomic clones (Section 3.2.3) were not included on the 952 walk blot (Figure 4.01A) as they were isolated after this part of the study had been completed. However they were restricted with *Hin* dIII / *Sal* I and Southern blotted and a reverse Southern analysis carried out. The results are shown in Figure 4.06. Due to time constraints it was not possible to hybridise the repetitive fragments from these clones to a genomic blot so actual levels of repetition could not be assessed further. The repetitive region in the 6kb p94.R1 fragment was mapped further using the same technique. Only those regions repeated several times in the genome will be visible upon autoradiography. Non-repetitive DNA will have been labelled but to such a low level as to be invisible or at best, visible only weakly depending upon the actual concentration of the sequence. The intensity of the signal is dependent upon the number of copies of the repeat. For

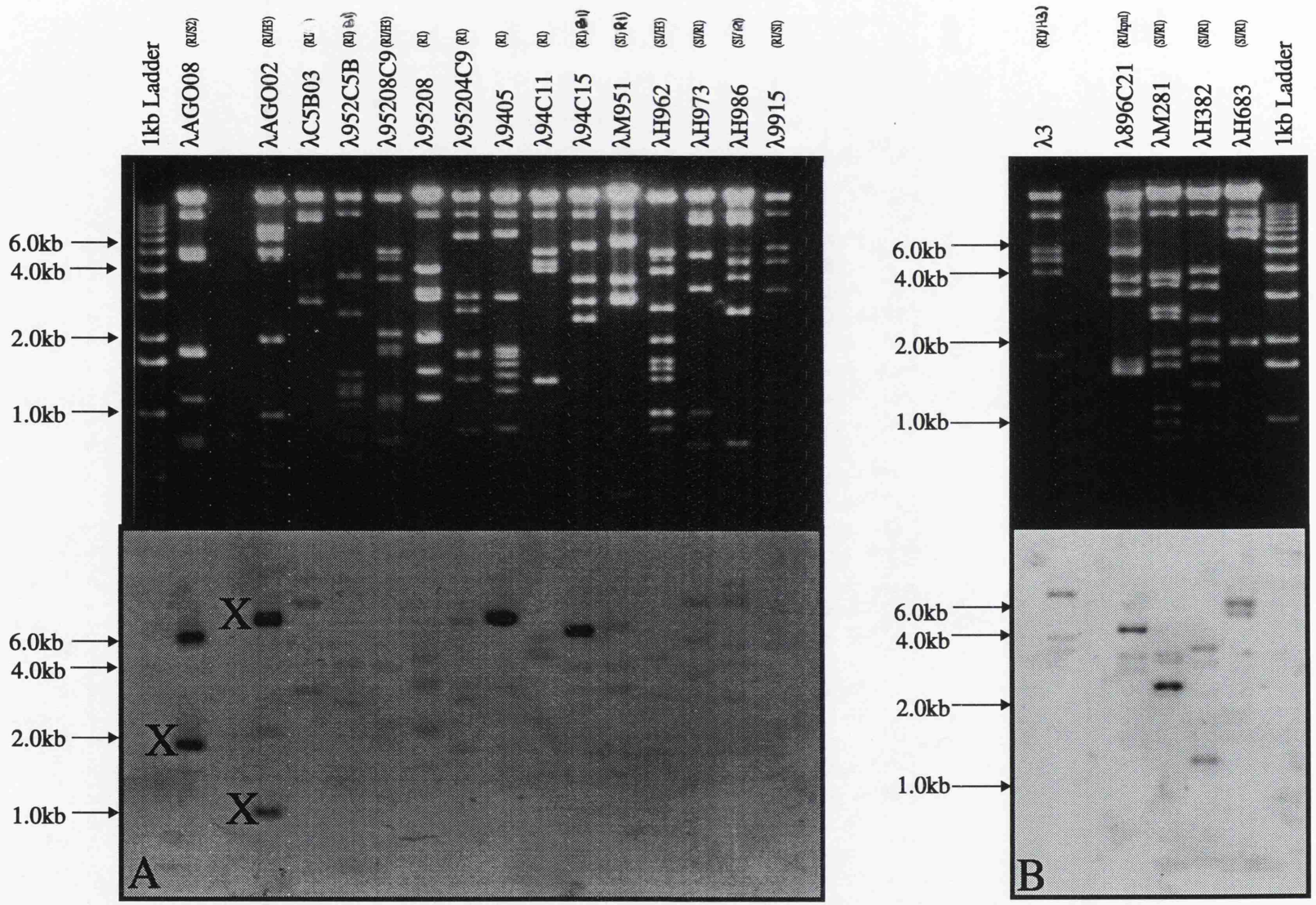


Figure 4.01a and b- Ethidium bromide stained gels and corresponding autoradiographs of the genomic phage clones from the (A) 952 walk and (B) 896 walk, probed with radio-labelled genomic DNA in a 'reverse Southern' analysis. The enzymes used to restrict the phage clones are indicated in brackets. The fragments labelled X are known to contain *opa* repeats. See text (4.2.1) and appendices 1 and 2 for location of repetitive areas.

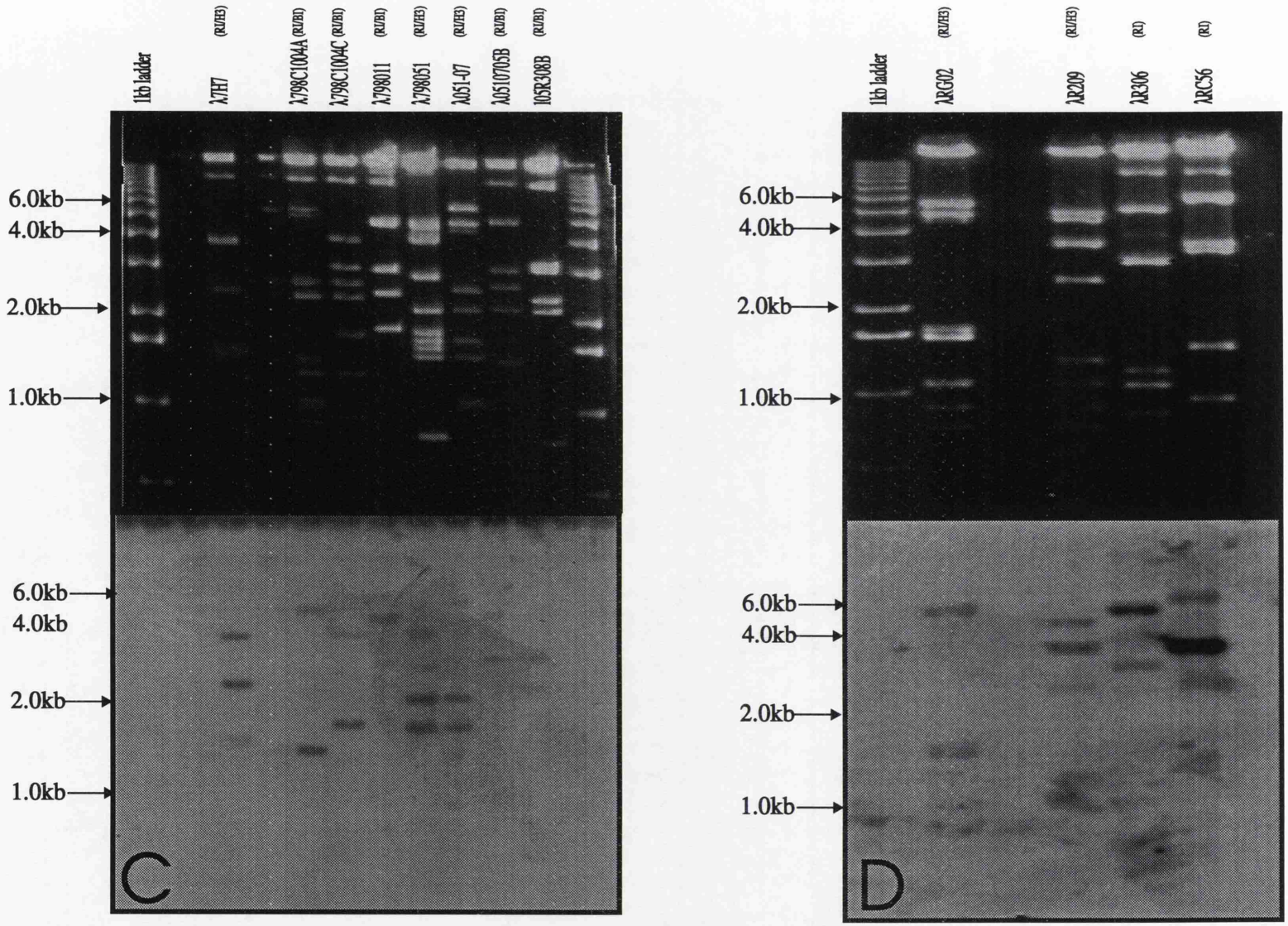


Figure 4.01c and d- Ethidium bromide stained gels and corresponding autoradiographs of the genomic phage clones from the (C) 798 walk, (D) *runt* walk, probed with radio-labelled genomic DNA in a 'reverse Southern' analysis. The enzymes used to restrict the phage clones are indicated in brackets. See text (4.2.1) and appendices 3 and 4 for location of repetitive areas.

example the fragments (labelled X in Figure 4.01a) contain *opa* repeats (Section 1.4.2.3) which occur hundreds of times in the genome (Wharton *et al.*, 1985) thus giving a strong signal in the hybridisation. It is difficult to give an exact figure as to how many times a sequence must be repeated, for it to be detectable with the reverse-Southern technique, due to the unknowns and variables involved *e.g.* length of repetitive DNA, hybridisation rates *etc.* However, it is possible, to estimate the degree of repetition that is detectable from the number of bands that were obtained when the repetitive fragments were hybridised to genomic blots. (It should be noted that some fragments must be excluded from this analysis as when hybridised to a genomic blot they gave only the number and size of bands expected from their size (see below for an explanation of this). When the genomic blots are studied (Section 4.2.2 and Table 4.1), it is possible to say that sequences repeated more than four times were identified *i.e.* the minimum number of times a region must be repeated to be detected by this study was four. This is provided that the fragments do not contain short or weakly homologous sequences (see below).

Using the reverse Southern technique many repetitive regions were identified, 16 of these (the strongest hybridising and those known to be repetitive from use in other parts of this study) were then characterised in more detail. See Figure 4.01a, b, c and d; Table 4.1 and Appendixes 1, 2, 3 and 4 for the locations of the repetitive regions studied.

4.2.2 Determination of the Repetitiveness of the Regions.

Genomic regions identified as containing repetitive DNA were isolated from their parent phage clones by electroelution and radiolabelled using random priming (Sections 2.5.10, 2.5.13.2). The labelled fragments were then hybridised under standard conditions (Sections 2.5.16), to genomic blots of *OrR, CS* and *Sierra Leone* genomic DNAs which had been digested with three restriction endonucleases, *Eco* RI, *Bam* HI and *Hin* dIII. Three different restriction enzymes were used to alleviate the problem of one enzyme cutting out the same size fragment from multiple copies of the same repeat but apparently giving only one band. The concentrations of the probes were adjusted such that the specific activity per μg and per kb of the labelled fragments were equivalent. The regions used as probes to challenge the restricted genomic DNAs were always the smallest possible restriction fragments that

Repetitive region:-	Class:-	Identified as:-	Transposable	Comments:-	Reference:-
1.0kb λAGO02	Class I	*opa*	No	Only *opa* repeat	Figure 4.02a
6.6kb λAGO02	Class I	*opa*	Yes	Multiple copies of the *opa* repeat or other repeat(s) probably present	Figure 4.02b
6.0kb λ9405	Class I	β-heterochromatin repeat	Yes	Other repeat(s) probably present	Figure 4.02c
5.7kb λ94C15	Class I	**N/I**	No	**N/D**	Figure 4.02d
1.4kb λS2	**N/D**	Unidentified retrotransposon	**N/D**	Sequence suggests that it is a retrotransposon	Section 4.2.4
5.0kb λ896C21	Class III	*opa*	No	Probably only *opa* repeat	Figure 4.03a
7.4kb λM283	Class III	*opa*	No	Probably only *opa* repeat	Figure 4.03b
2.6kb λM281	Class III	*opa + 412 + 297*	No	See text	Figure 4.03c
3.5kb λH683	Class III	**N/I**	No	**N/D**	Figure 4.03d
5.4kb λH385	Class III	**N/I**	No	**N/D**	Figure 4.03e
11.2kb λ7H7	Class II	*opa*	Possibly	Multiple copies of the *opa* repeat or other repeat(s) probably present	Figure 4.04a
5.6kb λ798C10O4A	Class I	**N/I**	Possibly	**N/D**	Figure 4.04b
2.6/1.4kb λ798C10O4A	Class III	*opa*	No	Probably only *opa* repeat	Figure 4.04c
1.8kb λ798051	Class II	**N/I**	No	**N/D**	Figure 4.04d
5.2kb λ798051	Class III	**N/I**	No	**N/D**	Figure 4.04e
1.0,1.6,6.0kb λ79805107	Class II	*copia*	Yes	See text (4.2.5 and 4.3.4)	Figure 4.04f
3.5,5.2kb λ0510705B	Class I	**N/I**	Possibly	**N/D**	Figure 4.04g
0.8,3.1kb λ05R308B	Class III	Faint *copia* hybridisation	No	See text (4.2.5 and 4.3.4)	Figure 4.04h
7.5kb λR03	Class III	Faint *opa* hybridisation	No	*opa* hybridisation in a lambda containing homologous sequences	Figure 4.05a
4.3kb λR03	Class III	**N/I**	No	**N/D**	Figure 4.05b
6.6kb λR209	Class III	**N/I**	No	**N/D**	Figure 4.05c
3.5,4.0kb λRC56	Class I	*opa*	No	Multiple copies of the *opa* repeat or other repeat(s) probably present	Figure 4.05e

Table 4.1 Summary of results obtained from the analysis of repetitive regions from the 952, 896, 798 and *Runt* chromosomal walks. **N/D**=No data; **N/I**=No information. The column labelled **Class** indicates the repetitiveness of the fragments in the Genomic southern blot analysis; Class 1: >15 bands, Class 2: 3-14 bands, Class 3: <3 bands. (See Section 4.2.2 for more details). The column labelled **Transposable** indicates whether the restriction pattern from the Genomic southern blot analysis suggests transposabilty of the fragment used in the analysis (Section 4.2.3 for more details). Those repetitive areas that extend over several identified restriction sites were individually isolated and radiolabelled and are indicated as being several fragments in column one of the table.

completely contained the repeats, as judged by the reverse Southern analysis (Section 4.2.1). If the repetitive region extended over a large area, with several identified internal restriction sites *e.g.* the region in the genomic phage clone λ798-051-07 (Appendix 3 for map and location) then subfragments were individually isolated and used as probes. The largest fragment used in this study was 11.2kb and the smallest 1.0kb; the average probe size was 3.5kb. All the probes used in this part of the study had an average specific activity of 1.3×10^9 cpm μg^{-1}, which was judged to be satisfactory.

All filters were hybridised and washed under standard conditions (0.1X SSC, 0.1% SDS at 65°C). The filters were then autoradiographed (Figures 4.02, 4.03, 4.04 and 4.05). After exposure the number of bands were counted and the repetitive fragments placed into one of three groups. Group I contained repetitive regions that yielded 15 bands or more, Group II fragments yielded between 4 and 14 bands whereas Group III fragments yielded three bands or less. Table 4.1 contains a summary of the results.

4.2.3 Transposability of the Repetitive Sequences.

The task of identifying a repetitive sequence as a transposable element is normally not difficult. If the sequence is observed to hybridise to multiple locations in the genome, and if these locations are different in different strains of *Drosophila melanogaster* then a conclusion of transposition is warranted. The most direct method of showing transposability is of course hybridisation to a panel of previously identified transposable elements, which was one approach utilised in this study (Section 4.2.5). However, even if all the transposable elements so far reported are used, this approach can never be said to be complete as new elements are continuously being reported *e.g. Bari-1* (Caizzi *et al.*, 1993).

An alternative approach is to perform *in situ* hybridisations of unknown repetitive sequences to polytene chromosomes. However, it has been shown previously (for example see Wharton *et al.*, 1985) that short repetitive regions such as the *opa*, *pen*, and *ala* repeats yield poor quality results with *in situ* hybridisations. This is presumably due to the small size of the repeats (*ca.* 150bp or less) giving low regions of homology. As a large number of the repetitive regions were known to contain *opa* repeats (Section 4.2.5) this approach was not pursued.

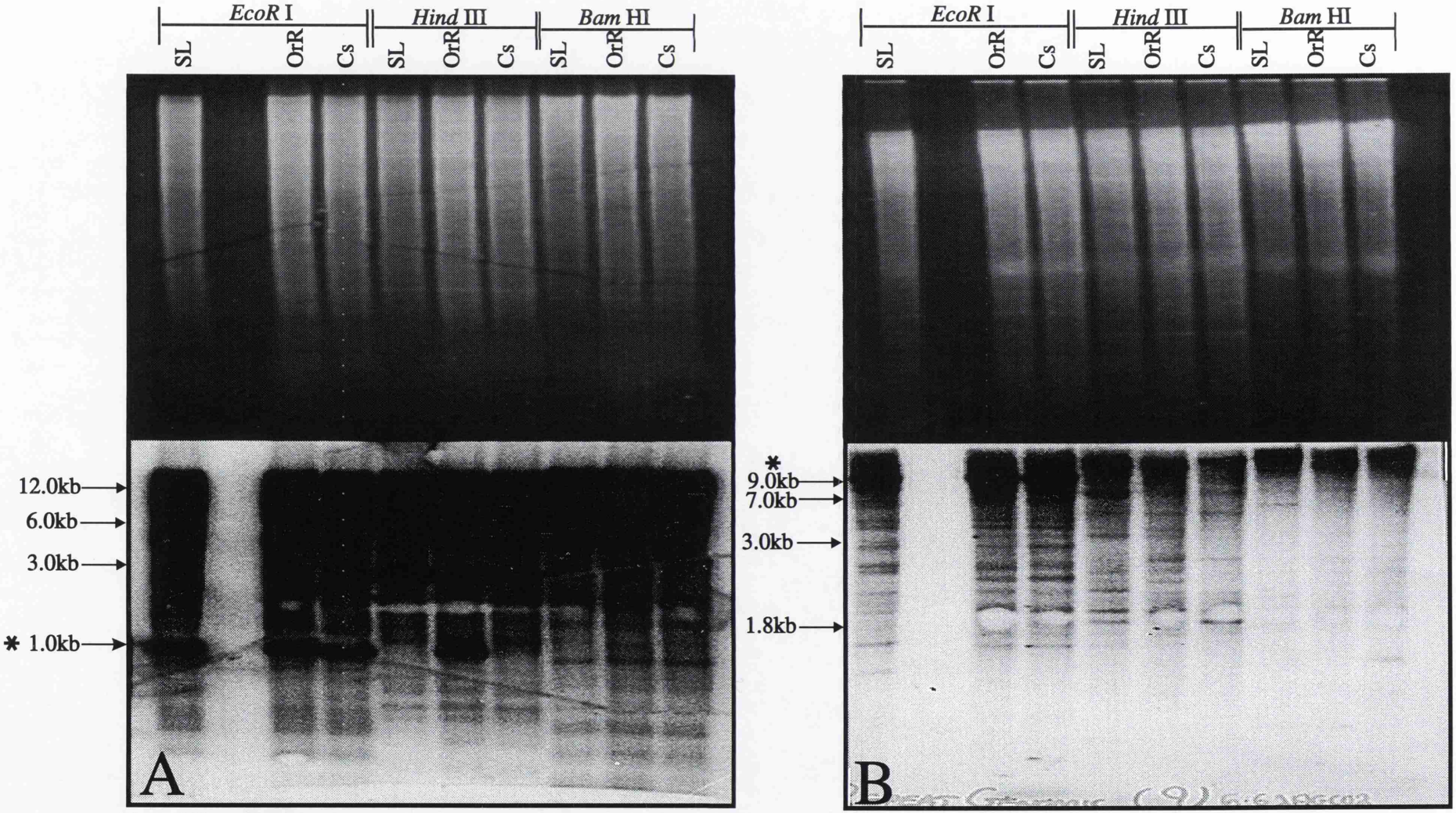


Figure 4.02a and b- Ethidium bromide stained gels and corresponding autoradiographs from genomic Southern blots of restricted wild-type DNAs hybridised with A)1.0kb *Eco* RI and B) 6.6kb *Eco* RI *Hin* dIII fragments from λAGO02. See Table 4.1 and text for an interpretation of the results. The sizes of the restriction fragments to which the regions used as probes will hybridise are indicated by asterixs.

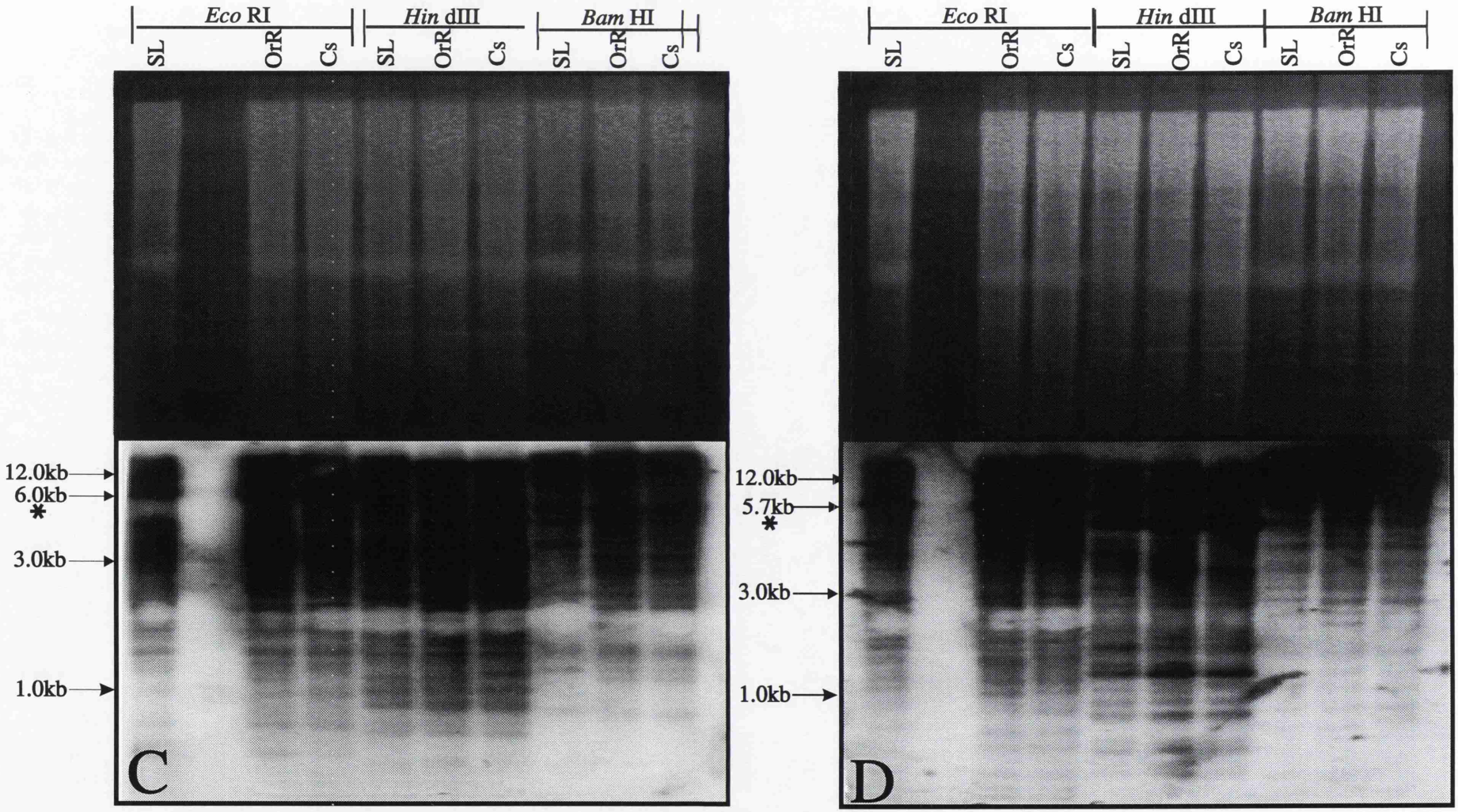


Figure 4.02c and d- Ethidium bromide stained gels and corresponding autoradiographs of genomic Southern blots of restricted wild-type DNAs hybridised with C) p94.R1, a 6kb *Eco* RI fragment from λ9405 and D) p94C11R2, a 5.7kb *Eco* RI fragment from λ94C15. See Table 4.1 and text for interpretation of the results. The sizes of the restriction fragments to which the regions used as probes will hybridise are indicated by asterixs.

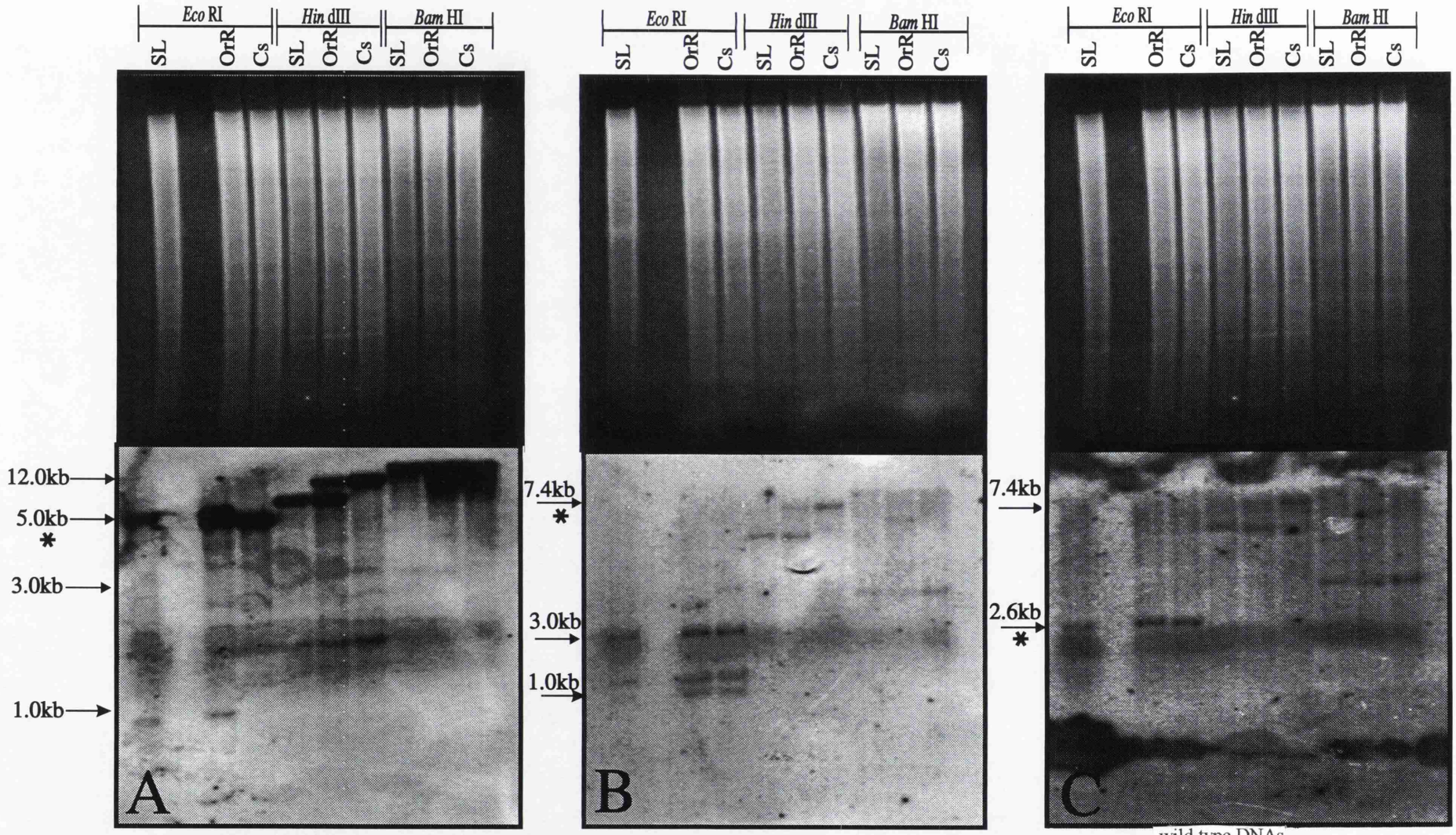


Figure 4.03a,b and c-Ethidium bromide stained gels and corresponding autoradiographs from genomic Southern blots of restricted wild type DNAs hybridised with A) 5.0kb *Kpn* I fragment from λ896C21, B) 7.4kb *Hin* dIII fragment from λM283 (a phage that overlaps λM281 and contains the same repetitive region) and C) 2.6kb *Eco* RI fragment from λM281. See Table 4.1 and text for an interpretation of the results. The expected sizes of the fragment used as probes are indicated by asterixs.

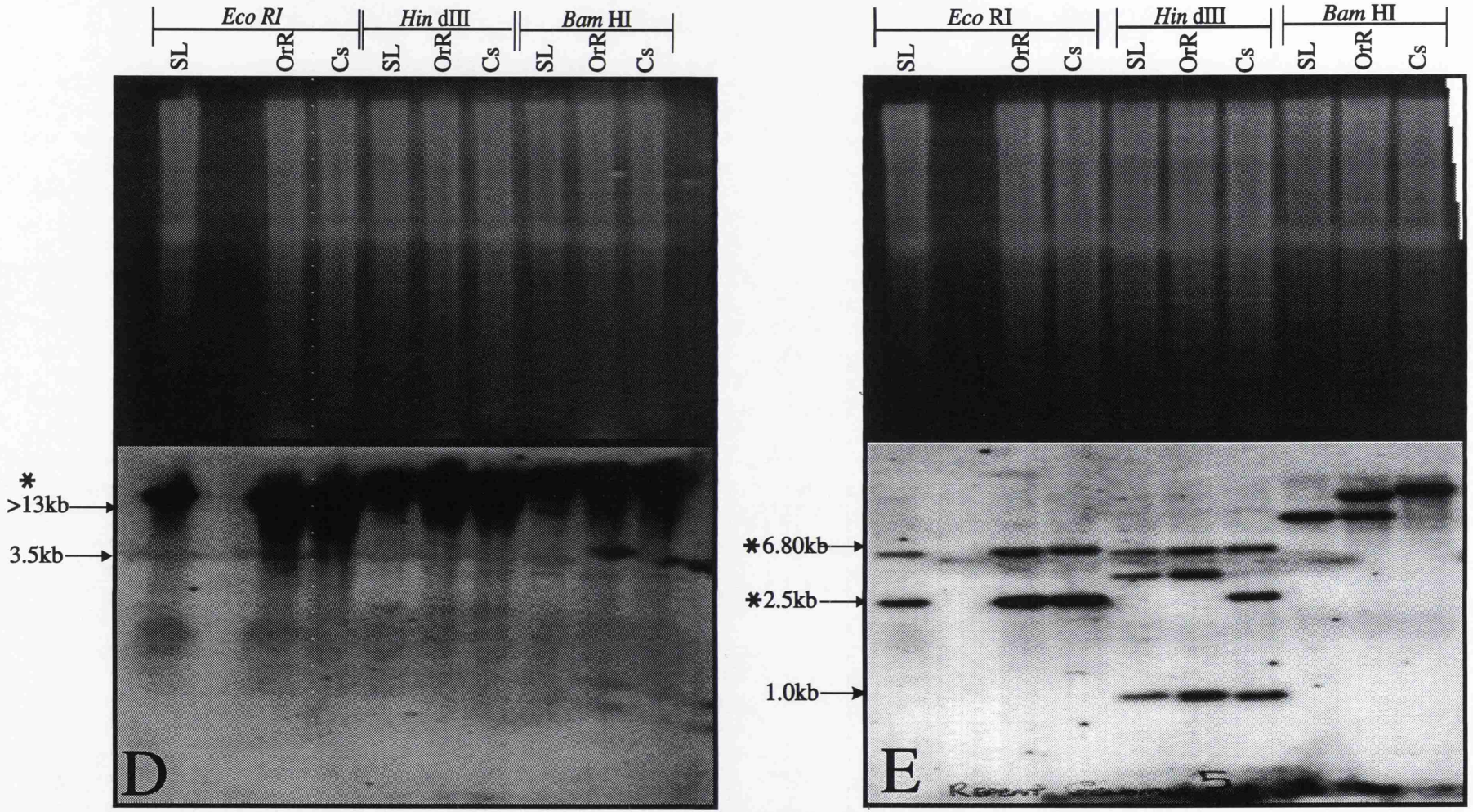


Figure 4.03d and e- Ethidium bromide stained gels and corresponding autoradiographs from genomic Southern blots of restricted wild-type DNAs hybridised with D) 3.5kb *Ava* I fragment from λH683 and (E) 5.4kb *Sal* I fragment from λH385 (a phage that overlaps λH382 and contains the same repetitive region). See Table 4.1 and text for an interpretation of the results. The sizes of the restriction fragments to which the regions used as probes will hybridise, are indicated by asterixs.

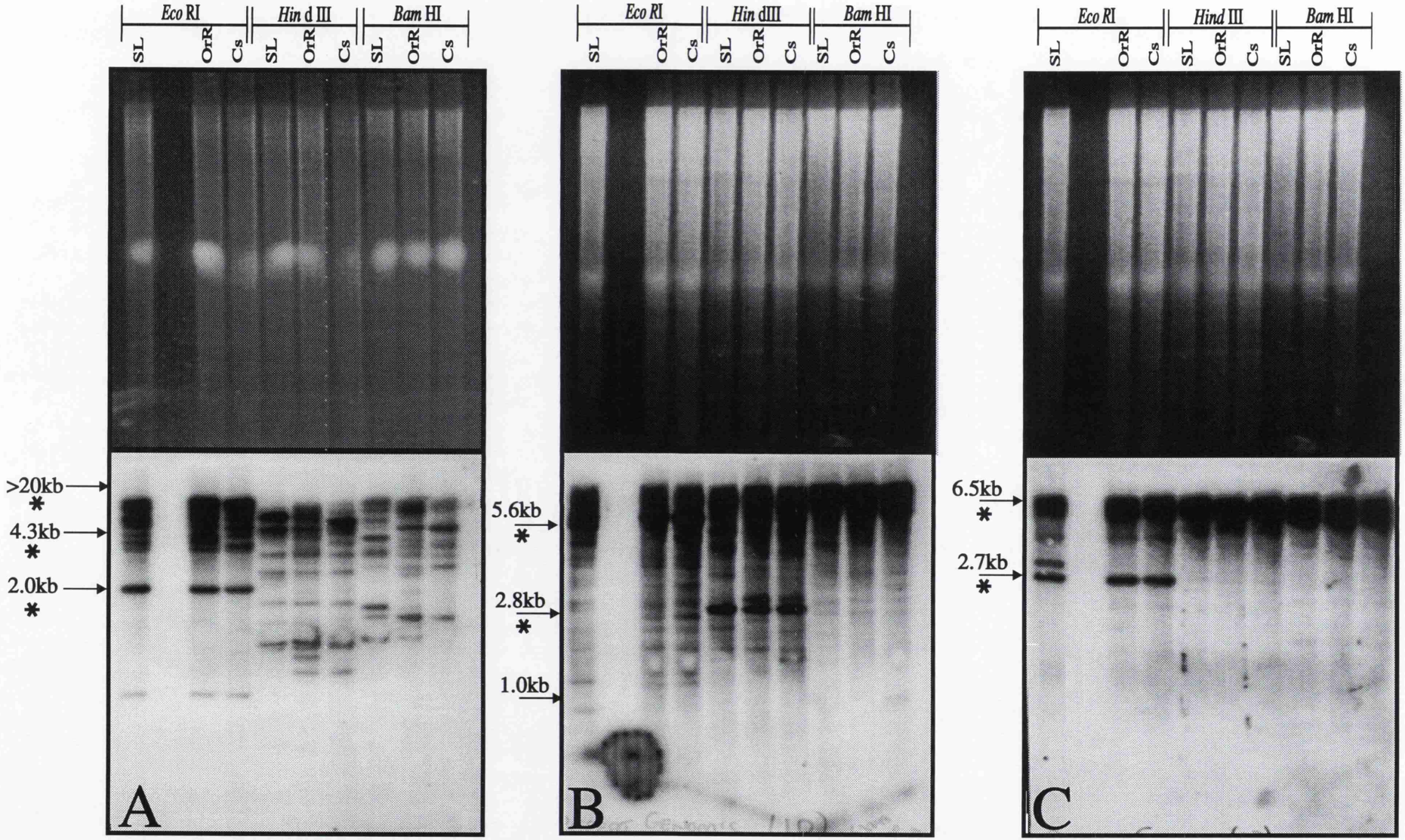


Figure 4.04a, b and c- Ethidium bromide stained gels and corresponding autoradiographs from genomic Southern blots of restricted wild-type DNAs hybridised with A) 1 1.2kb *Bam* HI fragment from λ7H7, B) 5.6kb *Eco* RI fragment from λ798C10O4A and C) 2kb *Eco* RI and a 1.4kb *Eco* RI - *Bam* HI fragment from λ798C10O4A. The sizes of the restriction fragments to which the regions used as probes will hybridise are indicated by asterixs. See Table 4.1 and text for interpretation of the results.

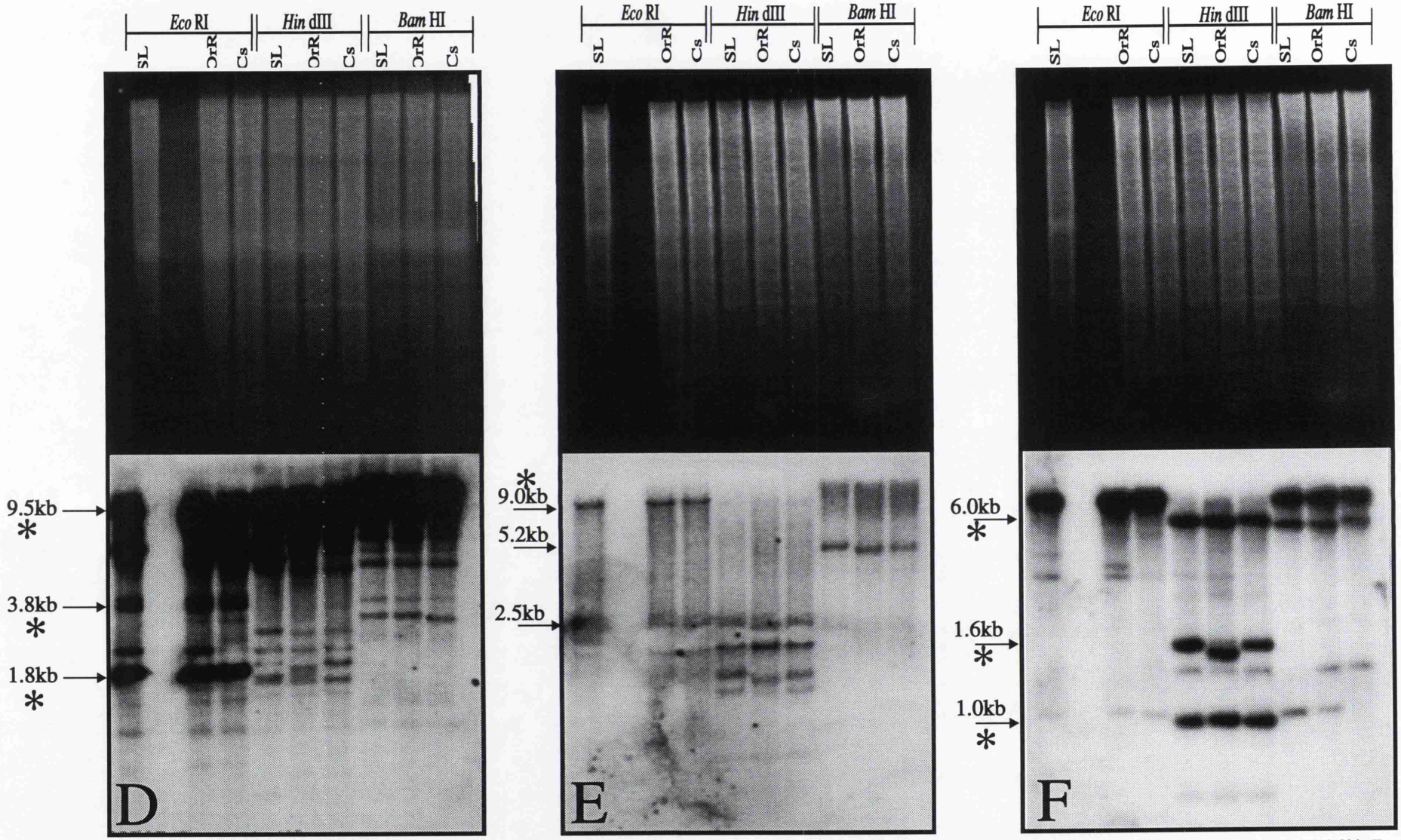


Figure 4.04d, e and f-Ethidium bromide stained gels and corresponding autoradiographs from genomic Southern blots of restricted wild-type DNAs hybridised with D) 1.8kb *Eco* RI fragment from λ798051, E) 5.2kb *Eco* RI fragment from λ798051 and F) 1.0kb, 1.6kb and 6.0kb *Hin* dIII fragment from λ05107. The sizes of the restriction fragments to which the regions used as probes will hybridise are indicated by asterixs. See Table 4.1 and text for an interpretation of the results.

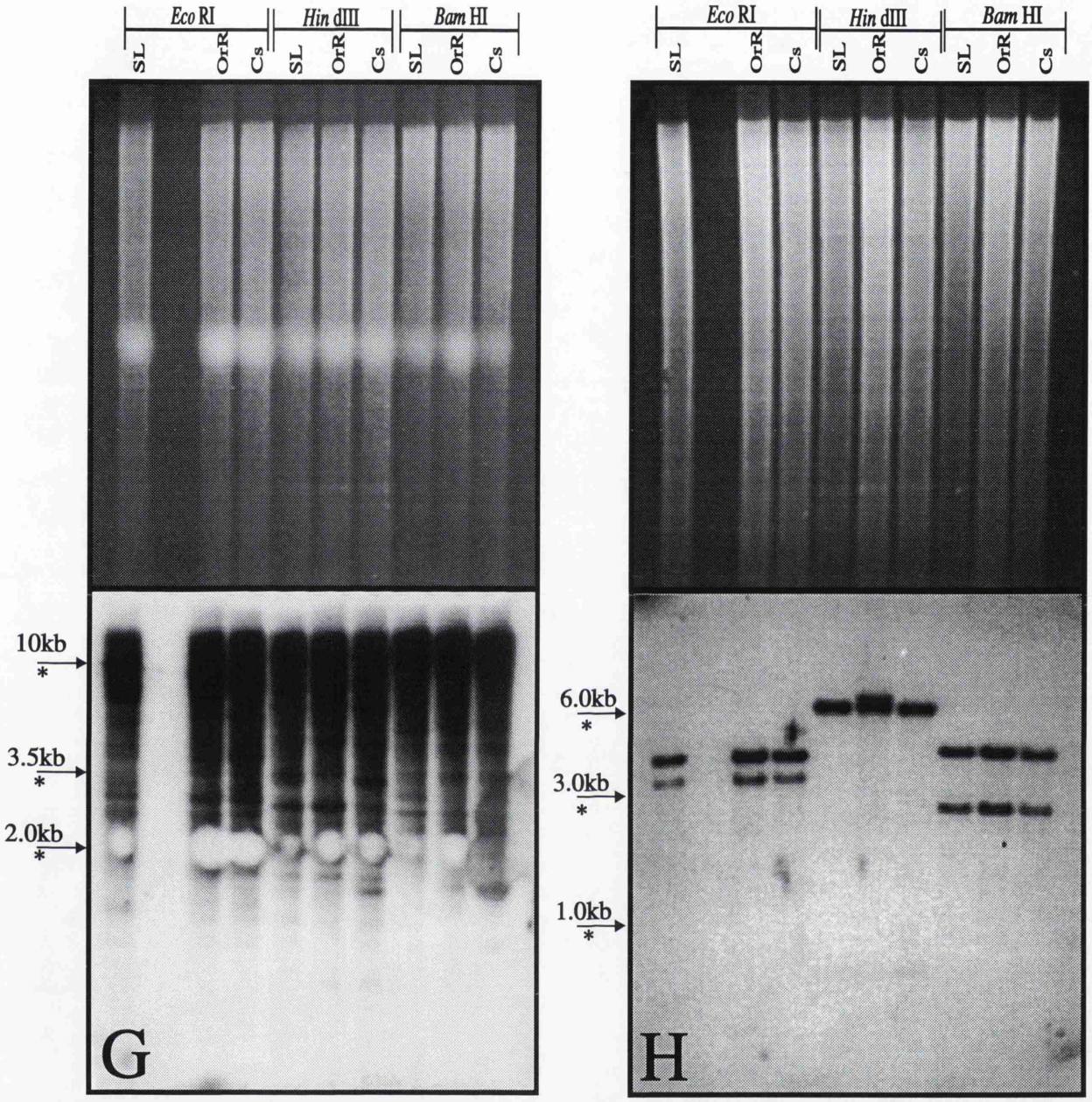


Figure 4.04g and h-Ethidium bromide stained gels and corresponding autoradiographs from genomic Southern blots of restricted wild type DNAs hybridised with G) 3.5kb *Hin* dIII and 5.2kb *Eco* RI-*Bam* HI fragments from λ0510705B and H) 0.8kb *Eco* RI-*Bam* HI and 2.1kb *Hin* dIII-*Bam* HI fragments from λ05R308B. The sizes of the restriction fragments to which the regions used as probes will hybridise, are indicated by asterixs. See Table 4.1 and text for interpretation of the results.

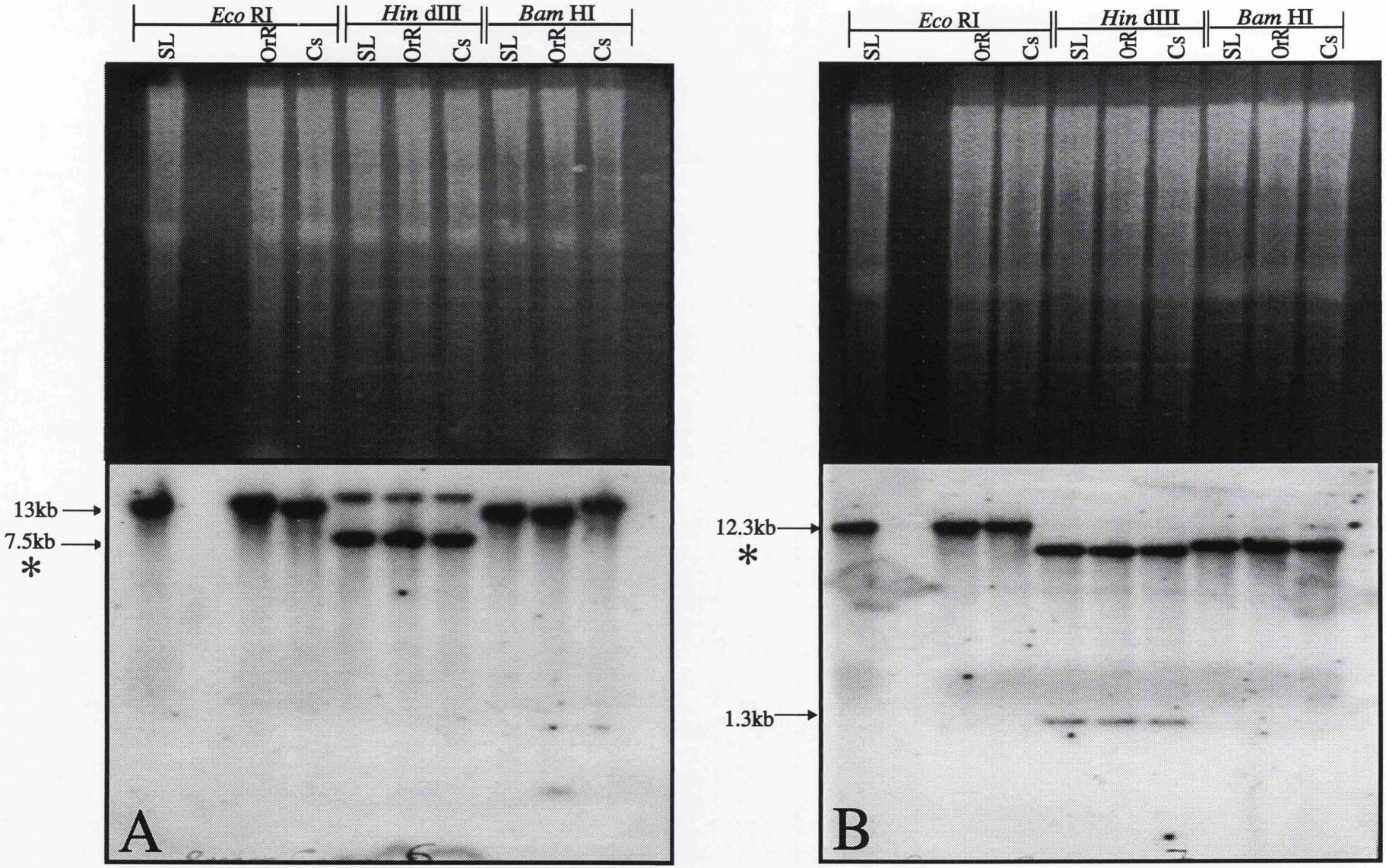


Figure 4.05a and b-Ethidium bromide stained gels and corresponding autoradiographs from genomic Southern blots of restricted wild-type genomic DNAs, hybridised with A) 7.5kb *Hin* dIII fragment from λR03 (a phage that overlaps λRG02 and contains the same repetitive region) and B) 4.3kb *Eco* RI-*Hin* dIII fragment from λR03. The sizes of the restriction fragments to which the regions used as probes hybridise, are indicated by asterixs. See Table 4.1 and text for an explanation of the results.

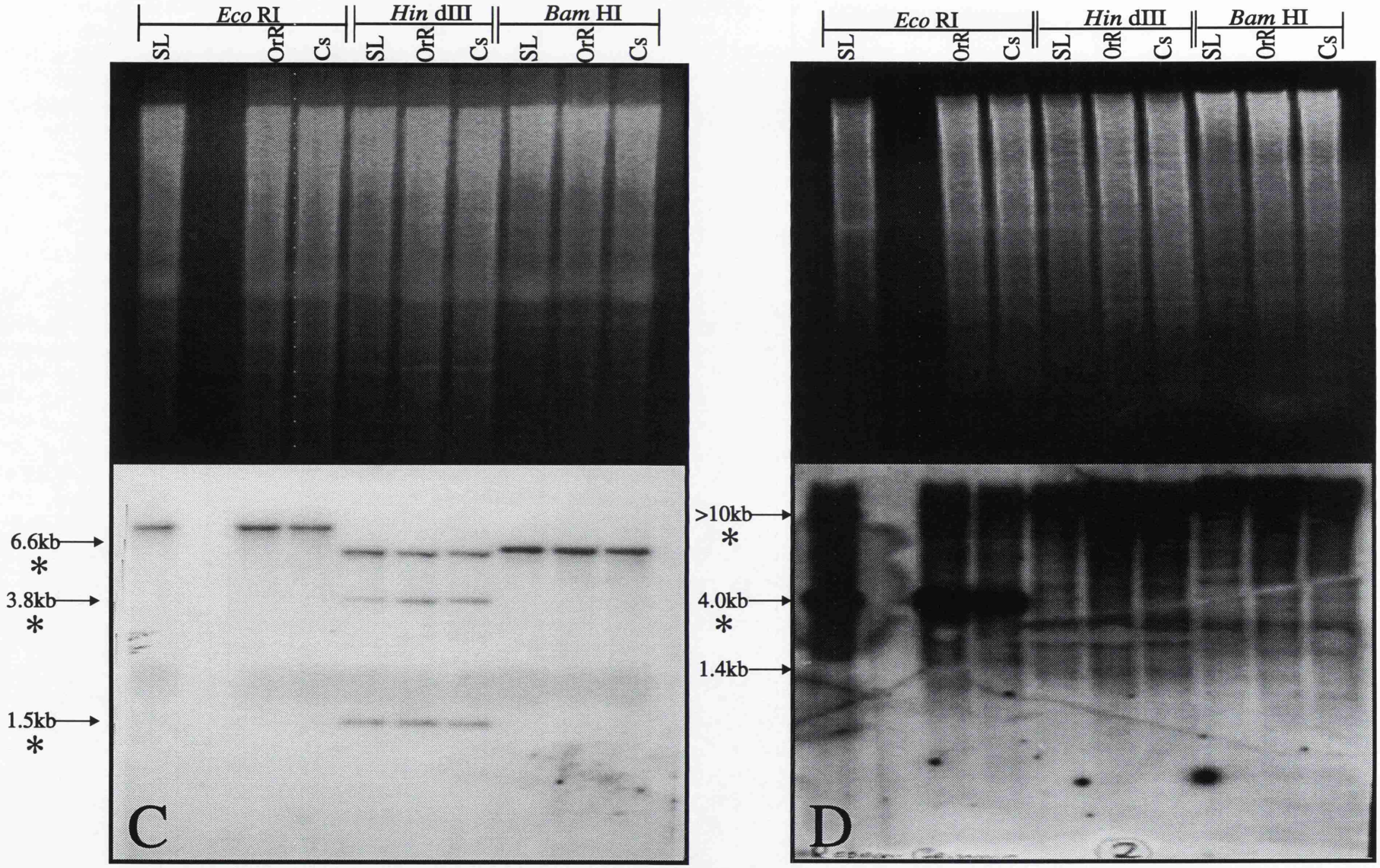


Figure 4.05c and d- Ethidium bromide stained gels and corresponding autoradiographs from genomic Southern blots of restricted wild-type genomic DNAs hybridised with C) 6.6kb *Bam* HI fragments from λR209 and D) 3.5kb and 4.0kb *Eco* RI fragments from λRC56. The sizes of the restriction fragments to which the regions used as probes will hybridise are indicated by asterixs. See Table 4.1 and text for an explanation of the results.

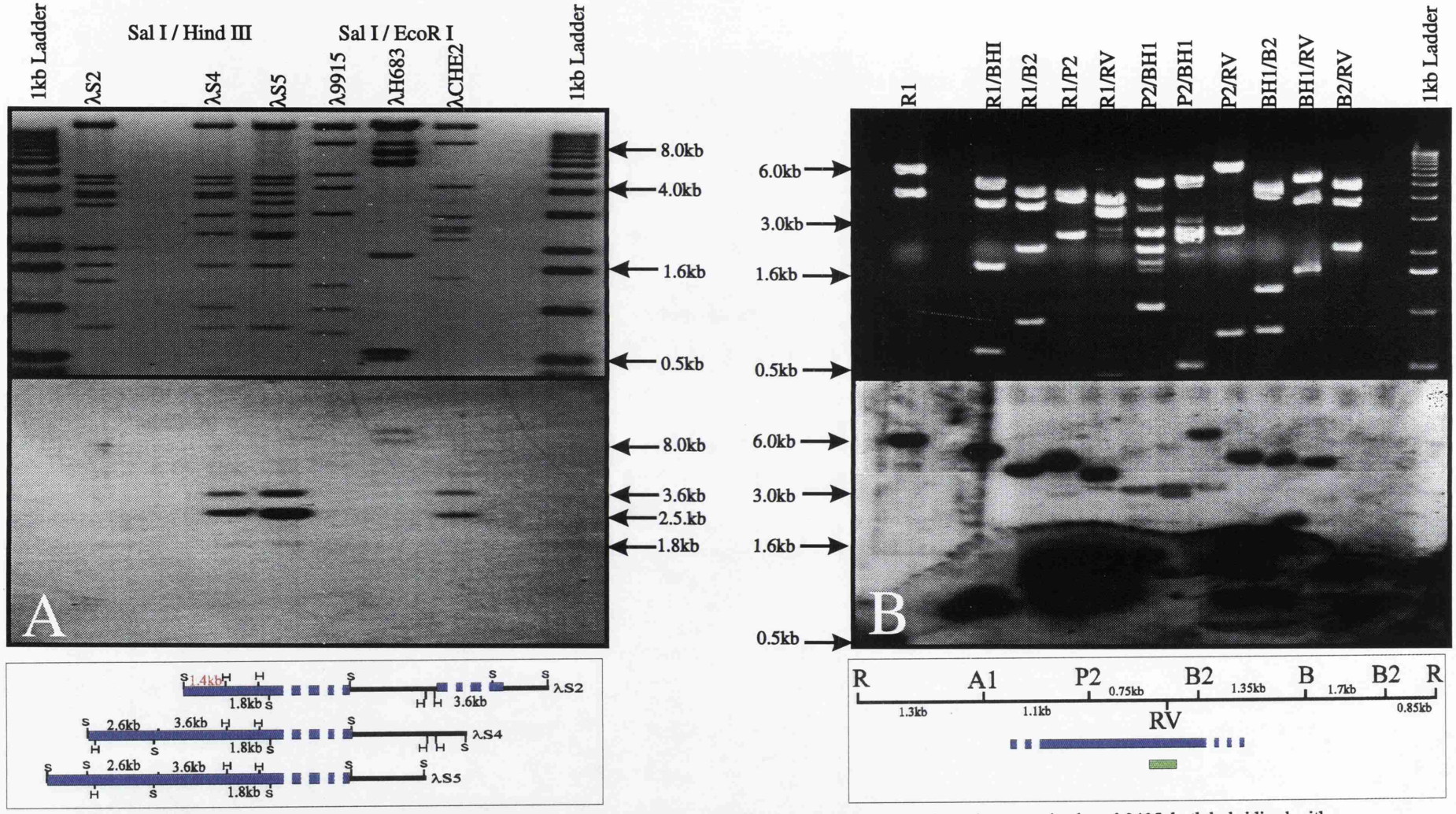


Figure 4.06-Ethidium bromide stained gels and Southern blots of (A) the λS phages, and (B) the p94.R1 subclone from the genomic clone λ9405, both hybridised with radiolabelled genomic *OrR* DNA in a reverse Southern analysis. The diagrams underneath are a schematic representation of the results. Repetitive regions are shown in blue, hatching indicates a less repetitive sequence. The green indicates the approximate position of the fragment from p94.R1 used in the sequence analysis (4.2.4). The 1.4kb λS2 fragment is repetitive but a transfer artifact has stopped the phage DNA from transferring correctly. It is possible however to see the repetition in the 3.6kb fragments with which it shares homology. (Key to restriction enzymes: R= *Eco* RI, S= *Sal* I, H= *Hin* dIII, A1=*Ava* I, P2= *Pvu* II, RV= *Eco* RV and B2= *Bgl* II).

The approach described in Section 4.2.3 (*i.e.* hybridising unknown repetitive sequences to restricted genomic DNA prepared from three different *Drosophila melanogaster* strains) will also indicate transposability (Strobel *et al.,* 1979). It has been shown that although transposable elements can have very different genomic locations between *Drosophila melanogaster* strains (for example Ananiev *et al.,* 1984), within isochromosomal lines most elements transpose at very low rates (Eggleston *et al.,* 1988) although some are known to transpose more frequently when flies are exposed to detrimental environmental conditions such as heavy heat shock (Ratner *et al.,* 1992). Thus by comparing the pattern of bands obtained when hybridising the unknown repetitive sequence to genomic blots it should, if the sequence is transposable, be possible to observe differing banding patterns (and hence chromosomal locations) between the three strains.

This method of inferring transposability employs an important assumption. This is that none of the repetitive regions under investigation are sequences that can expand or contract within populations such as seen with satellite sequences (Waye and Willard, 1985) as these will be indistinguishable from transposable elements using this approach. It is also difficult to distinguish between a single transposable sequence and two unrelated sequences lying next to each other, only one of which is transposable. It is therefore only possible to say that at least some of the sequences in the repetitive regions have homology to regions that are repeated many times throughout the genome and so are likely to be transposable elements. As it is known that many element families have members that are heterogeneous, in size and restriction pattern (*e.g.* Finnegan and Fawcett, 1986), it is not possible to estimate the number of different transposable elements amongst the repetitive sequences reported here. This is because homologous sequences could yield apparently differing results so leading to an overestimation of the number of types of transposable elements amongst the repetitive sequences. Time constraints meant that internal redundancy experiments were not performed which would gone some way to answering this question.

4.2.4 Sequencing of Repetitive Fragments.

For this part of the study four repetitive regions were chosen for sequence analysis. The main criteria for choosing these fragments was the proximity of the repetitive

sequences to an identified restriction site. Small fragments were sequenced from both sides whereas those repetitive sequences with an identified internal restriction site were subcloned and sequences immediately next to this restriction site obtained. A fifth region was sequenced during the course of an unrelated study (D. Crompton, personal communication). Sequencing rationales for each fragment are shown (with full sequence obtained) in Appendix 5. Two fragments from the distal end of the 952 walk, a 1.0kb *Eco* RI - *Eco* RI fragment and a 3.2kb *Bam* HI - *Eco* RI fragment from the distal ends of λAGO02 and λAGO01 respectively were sequenced. A 3.2kb *Hin* dIII - *Hin* dIII fragment from the clone λR209 which is situated in the middle of the *runt* walk was also sequenced as was a 1.4kb *Sal* I fragment from the proximal end of λS2. The fragment sequenced by D. Crompton was from a 6.0kb *Eco* RI - *Eco* RI region situated within the 952 walk clone λ9405. All four 952 walk repetitive regions analysed are from Group I (see Section 4.2.2), *i.e.* they are very repetitive, whereas the *runt* walk fragment was placed in Group III *i.e.* it is likely that it contains a region which is weakly homologous to other sequences in the genome, or it contains a short repetitive region that is not long enough to incorporate adequate radioactivity to be visible upon autoradiography. Similar regions have been described previously (Hall *et al.*, 1983; see Section 4.3.5).

The 1.0kb λAGO02 and 3.2kb λAGO01 fragments were sequenced from both sides using T3 and T7 primers. The 3.2kb fragment from λR209 was separated into two fragments of 1.8kb and 1.4kb, by restriction with *Bam* HI and the sequences adjacent to the 'internal' *Bam* HI site determined using the relevant primer. The outer 'ends' were not sequenced as it was known that both the 1.8kb and 1.4kb regions were repetitive and thus it was likely that the repetitive sequence spanned the middle of the 3.2kb fragment. The 1.4kb λS2 fragment is from a highly repetitive region which has been mapped. This fragment was subcloned into pBluescript and sequencing reactions performed as described (Section 2.15.1).

The sequences obtained (Appendix 5), were used to search the GENBANK and EMBL databases (Devereux *et al.*, 1984) as described (Section 2.16). Both the 1.0kb λAGO02 and 3.2kb λAGO01 sequences displayed strong homology, over an approximately 100bp region, to a large number of developmentally important genes in a wide range of species. Upon further investigation both regions of homology were

shown to be due to the presence of *opa* repeats (Wharton *et al.,* 1985). The comparisons between the *opa* repeat, the 1.0kb λAGO02 sequence and the 3.2kb λAGO01 sequences are shown in Figure 4.07a and b. The 1.4kb λS2 sequence had no homology at the nucleotide level to any sequence in the databases but one of the TFasta (Section 2.16) searches with the frame 1 translation, detected low level homology to the *Ulysses* retrotransposon from *Drosophila virilis* (Figure 4.08). No significant homology was detected in the database search to any sequence from the 3.2kb λR209 region (in either the 1.4kb and 1.8kb fragment). The nucleotide sequence found in the 6.0kb *Eco* RI - *Eco* RI fragment, from the middle of the 952 walk, showed homology to a large number of sequences in the database. The homology between the sequence from p94.R1 and three of these is shown in Figure 4.9. This homology is restricted to approximately 150bp of the sequence and the similarities range from 60-80% at the nucleotide level.

After examination of the initial sequence obtained from the 1.0kb *Eco* RI subclone from λAGO02 it was decided to attempt to sequence the rest of the fragment. To do this, *Exo* III deletions were prepared across the 1.0kb insert . This sequencing was done for two main reasons. Firstly it was necessary to show that the repetitiveness of the fragment was due solely to the *opa* repeat and no other repetitive sequence existed within the fragment. This was because the region was used as a probe to detect other *opa* repeats in the walks and it was therefore necessary to confirm that no other repetitive sequences would be detected. Secondly the sequencing was also performed to determine if the *opa* repeat was situated in an open reading frame. The sequencing rationale and sequence itself is shown in Appendix 5. The majority of the plasmid was sequenced using these overlapping *Exo* III deleted transformants. One region from the middle of the clone could not be sequenced however. The reasons for this are unknown but it was thought likely that the *Exo* III enzyme had 'chewed back' past the 3′ overhang which was supposed to act as a block to the exonuclease activity (*i.e. Exo* III can delete the overhanging strand from a 5′ DNA overhang but not the 3′ overhang). This 'chewing back' therefore deleted the priming sites. Other sequencing primers were used, none however would give satisfactory sequence implying the enzyme had removed most, if not all of the polylinker. From the size of the fragment on agarose gels and from the amount of

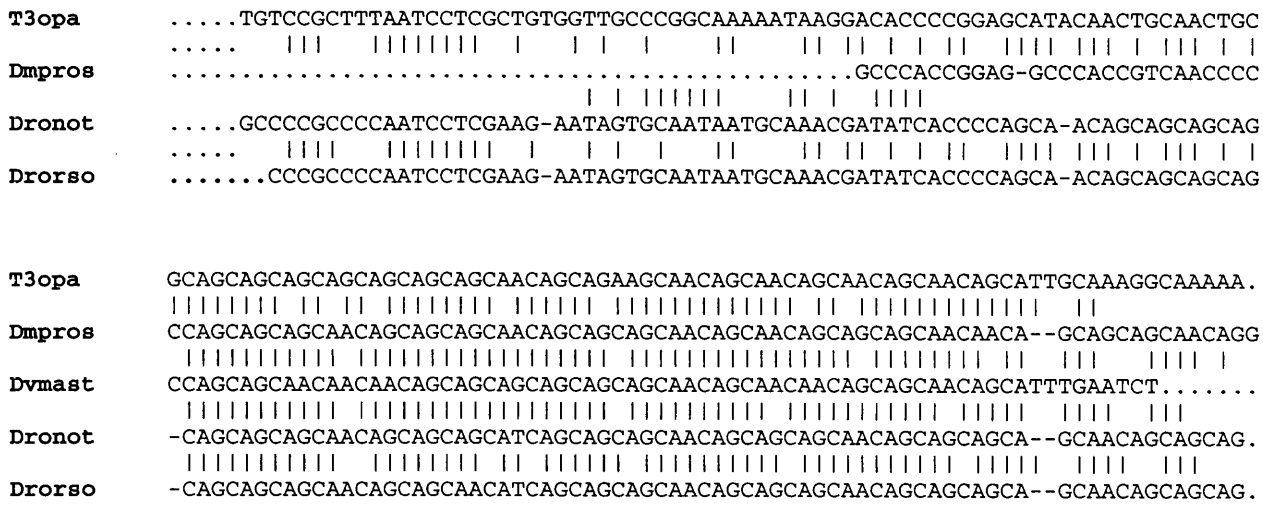


Figure 4.07a-Regions of sequence homology between the *opa* repeat within the 1.0kb *EcoR* I -*EcoR* I fragment from λAGOO2 and *opa* repeats from *prospero* (**Dmpros**); *mastermind* (**Dromas**); *notch* (**Dronot**) and the original isolate of the repeat (**Drorso**). (|) indicates identity to the *notch opa* repeat. The overall identity varies from *ca.* 75-80%.

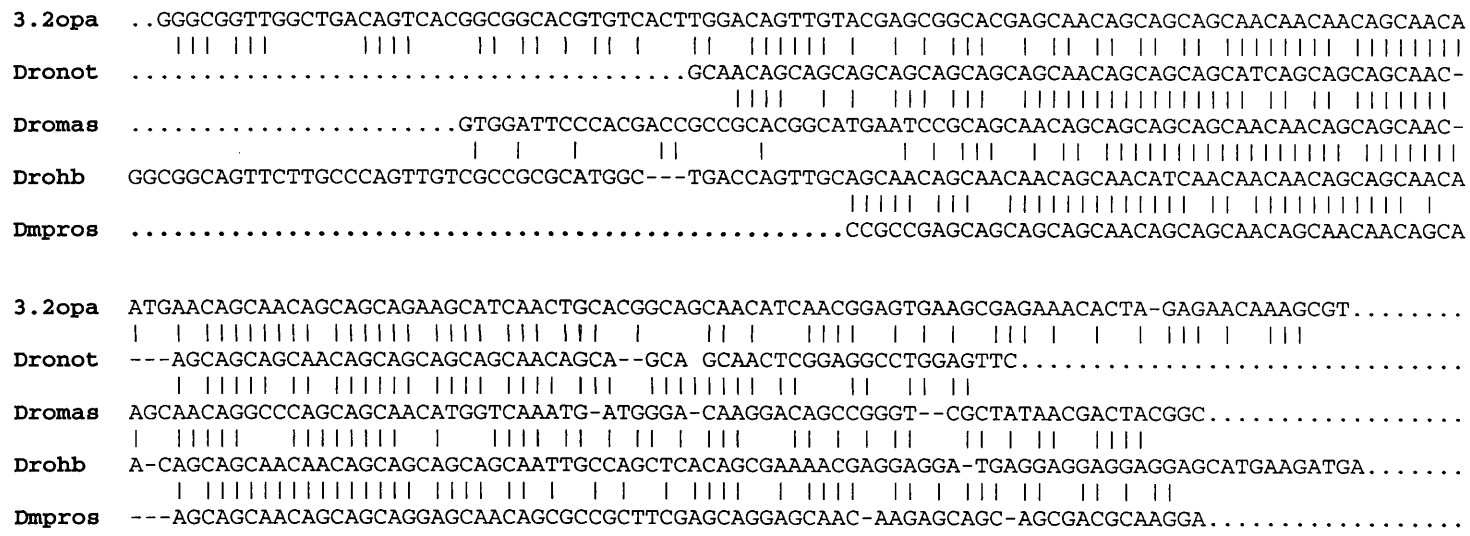


Figure 4.07b-Regions of sequence homology between the *opa* repeat within the 3.2kb *Bam* HI-*EcoR* I fragment from λAGOO1 and *opa* repeats from *notch* (**Dronot**); *mastermind* (**Dromas**); *hunchback* (**Drohb**) and *prospero* (**Dmpros**). (|) indicates identity to the *hunchback opa* repeat. The overall identity is *ca.* 75% in each case.

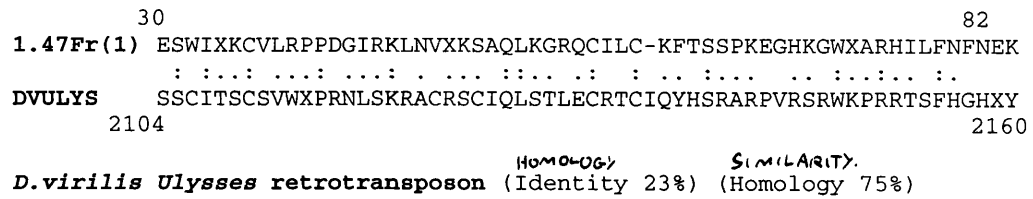


Figure 4.08--Homology of the Frame 1 translation (**1.47Fr(1)**)of the 1.45kb λS2 *Sal* I fragment with the *D. virilis, Ulysses* retrotransposon (**DVULYS-**Evgen'ev *et al.*, 1992). No other homologies were detected with the other five frames searched or at the nucleotide level. The homology is within the coding region of the element.

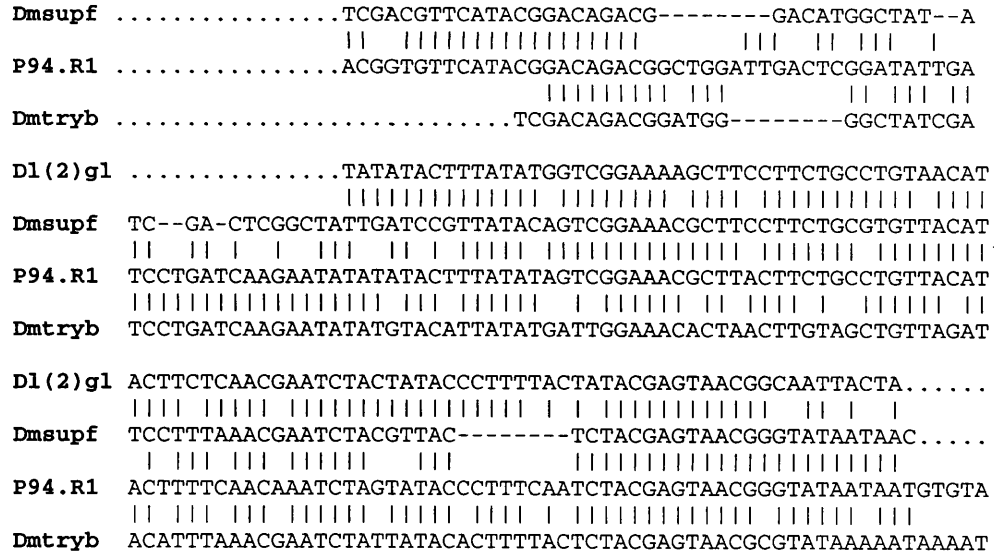


Figure 4.09--Homology of the repetitive sequence from the genomic clone λ9405, to *lethal (2) giant larvae*, (**Dl(2)gl-**Mechler *et al.*, 1985); *suppresser of forked* (**Dmsupf-**Langley *et al.*, 1993) and *trypsin alpha* (**Dmtryb-**Magoulas and Hickey, 1992). (|) indicates identity to the p94.R1 sequence. The levels of homology vary from 70-80%.

sequence obtained, it was estimated that 150bp of sequence could not be determined because of this.

4.2.5 Hybridisation of Previously Characterised Transposable Elements to the Chromosomal Walks.

The inserts of the transposable element clones listed in Table 2.3a were isolated and labelled by random priming. The probes were used to challenge Southern blots of restricted genomic phage clones from all four walks. A dilution series of the transposable element, from 1ng - 0.1pg, was dotted out onto nitro-cellulose as a positive control for hybridisation. The transposable elements, *Doc*, *pogo*, *P*, *mdg1*, *jockey*, *Springer*, *mdg4*, *I* and *hobo* failed to hybridise to any sequences within the walks. The *copia* probe had homology to a 1.4kb *Eco* RI - *Kpn* I fragment from the λ896-C21 phage clone and to a group of fragments at the proximal end of the 798 walk (Figures 4.11c and d). The *412* and *297* elements had homology to the same 2.4kb *Eco* RI - *Eco* RI fragment from the λM281 clone (Figures 4.11a and b). The *opa* repeat hybridised to several fragments from each walk to varying extents (Figures 4.12a and b).

4.3 Discussion.

4.3.1 Repetitiveness of DNA Within the Four Walks.

Other studies at the base of the X chromosome in *Drosophila melanogaster* have identified "thick nests of repetitive sequences" in which several walks have terminated (*e.g.* Miklos *et al.,* 1984;1988; Russell *et al.,* 1992; Langley *et al.,* 1993; Mitchelson *et al.,* 1993). No large nests of repetitive sequences have been identified in this study, with the possible exception of the region between the 952 and 896 walks (see Section 3.4). However, 58% of the genomic phage clones isolated from 19E1-19E4 / 5 do contain repetitive regions of varying extents and degrees of repetition (see Figure 4.01a, b, c and d). 100% of the genomic phage clones from the 798, *runt* and 896 walks contained restriction fragments that had repetitive sequences within them, as judged by the reverse Southern analysis. Only 26% (4 out of 15) of the clones from the 952 walk were found to be repetitive.

The differences between the numbers of, and types of repetitive sequences between the four walks (summarised in Table 4.2) is of note. Of the eight repetitive regions

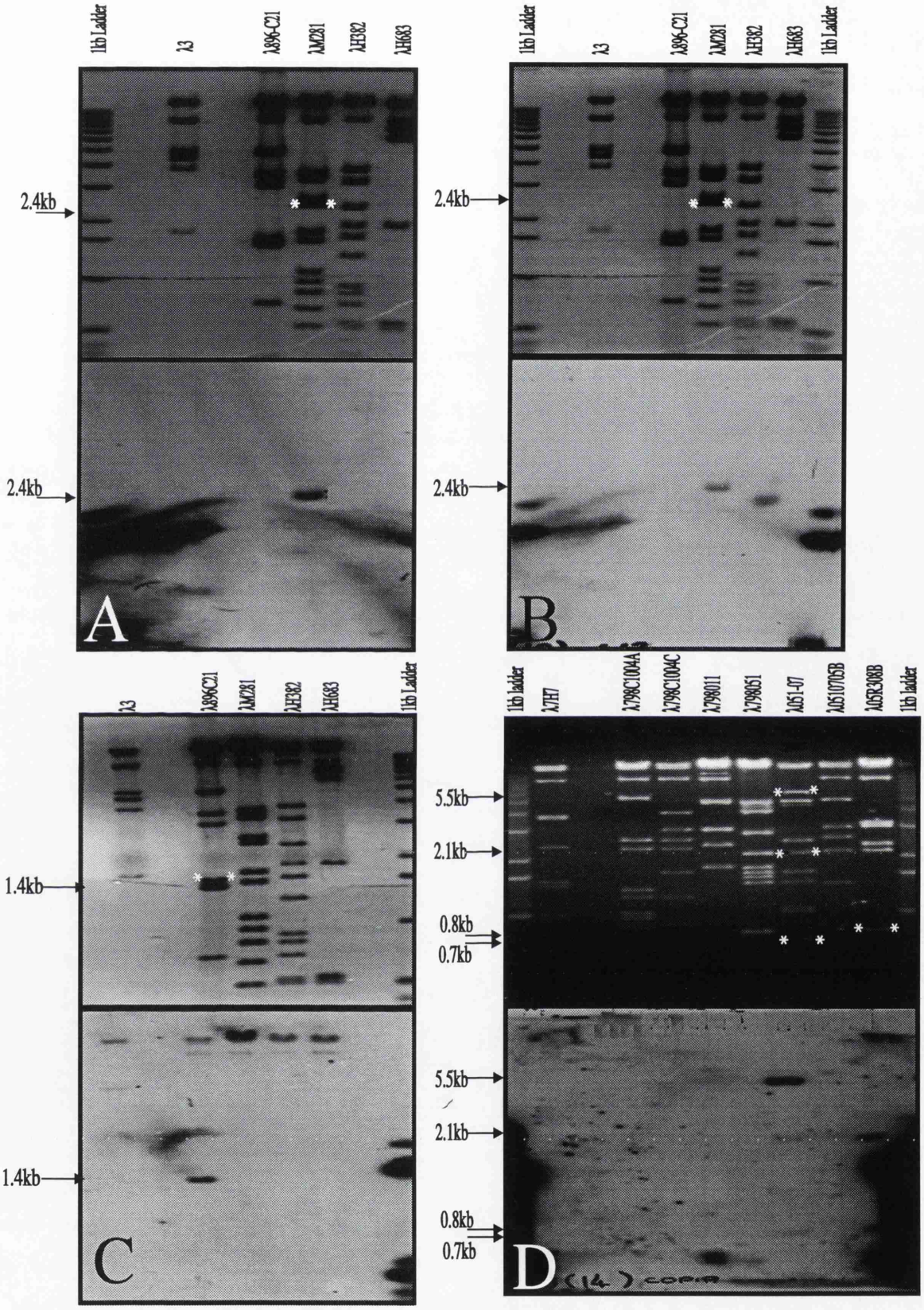


Figure 4.11 -Hybridisation of various transposable elements to the chromosomal walks. A) *412* element to the 896 walk; B) *297* element to the 896 walk, (There is some background hybridisation to the nitrocellulose filter but only a single band is apparent on each blot); C) *copia* element to the 896 walk; D) *copia* element to the 798 walk. See text (4.2.5) and appendices 2 and 3 for details of locations. Restriction fragments to which the elements are hybridising are marked with asterixs.

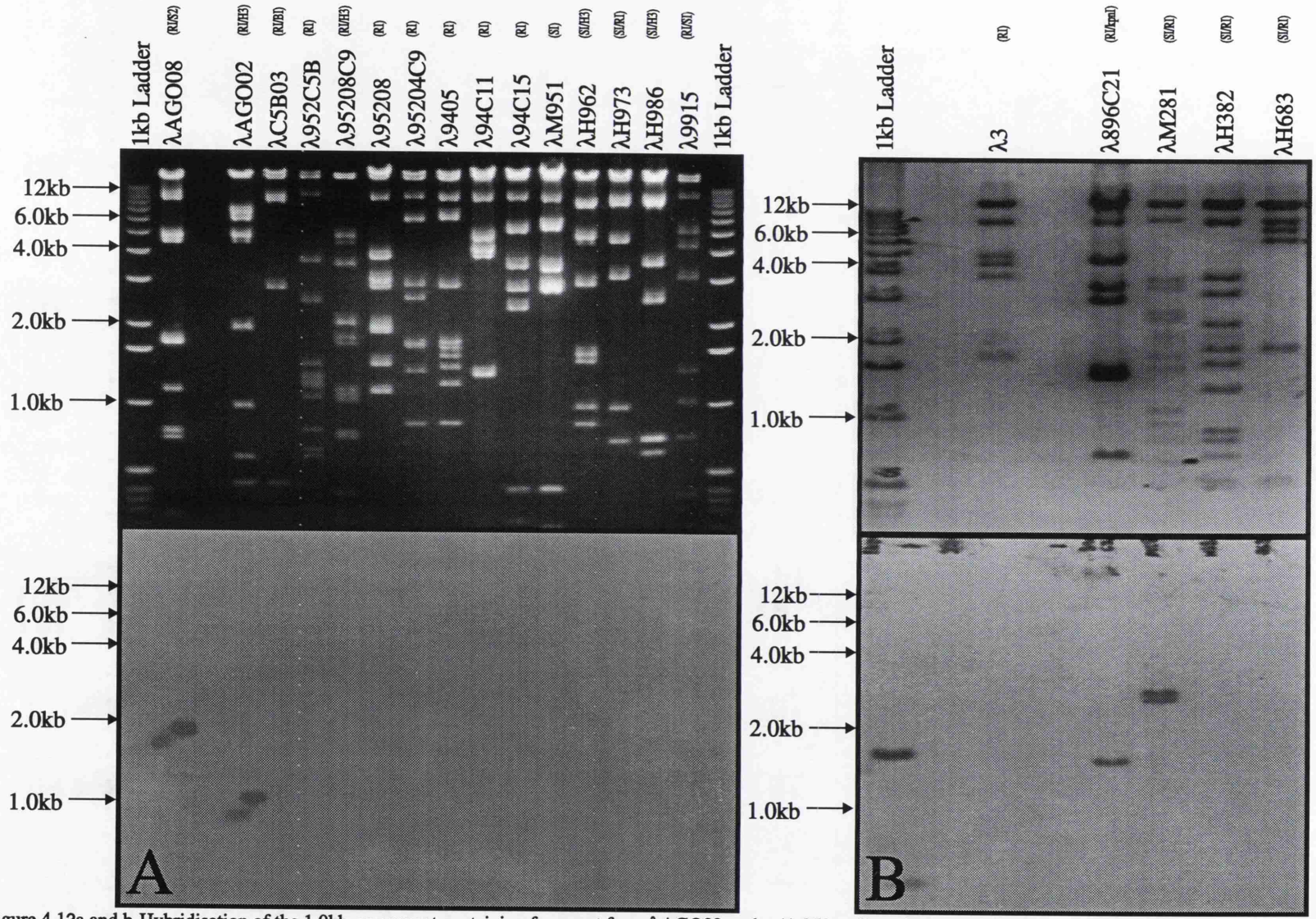


Figure 4.12a and b-Hybridisation of the 1.0kb *opa* repeat containing fragment from λAGO02 to the A) 952 walk and B) 896 walk. The enzymes used to restrict the phages are shown in brackets. See text (4.2.5) and appendicies 1 and 2 for locations of the hybridising restriction fragments .

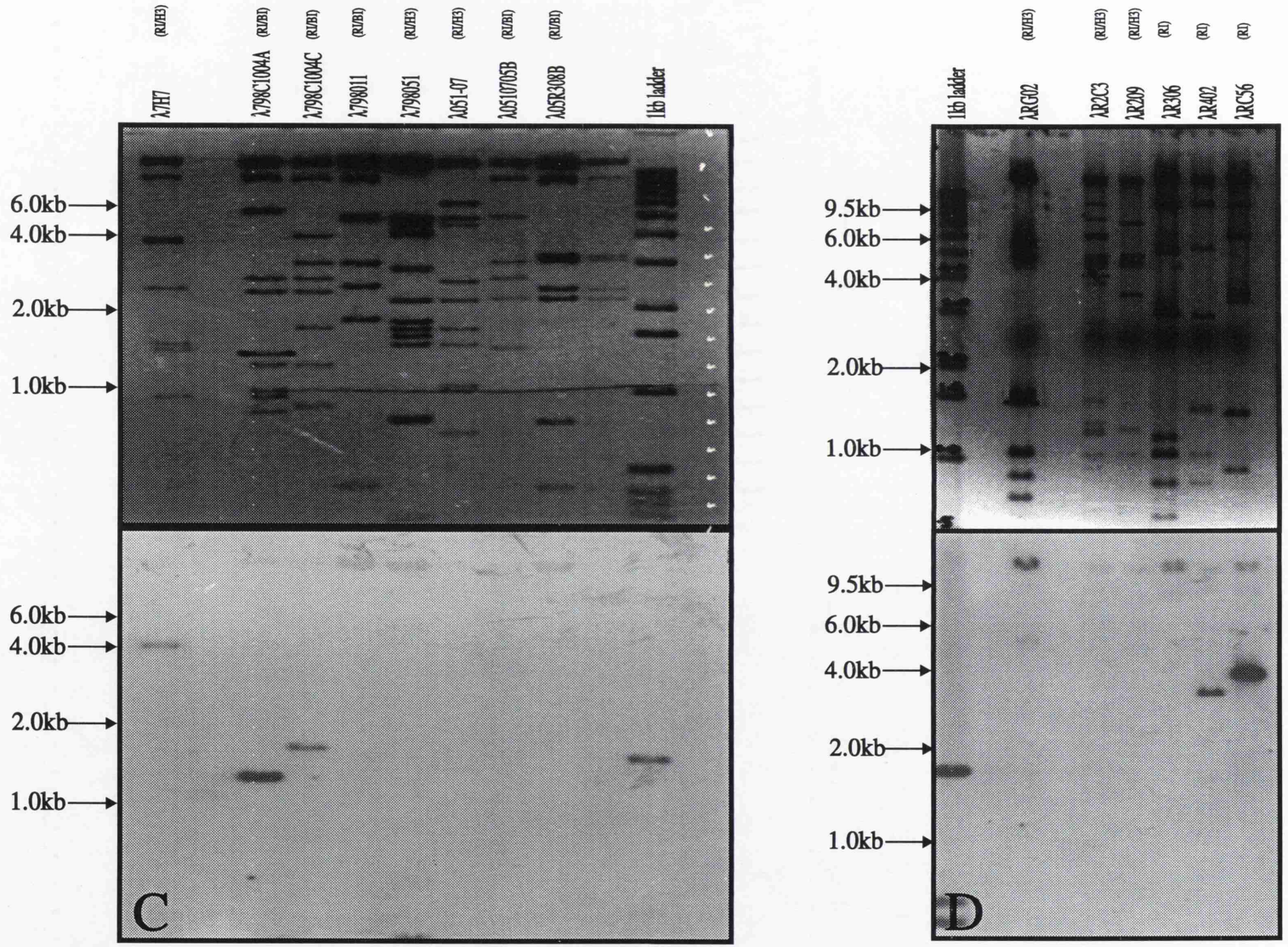


Figure 4.12c and d-Hybridisation of the 1.0kb *opa* repeat containing fragment from λAGO02 to the C) 798 walk and D) *runt* walk. The enzymes used to restrict the phages are shown in brackets. See text (4.2.5) and appendicies 3 and 4 for locations of the hybridising restriction fragments.

from the 798 walk that were hybridised to genomic DNAs, two regions were shown to fall into Class I (yielded more than 15 bands upon hybridisation to restricted genomic DNA), four into Class II (yielded between 4 -14 bands) and two were shown to fall into Class III (yielded less than three bands). The *runt* walk contains both Class I and III fragments but none from Class II, whereas in contrast the 952 and 896 walks contained fragments that only fall into Class I and III respectively. Thus the amount and type of repetitive DNA changes across the 19E subdivision. High, middle and low repetitive sequences occur relatively frequently across the 798 walk (at *ca.* 19E1); just low and highly repetitive sequences exist at *ca.*19E2. Whilst comparatively less frequently occurring but more highly repetitive sequences exist at 19E3, whilst frequently occurring low copy number sequences exist at 19E4. It is therefore possible to think of the molecular landscape of the region as a mosaic, with single copy sequences lying next to either low, medium or highly repetitive sequences. This lack of homogeneity in the microgenomic organisation across the 19E sub-division, has been observed previously; Miklos *et al.,* (1984) noted that commonly occurring middle and highly repetitive DNA exists at 19E8 whilst a chromosomal walk located at 19F6, six bands away, "extended for over 100kb without encountering any repetitive sequences of note". This lack of homogeneity aside, many other studies which have carried out chromosomal walks near to centromeric regions have reported repetitive sequences (see section 1.4). It is interesting that the frequency of the occurrence of the repeats in the walks does appear to have a correlation with the number of known single copy transcripts it also contains. For example, there are at least seven single copy transcribed regions across the 19E3 walk (see Chapter Five). None were identified in the other walks (although this may have been because they were not studied in as much detail). Therefore, as discussed in Chapter 5 it is possible that a large part of the 952 walk contains transcripts or their regulatory regions. Recombination events (or transposition events) that would bring repetitive DNA into the area would be very rare as the majority of them would be unstable as they would impinge upon at least one of the transcripts and would thus be probably deleterious for the fly. It should be noted however that this study did not distinguish *bone fide* transcripts that contain repetitive DNA, (in either their transcribed sequences or their introns), from transcribed repetitive elements. This is one possible reason why a single copy transcript is not apparent in the 798 walk, as chromosomal deficiency endpoints localised to the walk (*Df(1)mal*^{10}, *Df(1)GA40* and *Df(1)N77*) would suggest that the

Name of chromosomal walk.	**798**	***runt***	**952**	**896**
Approximate chromosomal location.	19E1-19E2	19E2-19E3	19E3	19E4
Amount of cloned DNA.	90kb	60kb	196kb	73kb
Number of lambda clones containing repeats.	8 / 8	5 / 5	4 / 15	5 / 5
Number of distinct repetitive regions	8	4	5	5
Amount of repetitive DNA (contained in *EcoR* I fragments)	46kb	21kb	29kb	23kb
Average distance between repetitive regions	7kb	10kb	31kb	9kb
Type of repetitive DNA	Class II and I	Class III	Class I	Class III

Table 4.2 Information concerning repetitive DNA within the four chromosomal walks in this study. The types of repetitive sequences were placed into three classes. Class I sequences produced >15 bands upon hybridisation to genomic Southerns, Class II had 3-14 bands, whereas Class III sequences had less than 3 bands.

melanized like (*mell*) locus (Section 3.) is contained in the region. *mell* however has only been defined in females, *trans* heterozygous for the two overlapping deficiencies *Df(1)mal*^{10} and *Df(1)N77*. It is not possible to tell whether *mell* is a separate locus or whether the phenotype is caused by the chromosomal breakpoints encroaching upon existing gene landscapes (Perrimon *et al.*, 1989b). It may be that the walk contains only regulatory sequences which would be undetectable by this study.

4.3.2 Identity of the Repetitive Sequences.

The 19E subdivision is a region where certain transposable elements preferentially re-enter the genome (Berg *et al.*, 1980; Bingham *et al.*, 1982; Green and Miklos, 1983; Ananiev *et al.*, 1984; Hey, 1987; Miklos and Costell, 1990). It is thought that areas such as 19E may be 'hot spots' for re-entry because of macro-structural features, such as chromosomal coiling, which guide transposable elements to the region. Micro-structural features, such as the target sequences, then allow the element to re-enter the genome (Craigie, 1992). In most regions the transposable element is able to jump out again, but in subdivisions 19 and 20 it appears that the element becomes 'stuck'; the reasons for this are unclear, although this may be a reflection of the lack of recombination at the base of the X chromosome which means that the effects are not deleterious *i.e.* recombination does not occur between elements and so chromosome abnormalities do not occur. In other regions recombination events would eventually remove the element; near to the centromere however recombination events are rare so that element numbers build up. Work by Sniegowski and Charlesworth (1994) suggests that natural selection actually opposes the build up of large numbers of transposable elements in euchromatic areas. If the number of transposable elements does build up, ectopic meiotic exchanges can occur between them *i.e.* exchanges can occur between regions of chromosome containing transposable elements. Such recombination events are nearly always deleterious and the fly die or fail to breed. Only those flies without large numbers of elements transposed to the euchromatin can survive and so most of the elements are maintained in their original, non-transposed locations. This constrains element numbers by reducing the fitness of those genomes with higher numbers of elements. This model would therefore predict a build up of transposable elements in regions with reduced levels of meiotic exchange, such as the base of the X chromosome.

31% of the repetitive sequences used in this study are thought to have homology to transposable elements. This homology was shown by hybridisation (Section 4.2.5); by nucleotide and amino-acid homologies (Section 4.2.4); or by inference from restriction pattern differences between strains (Section 4.2.3). Previous studies in this region have also identified a number of repetitive sequences as belonging to transposable element families. For example Miklos *et al.,* (1984;1988) identified *297*, *hobo* and *I* element sequences in microclones. Similar results have been reported by Ananiev *et al.,* 1984; Healy *et al.,* 1988; Yammamoto *et al.,* 1990; Russell *et al.,* 1992.

Other types of repetitive sequences have also been identified by other studies in this region. For example several tRNA genes were found at 19F1 / 2 (Russell *et al.,* 1992) and a cluster of type I-like sequences were identified at 19E8 (Miklos *et al.,* 1984;1988; Healy *et al.,* 1988). It is also known that subdivision 20 does not harbour 'significant amounts of the major satellite DNAs' (Yammamoto *et al.,* 1990) though it does contain a large proportion of repetitive non-satellite DNA sequences of uncertain origin (Young *et al.,* 1983; Donnelly and Keifer, 1986;1987; Miklos *et al.,* 1988; Yammamoto *et al.,* 1990).

Some of the regions in this study were shown to contain more than one type of repetitive DNA. Several regions from each walk contained *opa* repeats (Section 4.2.5) and some of these regions also possessed homology to transposable element clones used in this study. Table 4.1 lists the homologies identified within each region. The 1.4kb λS2 fragment that was sequenced had a low level of homology to the *Ulysses* retrotransposon at the amino acid level. There are several stop codons in the putative amino acid sequence. This does not prohibit it being an ancestral transposable element origin, however, as the remnants of a retrotransposon would be expected to accumulate multiple sequence changes and so the presence of stop codons are not surprising. The 2.4kb *Eco* RI - *Eco* RI fragment from λM281 had homology to both the *412* and *297* element probes used, as well as to the *opa* repeat probes (Section 4.2.5). It has been reported that the *297* element long terminal repeat (LTR) has extensive similarity with the *17.6* element LTR. However, no homology has been reported in the literature between the *297* and *412* elements. Nor could homology be detected between the *opa* repeat sequences reported here and the *297* and *412*

sequences currently in the sequence database. Thus, these three sequences exist within a 2.4kb fragment. The *opa* repeat is quite small (*ca.* 100bp), however complete *412* and *297* elements are 7.6kb and 7.0kb respectively (Finnegan and Fawcett, 1986), It follows therefore, that only fragments of these two elements can exist in this region. It is known that transposable elements have a tendency to insert in or near to other repetitive sequences (*e.g.* Dawid *et al.,* 1981; Scherer *et al.,* 1982; Di Nocera and Dawid, 1983; Young *et al.,* 1983) and it has been suggested that a large proportion of the β-heterochromatin arose in such a fashion *i.e.* "it represents the end product of insertion, deletion and amplification of mobile elements into each other as well as into regions containing unique sequences" (Miklos *et al.,* 1988). The discovery of three repetitive sequences in one small fragment would support this idea.

It is likely that the hybridisation observed to the 1.0, 1.6, 6.0kb fragments from λ05107 and with the 0.8kb, 3.1kb fragments from λ05R308B is also to fragments of an element, in this case the *copia* element. This is the only explanation that could account for the results obtained when these regions were hybridised to genomic Southern blots (Figures 4.04f and g). The *copia* element has an estimated 60 copies in the genome (Table 2.3a and Berg *et al.,* 1990), therefore a multitude of bands were expected. However, only restriction fragment sizes corresponding to those found in the genomic clones were observed (although several extra bands are visible with the λ05107 fragments). It is therefore likely that the homology detected is to small fragments of the *copia* element, a greater amount of which is located in the 6.0kb fragment. This would explain why a much greater signal intensity was observed when this 6kb fragment was challenged with the *copia* probe (Figure 4.11D). This would also explain why those hybridisations that included the 6kb fragment (Figure 4.04F) detected multiple signals on a genomic Southern, whilst hybridisation with a region homologous to the 1.6kb and 1.0kb fragments (a 5.2kb *Eco* RI fragment from λ798-051 (Figure 4.04E) which it is proposed contains less of the element sequence available to incorporate radioactivity, hybridised only to restriction fragments of the sizes corresponding to those expected from the fragments if they were single copy.

Several regions gave only the expected bands when hybridised to genomic Southern blots. It is likely that these contain sequences that are either very diverged, or that possess repetitive regions that are too short to incorporate enough radioactivity to be

visible when they hybridise to other bands. However, this short sequence must be sufficiently abundant, such that when the whole genome is randomly labelled enough sequences incorporate the label to allow the fragment to be detectable upon autoradiography. Similar repetitive sequences yielding only single copy bands have been reported previously (Pirotta *et al.,* 1983; Hall *et al.,* 1983; Miklos *et al.,* 1988; Russell *et al.,* 1992).

4.3.3 Correlation between Repetitiveness of a Region and the Number of Randomly Recovered, X-ray Induced, Deficiencies.

Cytogenetic studies (see Kramers *et al.,* 1983 for references) showed that divisions 19 and 20 are differentially susceptible to X-ray induced breakage (Lefevre, 1981). It was previously noted that there appeared to be a correlation between the repetitiveness of a region and the number of randomly recovered, X-ray induced, deficiencies (Miklos *et al.,* 1984). The results from this study also support such a correlation. The 798 walk at 19E1 / 2 is the most repetitive walk in this study and the region in which it lies has the highest number of X ray induced deficiency end-points. The reasons for this observation are uncertain. As mentioned previously Yammamoto *et al.,* (1990) suggest that genes encompassing long stretches of single copy DNA may be just as prone to breakage as an equally long stretch of repeated DNA. It is known that ectopic, meiotic, recombination between transposable elements can result in deficiencies (see Langley *et al.,* 1988, and Sniegowski and Charlesworth, 1994), and it has been postulated that similar mechanisms may contribute to the creation of several chromosomal features, such as the centromere (see Section 3.11.1 and Spradling *et al.,* 1993). If it is true that long stretches of unique DNA are more prone to breakage in some parts of the chromosome than in others *e.g.* the base of the X chromosome, then the events, or structures, that lead to these breakages are still unknown and must await further analysis.

4.3.4 The Small Repetitive Regions Identified by Sequencing.

4.3.4.1 The *opa* Repeats.

It is interesting to note that a large majority of the repetitive regions from the four walks contained sequences homologous to the *opa* repeat. It is thought unlikely however, that the degree of repetition detected in the Southern blot analysis (Section 4.2.2) was due solely to *opa* repeats for all of these fragments, unless there are

multiple copies of *opa* within the fragments. This is because it is presumed that the signal intensity observed when the fragments were hybridised to genomic Southern blots was too great to have been caused by the amount of radioactivity it would have been possible to incorporate into a single sequence, only 100bp to 150bp long, when this sequence is situated in at least ten times (and upto one hundred times) more surrounding DNA.

It would be worthwhile pursuing a sequence analysis of the region surrounding the *opa* repeats for several reasons. Firstly it would be interesting from the viewpoint of the *opa* repeats, as no research has been done on their occurrence in non-transcribed genomic DNA, relative to their occurrence in transcribed genomic DNA. Secondly, it is possible that at least some of the eleven *opa* repeat containing regions identified by this study, will be situated within a transcript.

No homology to the sequences obtained from the 1.0kb *opa* repeat containing region from λAGO02 at the distal end of the 952 walk (outwith the *opa* repeat itself) was detected in the database, thus this particular *opa* repeat may not be transcribed although it is situated in an open reading frame that extends for 135 residues and probably more as the coding frame continues to the end of the available sequence and thus continues an unknown distance past it (see Appendix 5 for sequence and open reading frame). Previous work has shown an extensive association of *opa* repeats with important, developmentally regulated genes which was one of the reasons why this *opa* repeat containing fragment was sequenced. The region was shown to be expressed at all developmental stages (except L3, section 5.4.1) however the actual pattern of expression may be being masked by the hybridisation of the *opa* repeat to the many *opa* repeat containing transcripts in the mRNA. It is known that a fragment just proximal to it contains a transcribed region that is differentially expressed in development (see Section 5.4.2), it is possible that the 1.0kb region contains a transcribed sequence other than the *opa* repeat. It may have been unfortunate that this particular *opa* repeat containing region from the 952 walk was chosen for analysis and not the one situated some 6kb proximal (and nearer to the single copy transcript in the 7.2kb *Eco* RI - *Eco* RI fragment from λAG001) as this *opa* repeat may be included in the transcript detected in the region. It will have to await the results of further research, on the occurrence of these repeats in non-coding and coding genomic DNA,

before the probability of finding an *opa* repeat in a transcribed region, as compared to a non-transcribed region, can be calculated.

4.3.4.2 The λ9405 Repetitive Sequence.

The 150bp region located within the 6.0kb fragment from the 952 walk is highly repetitive and has a wide range of homologies across the *Drosophila* group. This was shown by both the hybridisation results (Figure 4.02c) and the large number of "hits" in the database search. Nothing however, explains why it should be repetitive, as there are no obviously repetitive structures *e.g.* dinucleotide repeats or short regions like the *opa* sequences. In addition, there is no detectable homology to any transposable sequences. It is unlikely to have a structural function as the homologies to the database entries show it to be spread throughout the genome where it is located within transcripts, within introns, within non-transcribed DNA and next to a repetitive region that is specific for the *Drosophila virillis* phylad Zelentsova *et al.,* 1986; see Figure 4.09 and below for other references).

The database hits would seem to indicate that it is almost completely *Drosophila* specific. Added to this it is interesting to note that many of the homologies the database search identified, were within sequences mostly located near to, or actually within, regions of β–heterochromatin. For example, one of the homologies detected was to a tandem repeat in an intron of the *su(f)* gene (Mitchelson *et al.,* 1993; Langley *et al.,* 1993) which is the most proximal on the X chromosome; a second was to a region near to the Arg-tRNA genes at 42A on chromosome 2L. A third was to a region near to a tandem repeat from the *lethal (2) giant larvae* gene (Mechler *et al.,* 1985) which is the most distal gene on chromosome arm 2L. *Drosophila* telomeres are unusual compared to those in most other eukaryotes. Most eukaryotic telomeres contain short, simple repeats that are highly conserved. *Drosophila* does not have such sequences but carries at the ends of its chromosomes one or more *LINE*-like retrotransposable elements. Proximal to the terminal array of transposable elements are regions of tandem repeats that are structurally, (and probably functionally), analogous to the subterminal regions in other eukaryotes (Karpen and Spradling, 1992). The tandem repeat from *lethal (2) giant larvae* is probably one of these. It is possible therefore that this sequence from the λ9405 clone could be similar to repetitive sequences identified in several previous studies (*e.g.* Miklos *et al.,*

1984;1988; Yammamoto *et al.,* 1990) as being specific for certain types of repetitive DNA *e.g.* β–heterochromatin.

Heterochromatin was once thought of as 'junk' DNA but has since been shown to have several important roles in *Drosophila* and other organisms (see Section 1.4.1.1). In view of this it is probably premature to call *opa* repeats, or other repetitive sequences such as the sequence from λ9405, 'junk'. It would seem logical to assume that the fact that these repeats occur at such a high frequency within the genome suggests they have, (or may have had) a function for the fly. It is however possible that they exist as 'selfish' DNA, as has been postulated for transposable elements and their derivatives (Doolittle and Sapienza, 1980; Orgel and Crick, 1980; Charlesworth *et al.,* 1994) or have no function whatsoever.

Chapter Five

Chapter Five - Identification and Characterisation of Transcribed Regions Within the Chromosomal Walks.

5.1 Introduction.

Over 460kb of DNA has been cloned in this laboratory (see Chapter 3) from the 19E1 - 19E5 region at the base of the X chromosome. This region contains three lethal complementation groups, *runt*, formerly known as *legless*, (Kramers *et al.,* 1983; Gergen and Wieschaus, 1985), *R-9-28* (Lindsley and Zimm, 1992) and *shak-B*, (Baird *et al.,* 1990). It also contains at least one identified non-vital locus, *mal* (Schalet and Lefevre, 1976) and possibly a second, *mell* (Schalet and Lefevre, 1976). A deficiency and duplication breakpoint analysis (Chapter 3) indicated the probable locations of the four walks and showed where at least part of the lethal domain of the *shak-B* locus was situated. A more detailed analysis was then performed to identify regions of DNA transcribed at different developmental stages. This transcriptional analysis was performed primarily to identify transcripts from the *shak-B* locus, but would also identify previously uncharacterised transcripts from unidentified genes. This was thought to be an important consideration because although the base of the X chromosome has been saturated in screens for lethal complementation groups (see Section 1.3), it has probably not been saturated for non-vital loci which may possess subtle phenotypes which would have been 'invisible' in the screens used to date.

Several techniques exist for identifying transcribed regions within cloned genomic DNA. Genetic screens for open reading frames (Gray *et al.,* 1982), enhancers, (Weber *et al.,* 1984) and promoters (Allen *et al.,* 1988; Gossler *et al.,* 1989) are available but their value appears limited for identifying genes in long segments of cloned genomic DNA. The two most successful approaches to date involve either hybridising whole genomic phage clones to Northern blots (Section 2.5.8.2) or reverse Northern analysis. Reverse Northern analysis involves hybridising radiolabelled first strand cDNA prepared by reverse transcription of polyA$^+$ RNA, to Southern blots of genomic walk phages (see Sections 2.5.13.4, 2.5.16). Any transcribed region in genomic DNA will hybridise to its radiolabelled counterpart in the cDNA and so be detectable upon autoradiography (Section 2.2.6, 2.2.7). If RNA is isolated from different developmental stages of the fly's life cycle then it is possible to determine which parts of the chromosomal walk are expressed in what stage.

Three factors precluded the use of whole genomic phage clones, from the walks, as probes to challenge Northern blots. Firstly, the large numbers of genomic phage clones from the four walks would mean an unfeasible amount of work preparing and running Northern blots. Secondly, the use of whole phage clones as probes may cause problems with the detection of rarer transcripts. This is because of the difficulty in labelling the large segments of genomic DNA within each phage clone to an appropriately high degree of specific activity. Finally, the large number of repetitive sequences, known to be situated within the phage clones (Chapter 4), would give rise to spurious results if transcribed (*e.g.* Finnegan, 1989; Section 1.4.2.1).

It was therefore decided to carry out a reverse Northern analysis which, in conjunction with the repetitive DNA analysis, would show the locations of single copy transcribed regions. The regions identified could then be studied in more detail *i.e.* used as probes to screen cDNA libraries to isolate transcripts or to probe Northern blots. This approach has been used successfully (Hall *et al.*, 1983; Scott *et al.*, 1983; Knipple *et al.*, 1991) to identify transcripts within long segments of cloned DNA.

In this chapter, I shall discuss the characterisation of all four walks using this reverse Northern technique and the further characterisation of some of the transcribed regions in the 19E3 region (952 walk) using Northern blots.

5.2 Results.

RNA was prepared for use in a transcription analysis as described (Section 2.11) and the polyA$^+$ fraction isolated by oligo dT cellulose chromatography (Section 2.13). Northern blots were carried out under standard conditions (Section 2.2.8.2) as was the reverse transcription (Section 2.5.13.4) and associated procedures. All stages of the life cycle were used for this investigation with the exception of second instar (L2) larvae. It was decided to omit this stage due to the problems associated with reliably collecting L2s and also the low level of developmental gene activity that occurs during this time (Ashburner, 1992).

5.2.1 Reverse Northern Analysis.

The aim of this experiment was to produce radiolabelled first strand cDNA for use in identifying transcribed regions within the chromosomal walks. The walk phages were

digested with appropriate restriction enzymes (Section 2.5.3) and transferred onto nitrocellulose (Section 2.5.11). The nitrocellulose filters produced (hereafter called 'walk filters') were probed with radiolabelled first strand cDNA (Sections 2.5.16, 2.5.17) and autoradiographed for up to four weeks (Sections 2.2.2, 2.2.7).

5.2.2 Optimisation of Experimental Conditions.

5.2.2.1 Choice of Reverse Transcriptase.

Pilot experiments, using avian myeloblastosis virus reverse transcriptase (AMV RTase) and Moloney murine leukaemia virus reverse transcriptase (MMLV RTase), were performed to determine which enzyme gave optimal incorporation with optimal transcript length. Results from these pilot experiments, using standard conditions, contradicted published results in that in my hands MMLV RTase gave better incorporation and longer transcripts | MMLV Rtase was therefore used throughout the reverse Northern analysis.

5.2.2.2 Choice of Experimental Primer.

Control experiments were undertaken to determine the best method of priming first strand cDNA synthesis. Random hexamers have the benefit of priming throughout the length of the mRNA and not just at the 3' end, as is the case for oligo dT. This can therefore avoid the problems associated with secondary structure caused by GC rich regions. Oligo $dT_{12\text{-}16}$ mers have the benefit of only priming off $polyA^+$RNAs and so avoid transcription of contaminating ribosomal RNA. Both types of primer were used to prime cDNA production at concentrations of $0.6\mu gml^{-1}$ and $2\mu gml^{-1}$ respectively. Although both types of primers yielded probes of the appropriate size range (200 - 1000bp) and with approximately the same amount of incorporation, the probes primed with random hexamers always gave unacceptably high levels of background on hybridisation to the walk filters. No explanation could be found for this observation though it had been noted previously by several workers (S. Russell and S. Tomlinson personal communications). It was therefore decided to use oligo $dT_{12\text{-}16}$ mers as primers to initiate cDNA production.

It was discovered that the major factor determining the success of the reverse Northern analysis was probe length. Short probes, where the cDNA product was <400bp, failed to hybridise successfully to control DNA (see below) under standard

hybridisation conditions (5X SSC, 65°C). Probes with the majority of the product in the size range 200bp - 1000bp hybridised satisfactorily to the control DNA, thus this was the size range of the cDNA product used throughout this study. As probe size was so important controls for it were carried out for each hybridisation. A small amount of the cDNA probes were electrophoresed through alkaline denaturing gels (Section 2.5.8.3; Figure 5.02) together with appropriate marker end labelled with ^{32}P dATP (Section 2.5.13.3). Only those probes where the product was in the optimal size range were used.

5.2.3 Hybridisation Controls.

A hybridisation control (which doubled as a second control for probe length) was included in each experiment. This control consisted of a Southern blot of the *Drosophila melanogaster* Actin 5C plasmid restricted with *Hin* dIII. This control Southern blot was simultaneously hybridised to the cDNA probes with the walk filters. The Actin 5C gene was chosen as a positive control for reverse Northern analysis as it is known to be expressed in all tissues and it is expressed consistently throughout development (excepting a slight elevation in embryonic and early pupal tissues, Fyrberg *et al.*, 1983). The six *Drosophila melanogaster* actin genes all have highly homologous protein coding regions yet exhibit little homology in their 5′ and 3′ untranslated regions (Fyrberg *et al.*, 1981). The Actin 5C plasmid used was an 8.7kb *Eco* RI - *Eco* RI fragment originally isolated by Fyrberg *et al.*, (1981; 1983) in pBR322 and kindly donated to this study by Guy Wakefield, (Cambridge University). A *Hin* dIII digest yields fragments of 0.36kb, 1.60kb, 1.80kb and 4.9kb. The 1.60kb fragment contains a portion of the 3′ untranslated region, the 1.8kb fragment contains the remaining 3′ untranslated region, the complete protein coding region and a small part of the 5′ untranslated region. Both of these fragments should hybridise to radiolabelled first strand cDNA provided the probe is of an appropriate length, (*i.e.* longer than the 3′ untranslated region) and the hybridisation is successful (Figure 5.01A to G).

Hybridisation to actin also gives an approximate indication of the relative amounts of mRNA homologous to the transcribed region. Actin is taken as 1% of the total poly A^+ transcripts at any stage of the fly's life cycle (Berger and Kimmel, 1987), therefore a

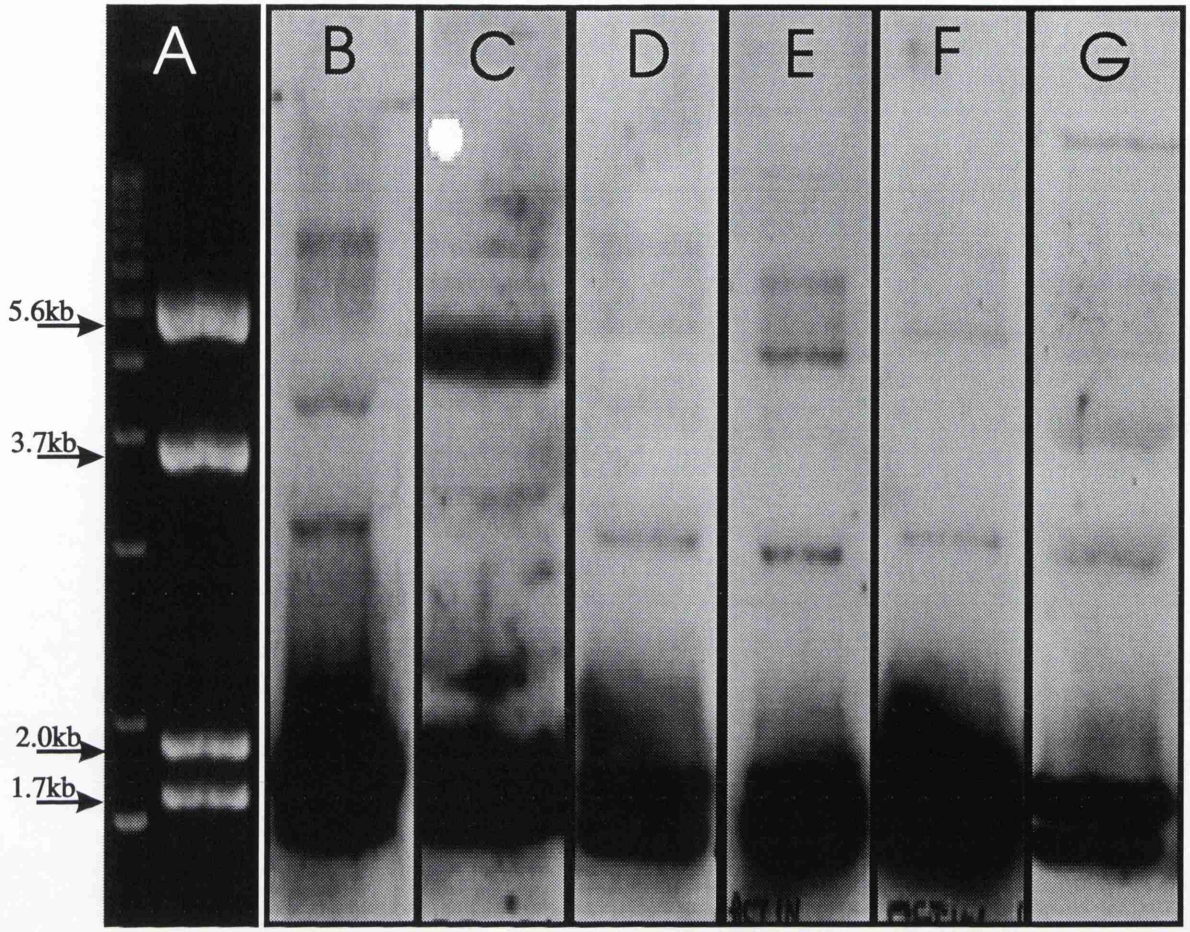


Figure 5.01-Ethidium bromide stained gels (A) and results of hybridising first strand cDNA to *Hin* dIII restricted, Southern blotted Actin 5C plasmid. The probes used are B=Embryonic cDNA; C=L1 cDNA; D=L3 cDNA E=Pupal cDNA; F=Adult cDNA; G=Body cDNA.

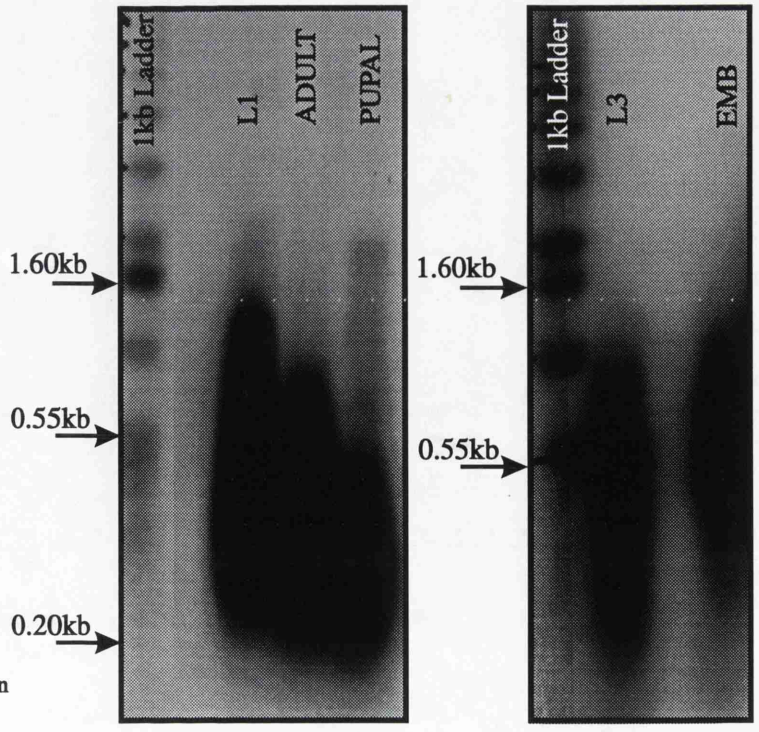


Figure 5.02-
Results of size fractionating first strand cDNA through alkaline denaturing gels. In all cases the majority of the product is between 200-1500bp.

direct comparison of the signal intensities produced from the Actin 5C and the transcribed region provides a rough estimate of the amount of radiolabelled transcript.

5.3 Detection of Genomic Regions Transcribed in Various Stages of the Flys Lifecycle.

Radiolabelled first strand cDNA was prepared from polyA$^+$ RNA from the relevant stage of development and from adult bodies and hybridised as described (Section 5.2 - 5.2.2). The Actin 5C Southern blots included as positive controls for the hybridisations are shown in Figure 5.01. The cDNA probes that were electrophoresed through an alkaline denaturing gel to determine their sizes prior to use are shown in Figure 5.02. Several bands were detected upon autoradiography from each walk filter (Figures 5.03, 5.04, 5.05, 5.06, 5.07 and 5.08). The results for the 952 walk are summarised in Table 5.1. Appendices 1, 2, 3, and 4 show the extent and location of each transcribed region. Each phage clone Southern blotted to similar walk filters was restricted with the same enzymes (**952 walk**: λAGO08 - *Eco* RI / *Sst* I; λAGO02 - *Eco* RI / *Hin* dIII; λC5B03 *Eco* RI; λ952C5B *Eco* RI / *Bam* HI; λ95208C9- *Eco* RI / *Hin* dIII; λ95208- *Eco* RI; λ95204C9- *Eco* RI; λ9405- *Eco* RI; λ94C11- *Eco* RI; λ94C15- *Eco* RI / *Bam* HI (In Fig 5.03A= *Eco* RI / *Hin* dIII);; λM951- *Eco* RI / *Sal* I; λH962-*Hin* dIII / *Sal* I (In Fig 5.08A= *Eco* RI / *Hin* dIII); λH973- *Eco* RI / *Sal* I; λH986- *Eco* RI / *Sal* I; λ9915- *Eco* RI / *Sal* I. **896 walk**: λ3-*Eco* RI / *Hin* dIII; λ896C21- *Eco* RI / *Kpn* I; λM281- *Eco* RI / *Sal* I; λH382- *Eco* RI / *Sal* I; λH683-*Eco* RI / *Sal* I. **798 walk**: λ7H7- *Eco* RI / *Bam* HI; λ7A1- *Eco* RI/ *Bam* HI; λ798C10O4A- *Eco* RI / *Bam* HI; λ798C10O4C- *Eco* RI / *Bam* HI; λ798011- *Eco* RI / *Bam* HI; λ798051- *Eco* RI / *Hin* dIII; λ05107- *Eco* RI /*Hin* dIII; λ0510705B-*Eco* RI / *Bam* HI; λ05R308B- *Eco* RI / *Bam* HI. ***runt* walk**: λRG02 - *Eco* RI / *Hin* dIII; λR03 - *Eco* RI / *Hin* dIII; λR04C - *Eco* RI / *Hin* dIII; λR209- *Eco* RI / *Hin* dIII; λR2C3 - *Eco* RI / *Hin* dIII; λRC42A - *Eco* RI / *Bam* HI; λR306- *Eco* RI / *Bam* HI; λR402 - *Eco* RI / *Bam* HI; λRC56- *Eco* RI / *Bam* HI. All bands that are transcribed and were detected by this study (including very faint ones) are marked with asterixes.

The clone labelled λ95208 in Figure 5.04A is not λ95208. However, as only 1.2kb of the walk was not included in the L1 analysis (due to overlap with the proximal and distal lambdas), it was decided to omit this clone from this part of the analysis. The aliquot of the λM951 clone used in Figures 5.05A and 5.06A was

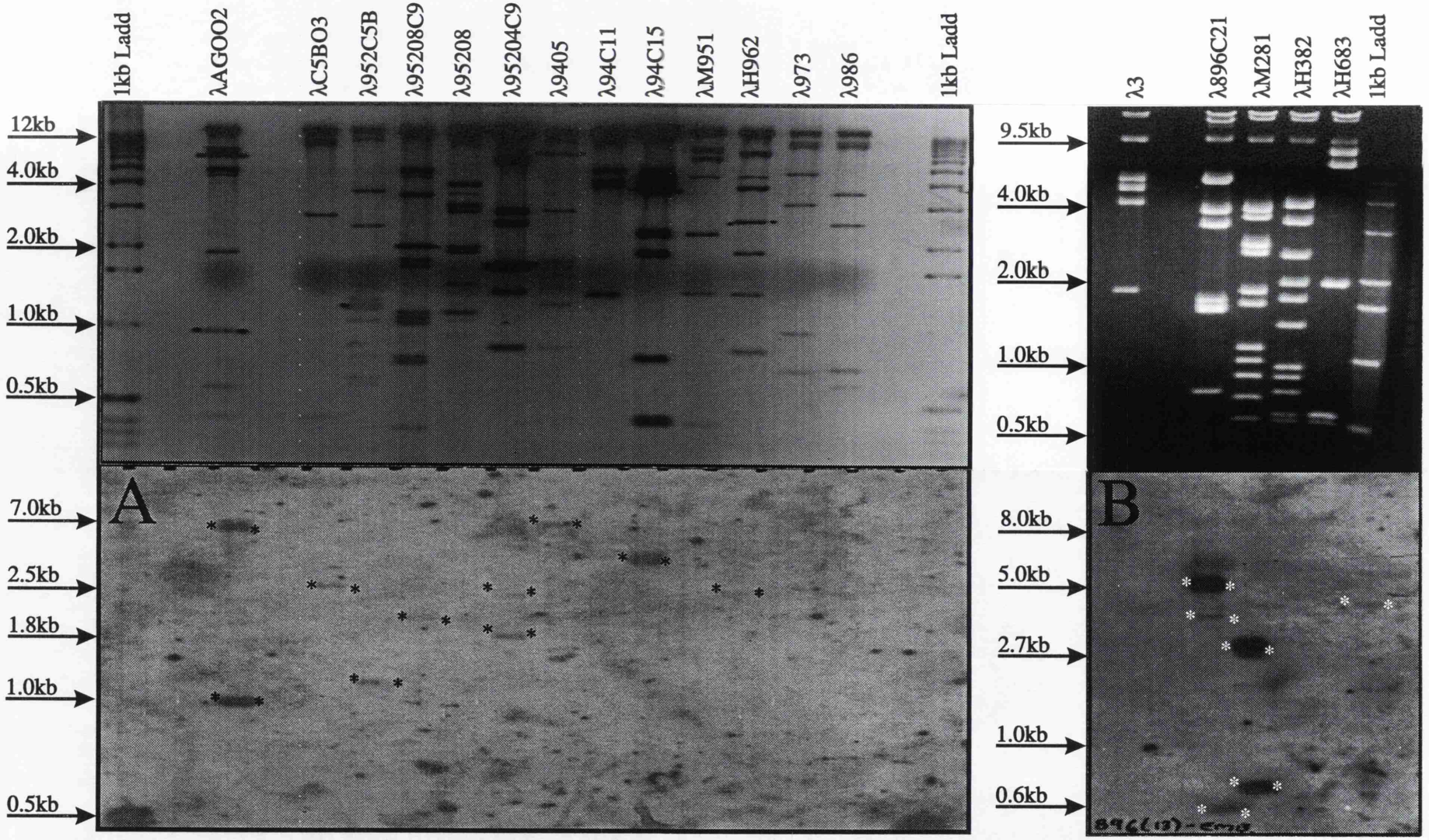


Figure 5.03A and B-Ethidium bromide stained gels and corresponding autoradiographs of the 952 (A) and 896 (B) walk genomic clones digested with appropriate enzymes (see text), Southern blotted and then hybridised with radiolabelled embryonic first strand cDNA. All of the hybridising restriction fragments are marked with asterixs. The results are summarised in Appendices 1 and 2. The 952 walk gel image is a negative and therefore looks different from the 896 gel photograph which is a conventional polaroid.

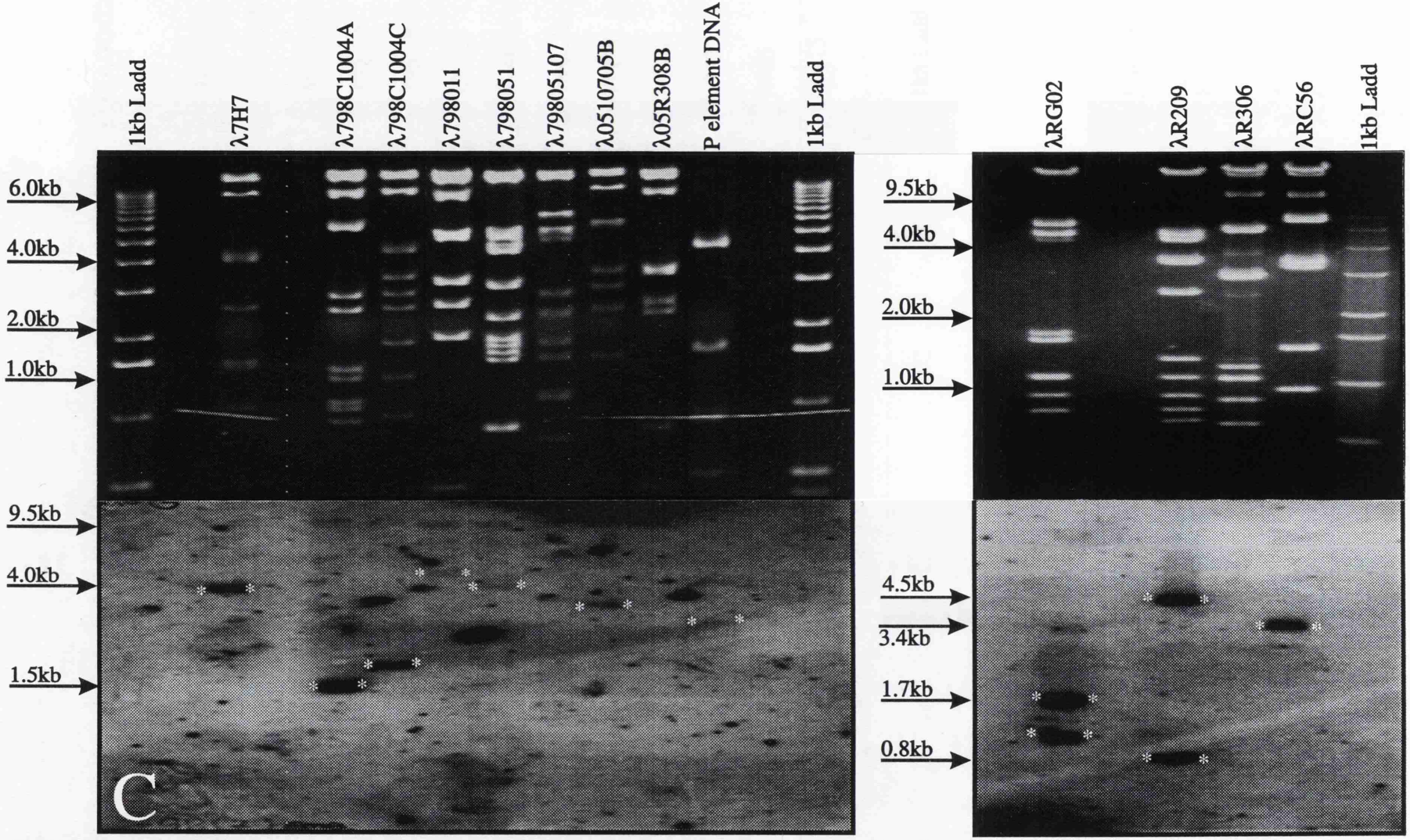


Figure 5.03 C and D-Ethidium bromide stained gels and corresponding autoradiographs of the 798 (C) and *runt* (D) walk genomic clones digested with appropriate enzymes (see text), Southern blotted and then hybridised with radiolabelled embryonic first strand cDNA. All of the hybridising restriction fragments are marked with asterixs. The results are summarised in Appendices 3 and 4.

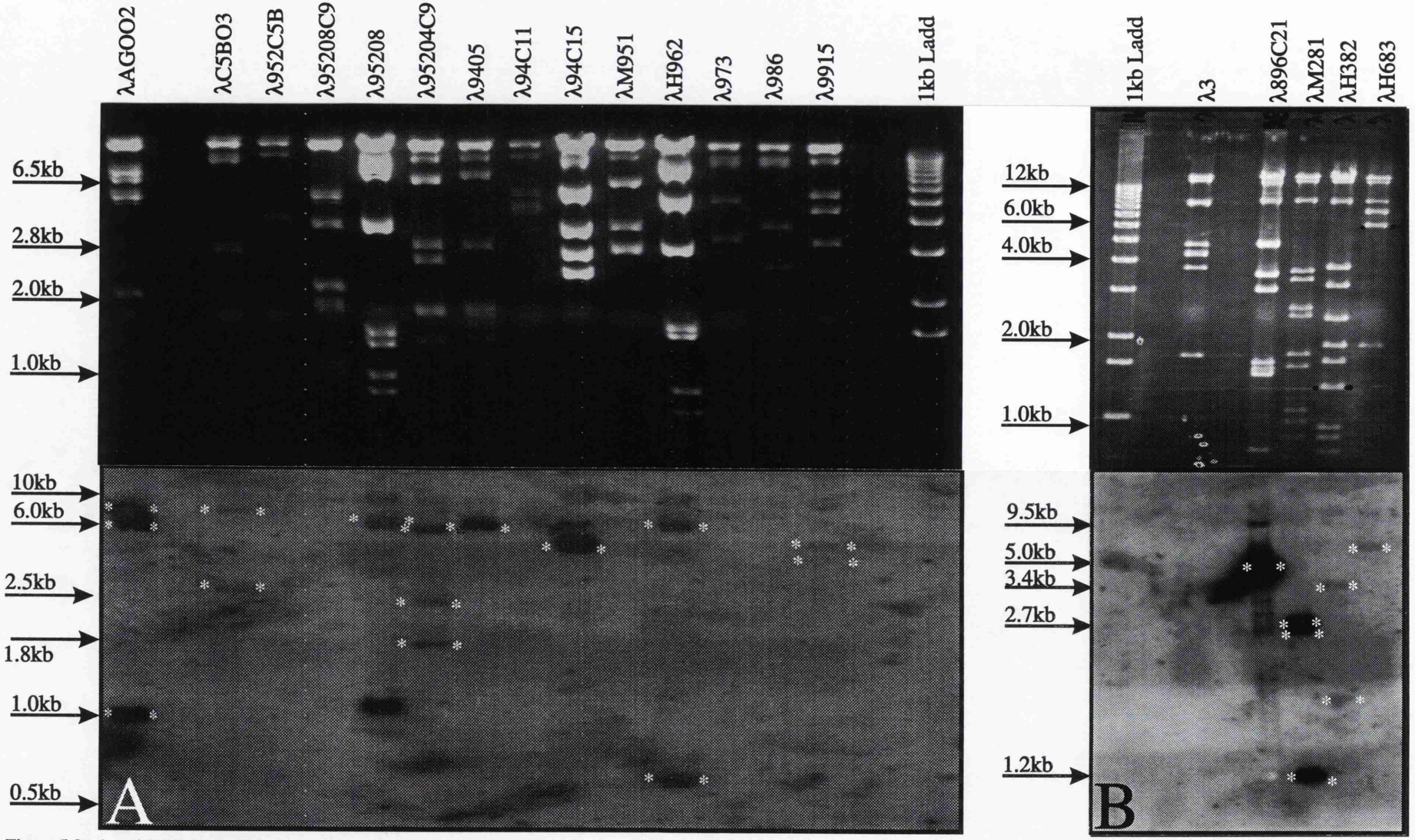


Figure 5.04 A and B-Ethidium bromide stained gels and corresponding autoradiographs of the 952 (A) and 896 (B) walk genomic clones digested with appropriate enzymes (see text), Southern blotted and then hybridised with radiolabelled first instar larval first strand cDNA. All of the hybridising restriction fragments are marked with asterixs. The results are summarised in Appendices 1 and 2.

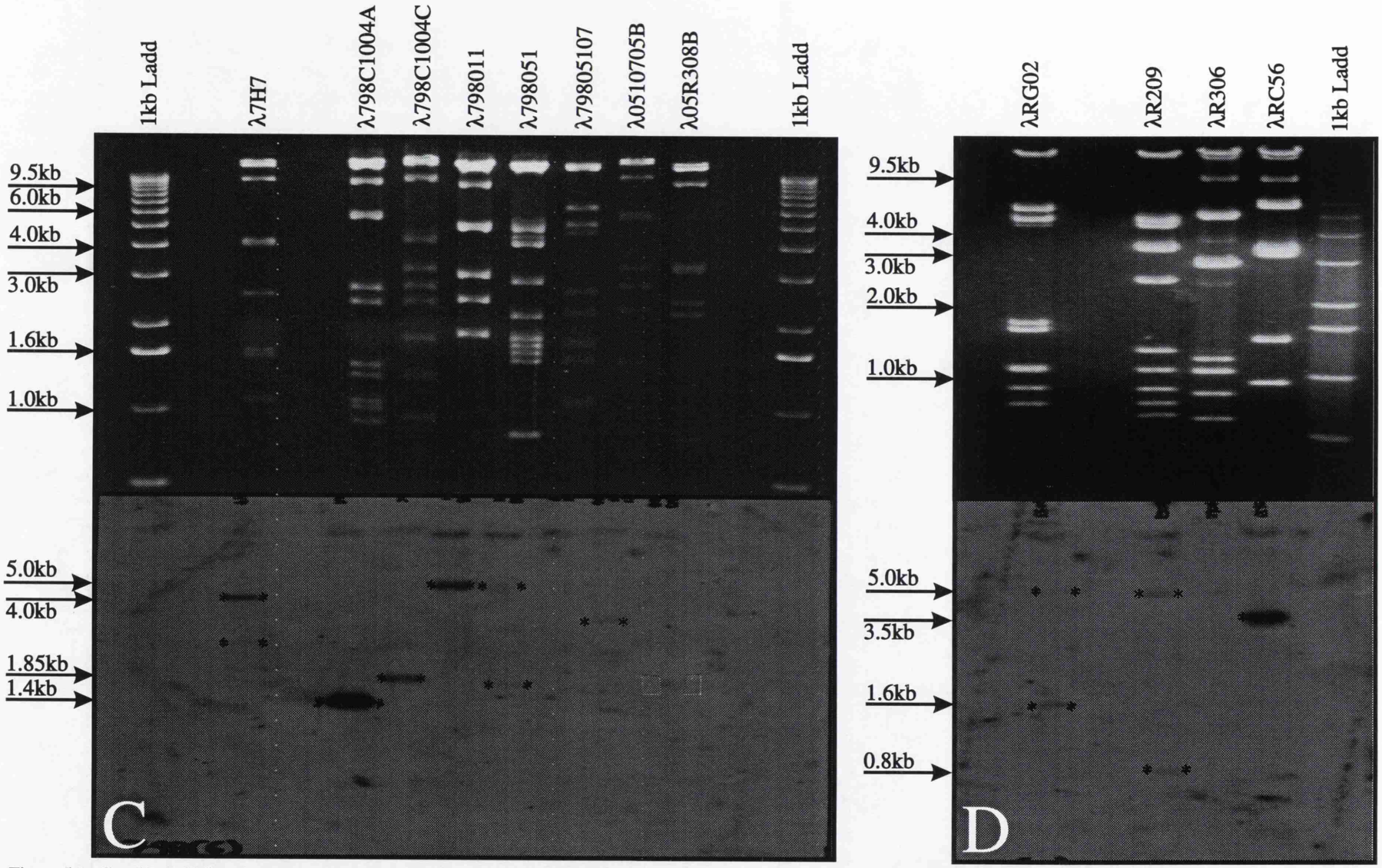


Figure 5.04 C and D-Ethidium bromide stained gels and corresponding autoradiographs of the 798 (C) and *runt* (D) walk genomic clones digested with appropriate enzymes (see text), Southern blotted and then hybridised with radiolabelled first instar larval first strand cDNA. All of the hybridising restriction fragments are marked with asterixs. The results are summarised in Appendicies 3 and 4.

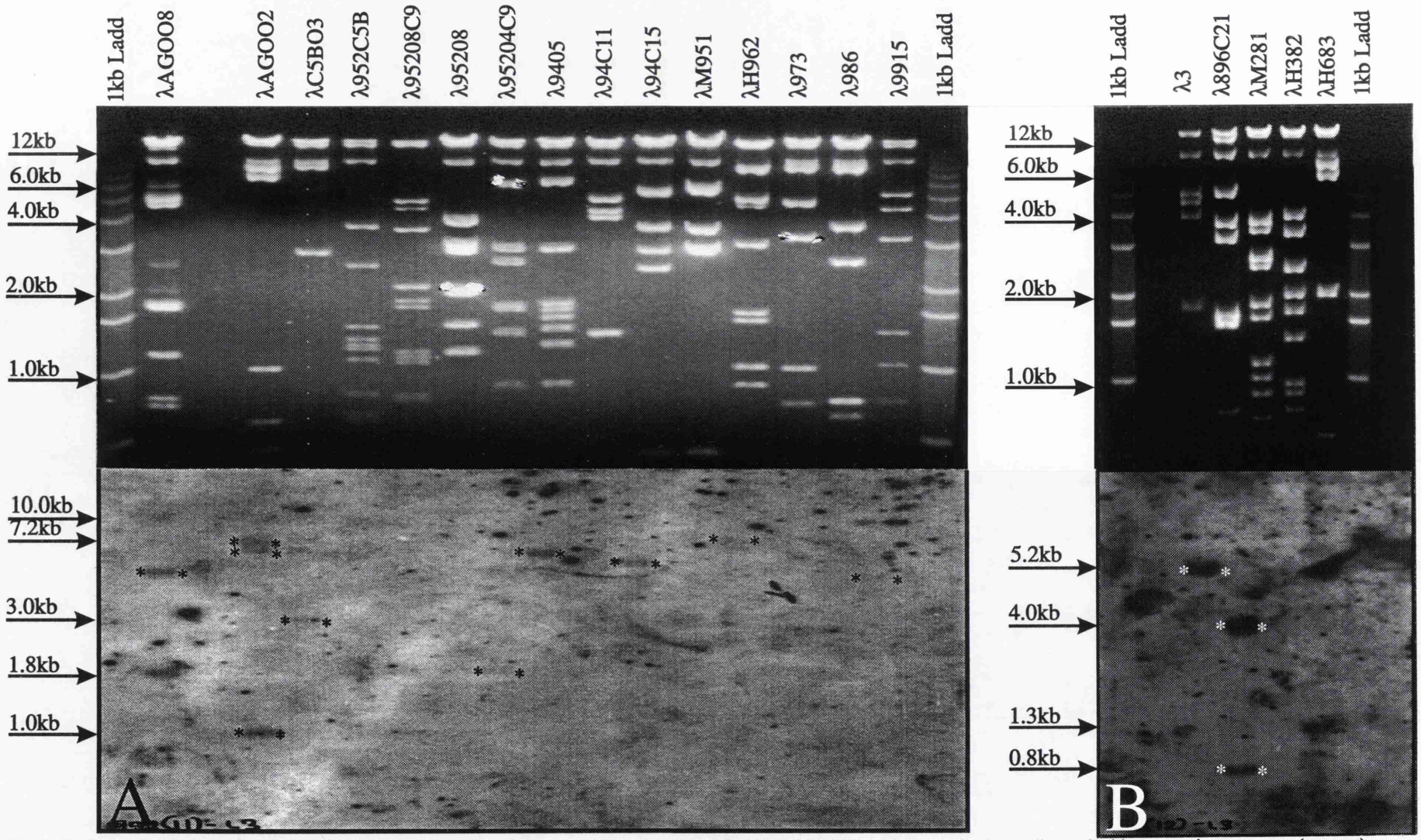


Figure 5.05A and B-Ethidium bromide stained gels and corresponding autoradiographs of the 952 (A) and 896 (B) walk genomic clones digested with appropriate enzymes (see text), Southern blotted and then hybridised with radiolabelled third instar larval first strand cDNA. All hybridising restriction fragments are marked by asterixes. The results are summarised in Appendices 1 and 2.

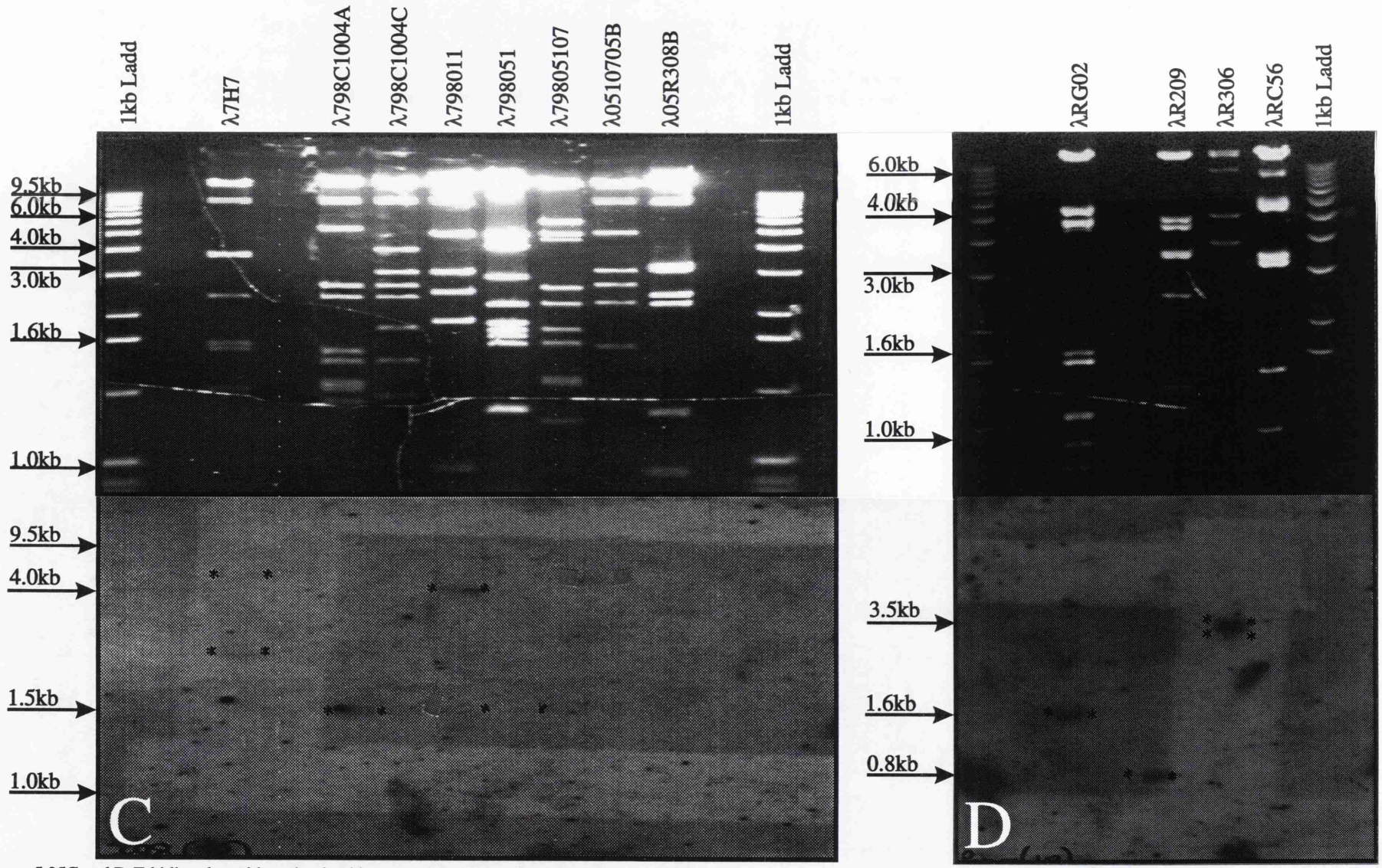


Figure 5.05C and D-Ethidium bromide stained gel images of the 798 and runt walk genomic clones digested with appropriate enzymes (see text), Southern blotted and then hybridised with radiolabelled third instar larval first strand cDNA. All hybridising restriction fragments are labelled with an asterix. The results are summarised in Appendices 3 and 4.

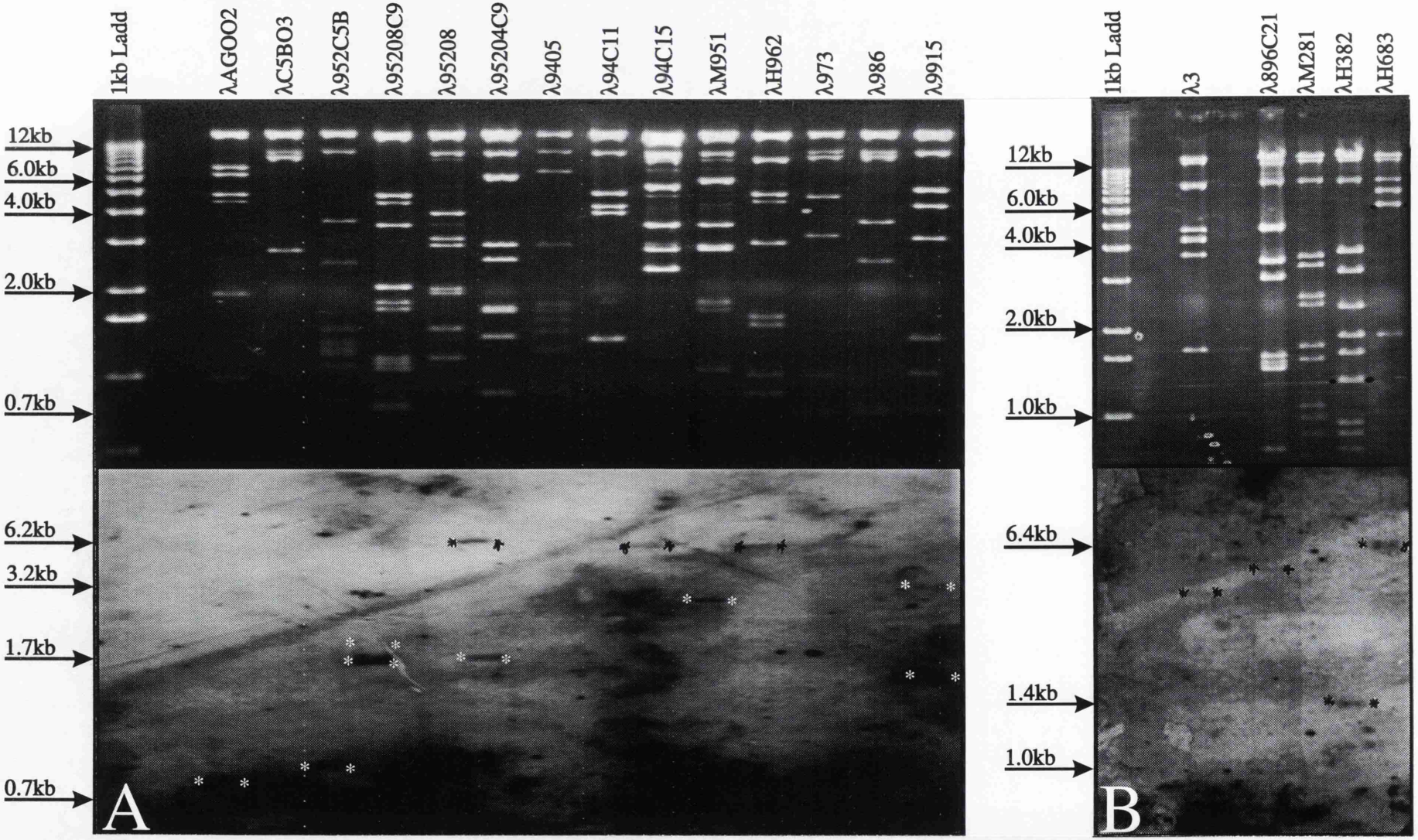


Figure 5.06 A and B-Ethidium bromide stained gels and corresponding autoradiographs of the 952 (A) and 896 (B) genomic walk clones, digested with appropriate enzymes (see text), Southern blotted and hybridised with radiolabelled pupal first strand cDNA. All hybridising restriction fragments are labelled with asterixes. The results are summarised in Appendices 1 and 2.

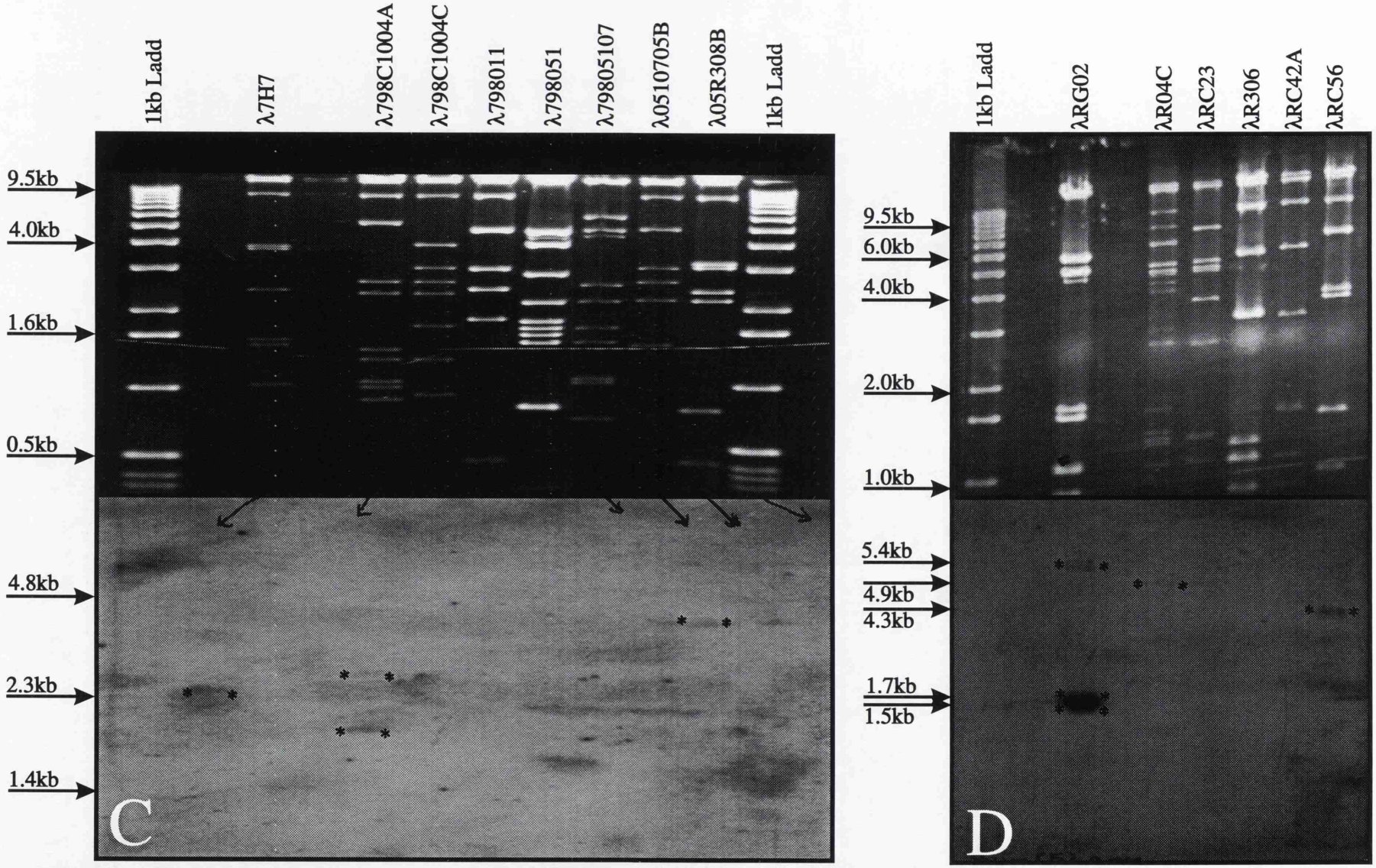


Figure 5.06C and D-Ethidium bromide stained gels and corresponding autoradiographs of the 798 (C) and *runt* (D) genomic walk clones, digested with appropriate enzymes (see text), Southern blotted, and then hybridised with radiolabelled pupal first strand cDNA. All hybridising restriction fragments are labelled with asterixs. The results are summarised in Appendices 3 and 4.

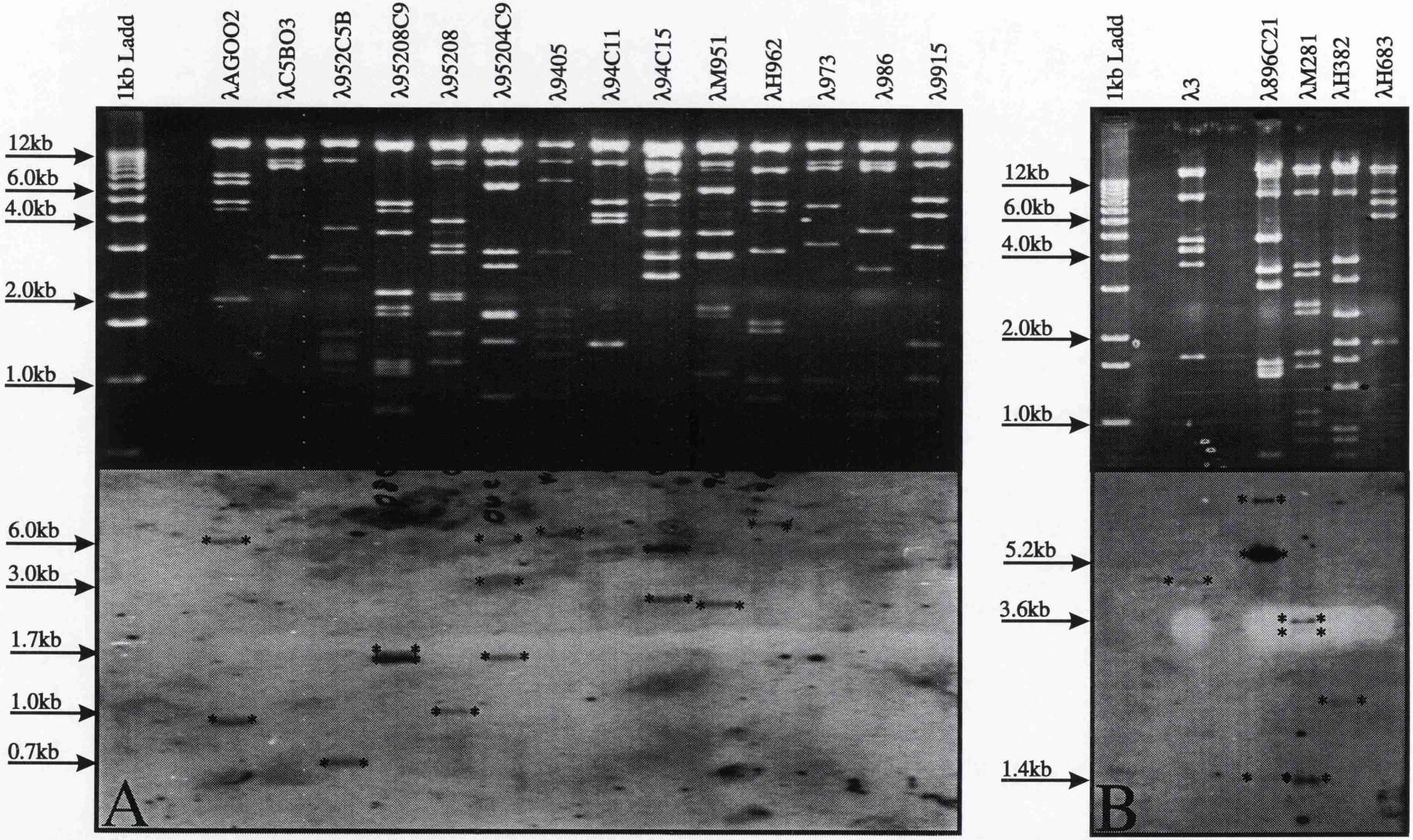


Figure 5.07A and B-Ethidium bromide stained gels and corresponding autoradiographs of the 952 (A) and 896 (B) genomic walk clones, digested with appropriate enzymes (see text), Southern blotted and then hybridised with radiolabelled adult first strand cDNA. All hybridising restriction fragments are labelled with asterixs. The results are summarised in Appendices 1 and 2.

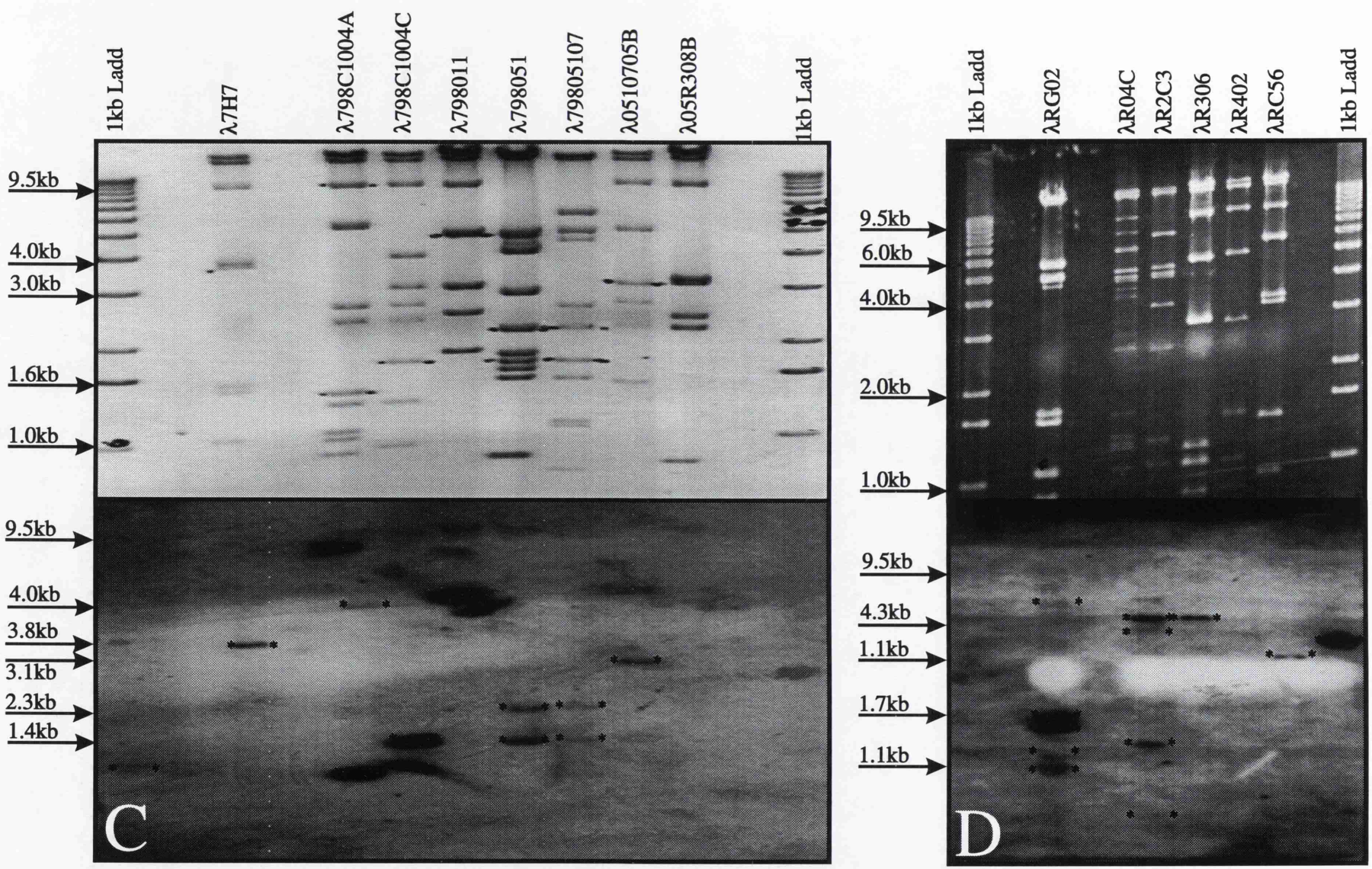


Figure 5.07 C and D-Ethidium bromide stained gels and corresponding autoradiographs of the 798 (C) and *runt* (D) genomic walk clones digested with appropriate enzymes (see text), Southern blotted, and then hybridised with radiolabelled adult first strand cDNA. All hybridising restriction fragments are labelled with asterixs. The results are summarised in Appendicies 3 and 4.

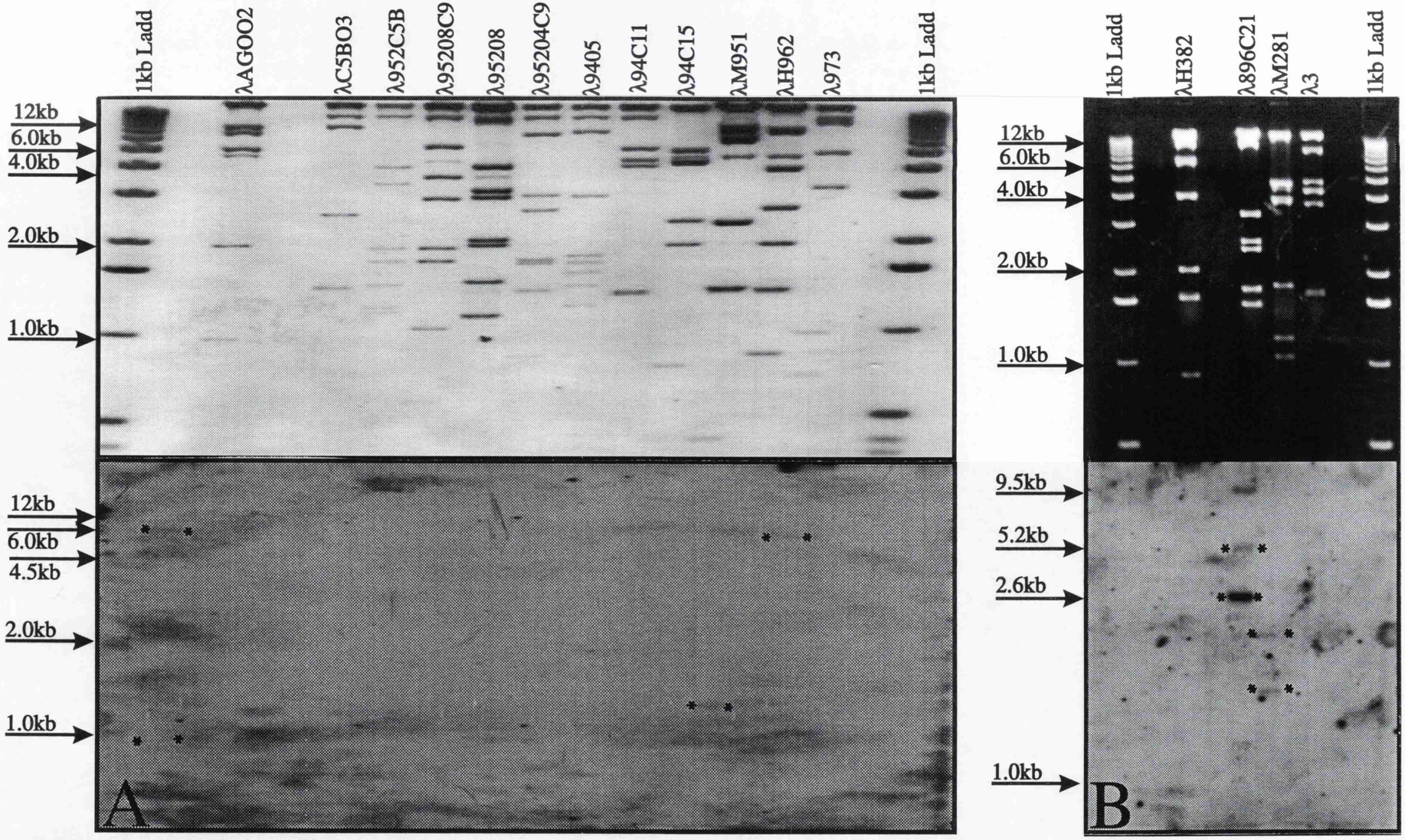


Figure 5.08A and B-Ethidium bromide stained gels and corresponding autoradiographs of the 952 (A) and 896 (B) genomic walk clones digested with appropriate enzymes (see text), Southern blotted, and then hybridised with radiolabelled body first strand cDNA (prepared from adult tissue). All hybridising restriction fragments are labelled with an asterix. The results are summarised in Appendices 1 and 2.

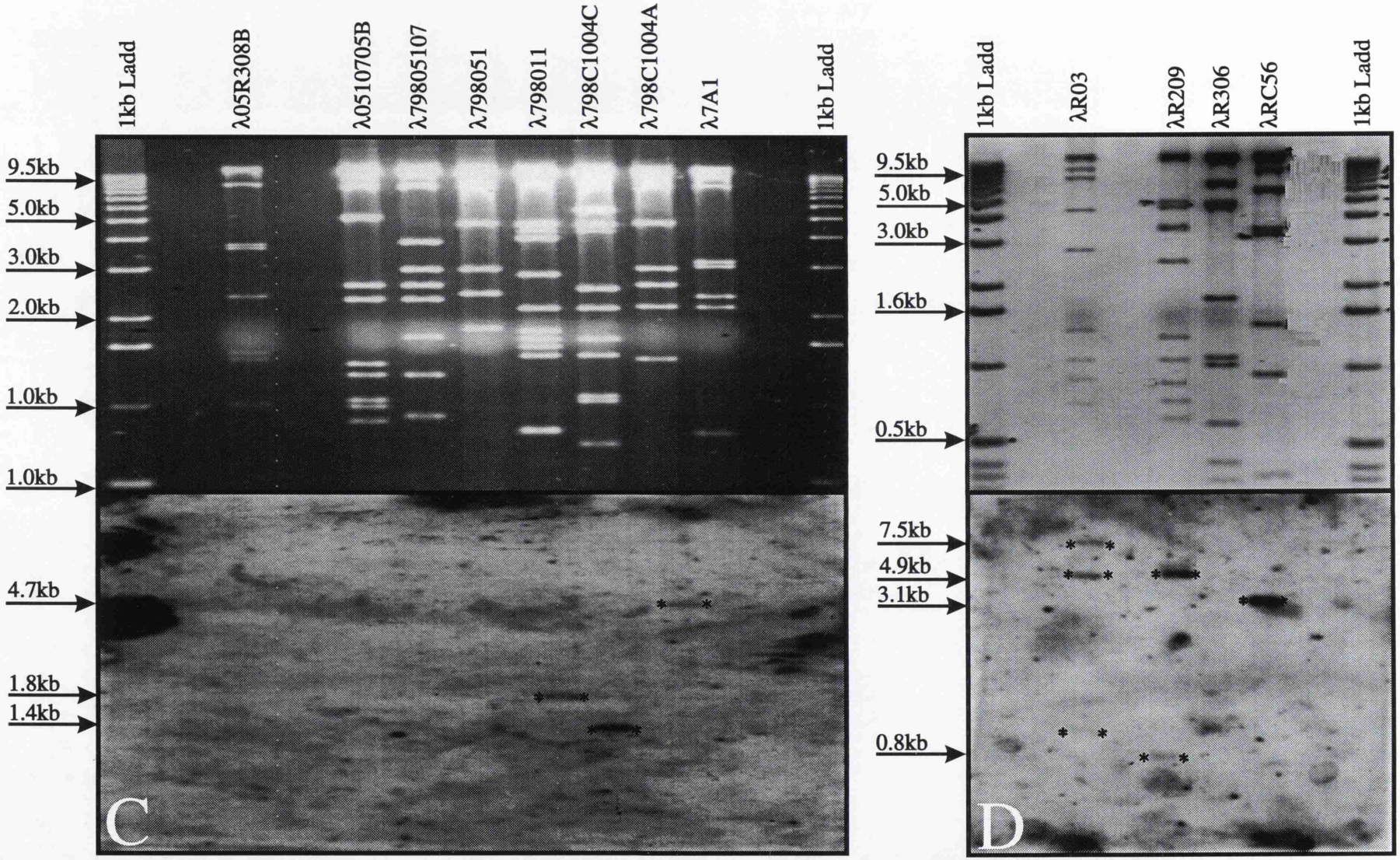


Figure 5.08 C and D-Ethidium bromide stained gels and corresponding autoradiographs of the 798 (C) and *runt* (D) genomic walk clones digested with appropriate enzymes (see text), Southern blotted, and then hybridised with radiolabelled body first strand cDNA (prepared from adult tissue). All hybridising restriction fragments are labelled with asterixs. The results are summarised in Appendicies 3 and 4.

Genomic region	Reverse Northern analysis	Repetitive.	Northern Analysis.	Figure Reference	Appendix 5 label.
1.0kb *Eco* RI - *Eco* RI λAGOO2	E+L1+L3+P+A+B	yes	Individual transcripts not seen.	Fig 5.09	A
6.6kb *Eco* RI - *Hin* dIII λAGOO2	E+L1+L3+A+B	yes	N/D	---	B
7.4kb *Eco* RI - *Eco* RI λAGOO2 (contains 2.45kb λC5B03 fragment).	E+L1+L3	no	2.0kb transcript in adult heads + pupae (appears to be head enriched).	Fig 5.10	C
7.2kb λC5B03	L1	n/d	N/D	---	D
1.7kb *Eco* RI - *Eco* RI λ952C5B	P+A	no	1.4kb adult, pupal and embryonic transcript.	Fig 5.11	E
2.0kb *Hin* dIII - *Hin* dIII λ95208C9	E	no	1.0 + 1.25 + 1.5kb transcripts in different various stages except embryos	Fig 5.12	F
1.7kb *Eco* RI - *Eco* RI λ95208C9	P+A	no			
1.0 *Eco* RI - *Eco* RI λ95208	A	no	N/D	Fig	G
6.0kb *Eco* RI - *Eco* RI λ95204C9	L1+P+A	no	1.6kb head specific transcript.	Fig 5.13.1	H
2.5kb *Eco* RI - *Eco* RI λ95204C9	E+L1+A	no	1.6kb head specific transcript.	Fig 5.13.2	I
1.8kb and 1.7kb *Eco* RI - *Eco* RI λ95204C9. (can't separate these).	E+L1+L3+P+A				J K
6.0kb *Eco* RI - *Eco* RI λ9405	E+L1+L3+A	yes	N/D	Fig 4.02c	M
5.0kb *Eco* RI - *Eco* RI λ94C15	E+L1+L3+P+A	yes	N/D	Fig 4.02d	N
2.5kb *Eco* RI - *Sal* I λM951.	P+A+B	no	No results *i.e.* no signal detected on a Northern although all controls worked.	Fig 5.16	O
8.0kb *Hin* dIII - *Hin* dIII λH962	E+L1+L3+A	no			P
5.3kb *Sal* I - *Sal* I λ9915	L1	no	N/D	---	Q
3.2kb *Sal* I - *Sal* I λ9915	P	no	2.15kb transcript in L3, P, A (head enriched)	Fig 5.14	R
1.4kb *Sal* I - *Sal* I λ9915	P	no			S
4.3kb *Eco* RI - *Eco* RI λ9915	L1	no	1.0kb transcript at all stages	Fig 5.15	T

Table 5.1-Transcribed regions identified across the 19E3 region (952 walk). **Key**; E=Embryonic; L1 =First instar larval; L3 = Third instar larval; P =Pupal; A = Adult and B =Body. N/D = No data *i.e.* the hybridisation was not performed..

contaminated with plasmid sequences. As no hybridisation occured to these sequences it was decided to include those results obtained with this clone in these experiments in the results and ignore the slight contamination.

5.4 Northern Analysis.

For this analysis between 1 - 5μg of staged polyA$^+$ RNA and up to 70μg of staged total RNA were electrophoresed through denaturing formaldehyde gels (Section 2.2.8.2). On every gel where poly A$^+$ RNA was used at least one track of total RNA was also included as a control for possible non-specific hybridisation to contaminating ribosomal RNA. Size markers, either an RNA ladder (0.24kb - 9.5kb) or 1kb DNA ladder (0.075kb - 12.2kb) which had been previously calibrated against the RNA ladder, were also run on each gel. Transcribed regions from across the 19E3 (952 walk) region (Section 5.3) identified in the reverse Northern analysis were used to probe developmental Northern blots. After the experiments were completed each Northern blot was hybridised with the ribosomal protein gene *rp49* (O'Connell and Rosbach 1984; Kongsuwan *et al.*, 1985). This is transcribed at an almost steady level throughout development and encodes a transcript of *ca.* 650bp. It acts as a good loading and transfer control for each blot.

5.4.1 1.0kb *Eco* RI - *Eco* RI Fragment from λAG0O2.

This fragment is known to be repetitive and has been shown (Section 4.2.5, 1.4.5.1) to contain an *opa* repeat (Wharton *et al.*, 1985). Results from the reverse Northern analysis suggested that at least part of this region is transcribed in all developmental stages of the fly. It was known that this *opa* repeat region (approximately 100bp) would have homology to a large number of transcripts in *Drosophila* RNA and thus when used to probe a Northern blot a 'background smear would be visible as the *opa* repeat region hybridised to the many different, variously sized, transcripts. It was hoped, that any signal from the 1.0kb region would be observable above the general background smearing however this was not the case and no significant signal was detected (Figure 5.09).

5.4.2 7.4kb *Eco* RI - *Eco* RI Region from λAG0O2.

This is a single copy fragment. Results from the reverse Northern analysis suggested that transcripts from at least part of this region are present in first instar larval, third

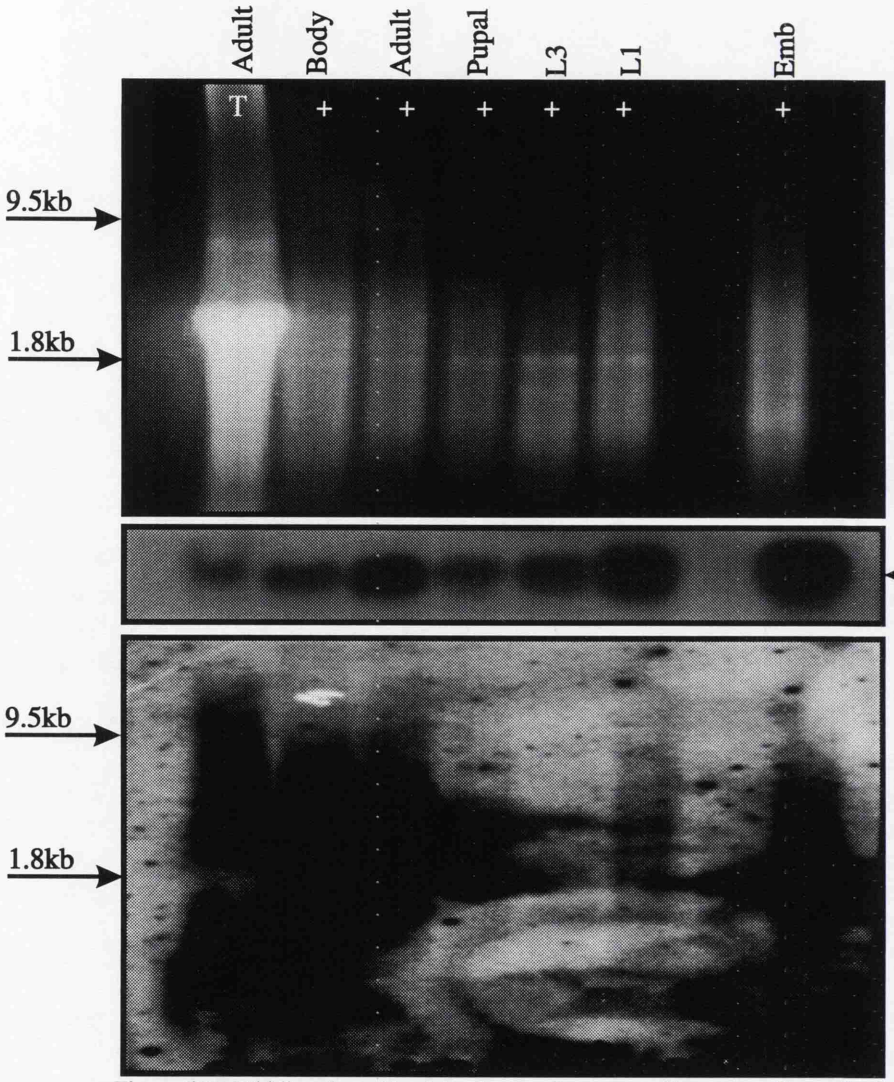


Figure 5.09-Ethidium bromide stained gel and corresponding autoradiograph of a developmental Northern blot, composed of various *Drosophila* staged RNAs probed with a 1.0kb *Eco* RI *opa* repeat containing fragment from λAG002 and a *rp49* loading control. No bands are visible above the background smear caused by the *opa* repeat. See text (Section 5.4.1) for more details.

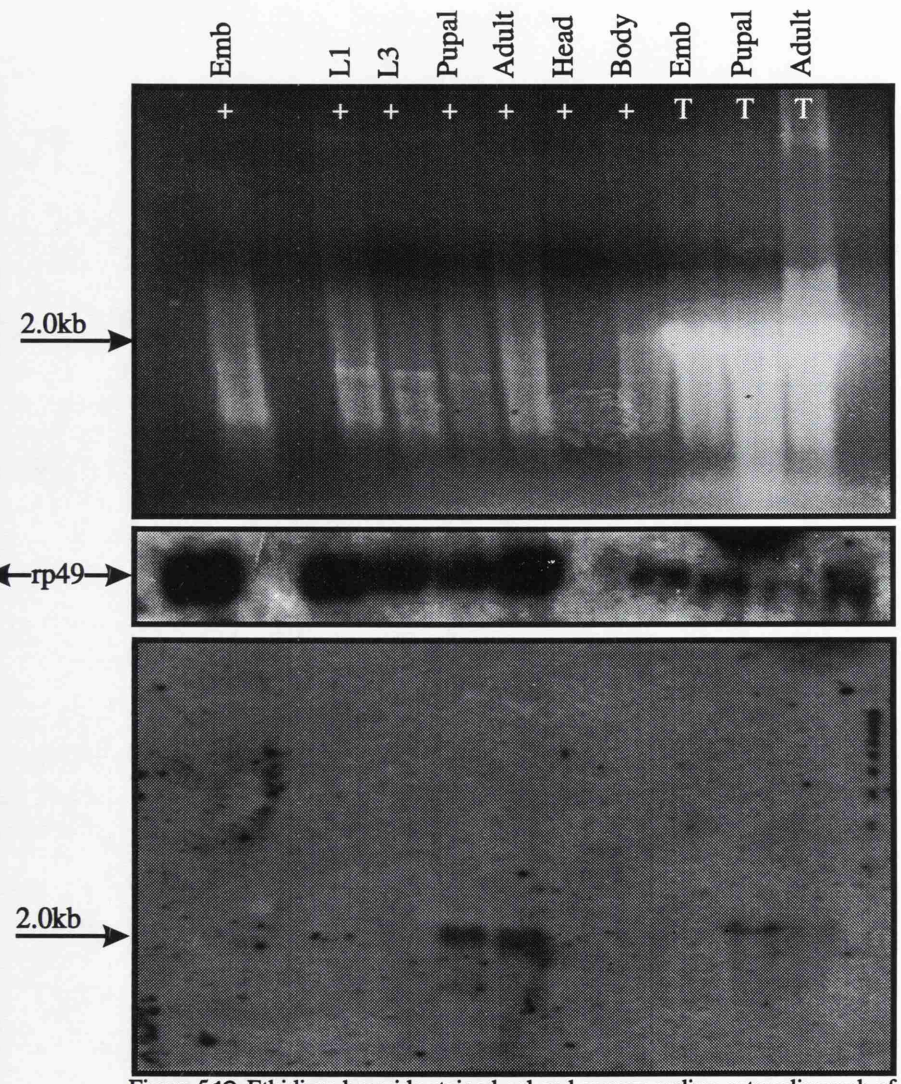


Figure 5.10-Ethidium bromide stained gel and corresponding autoradiograph of a developmental Northern blot, composed of various *Drosophila* staged RNAs probed with a 7.2kb *Eco* RI fragment from λAG002, and a *rp49* loading control. A single transcript of *ca.*2.0kb is visible in some of the tracks. See text (Section 5.4.2) for more details.

instar larval and embryonic RNAs. The fragment was hybridised to a developmental Northern blot containing embryonic, L1, L3, pupal, adult, adult head and adult body poly A^+ RNA together with embryonic, pupal and adult total RNA. The result (Figure 5.10) showed that at least part of the region is homologous to a 2.0kb transcript expressed in pupae and in adults. No expression was visible in the L1 or embryonic RNAs. As the head polyA^+ RNA is very under loaded (see *rp49* loading control) relative to the adult and body RNAs it is possible to say that this fragment is expressed predominantly in the head in the adult.

5.4.3 1.7kb *Eco* RI - *Eco* RI Region from λ952-08C9.

This fragment was shown to be single copy (Section 4.2.1). Results from the reverse Northern analysis (Section 5.3) showed that at transcripts from at least part of this region are found in pupal and adult RNAs. When hybridised to a Northern blot containing embryonic, L1, L3, pupal, adult, adult head and adult body poly A^+RNAs as well as embryonic, and adult total RNAs, the results (Figure 5.11) show that this region hybridises to a 1.4kb transcript expressed in pupae and adults, being expressed equally in both adult head and body tissues.

5.4.4 4.8kb *Eco* RI - *Eco* RI Region from λ952-08C9.

This single copy fragment (Section 4.2.1) is composed of three fragments. A distal 1.0kb *Eco* RI - *Hin* dIII fragment that appears not to contain any transcribed sequences; a 1.9kb *Hin* dIII - *Hin* dIII fragment that the reverse Northern analysis suggests contains sequences expressed in embryos and a 1.7kb fragment (the most proximal) that appears to contain sequences expressed in adults and pupae (Section 5.3). The whole 4.8kb region was hybridised to the Northern blot used for the analysis of 7.2kb *Eco* RI - *Eco* RI region from λAG0O2, (Section 5.4.2) and a complex pattern of transcripts was observed (Figure 5.12). A faint transcript (or transcripts) of *ca.* 1.0kb are visible in the L1 RNA. This signal is also visible in the L3 RNA with the addition of another transcript at *ca.* 1.4kb. In the pupal RNA a strong signal corresponding to the 1.0kb transcript(s) is visible, which is also the case in the adult RNA. There is however an additional transcript of *ca.* 1.25kb in the adult RNA that is also visible. Due to the under loading of the head RNA no signal is visible however transcripts are visible at *ca.* 1.0kb and 1.25kb in the body RNA. Weaker signals are visible in the pupal and adult total RNA tracks which correspond to the above

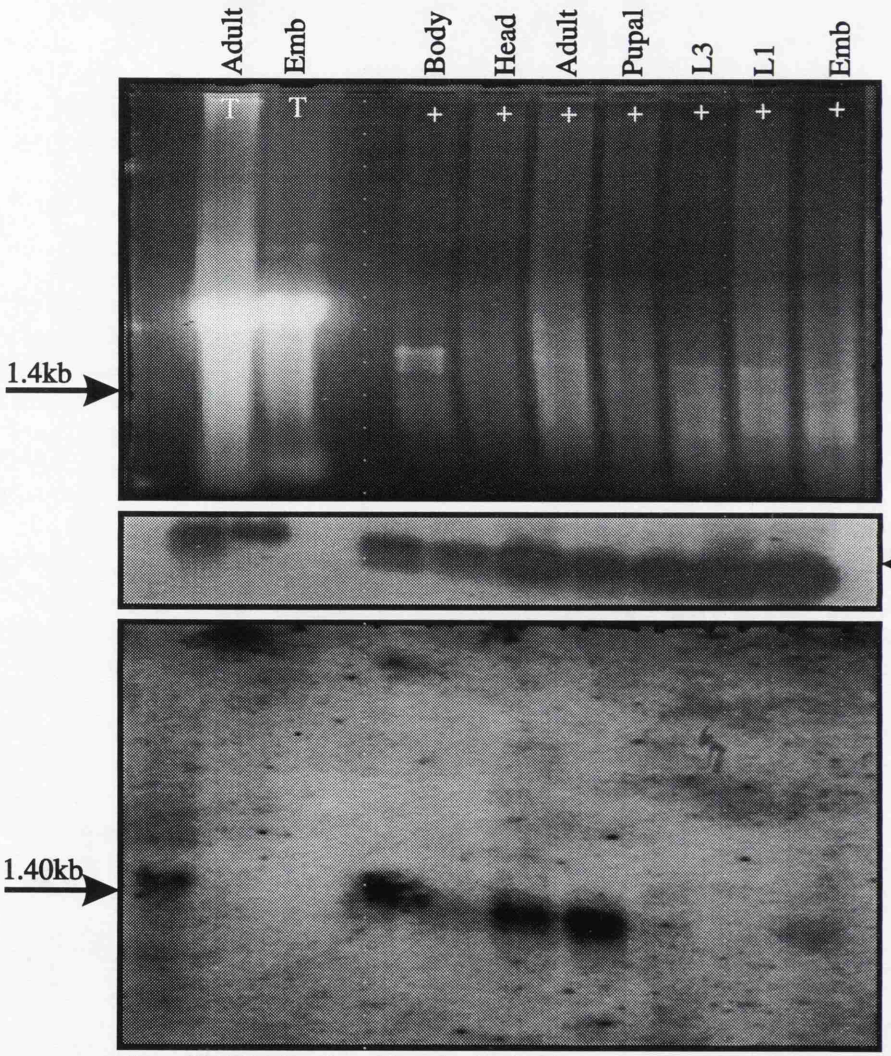


Figure 5.11-Ethidium bromide stained gel and corresponding autoradiograph of a developmental Northern blot, composed of various *Drosophila* staged RNAs probed with a 1.7kb *Eco* RI fragment from λ95208C9 and an *rp49* loading control. A single transcript of *ca.*1.4kb is visible in the pupal and adult tracks. See text (Section 5.4.3) for more details.

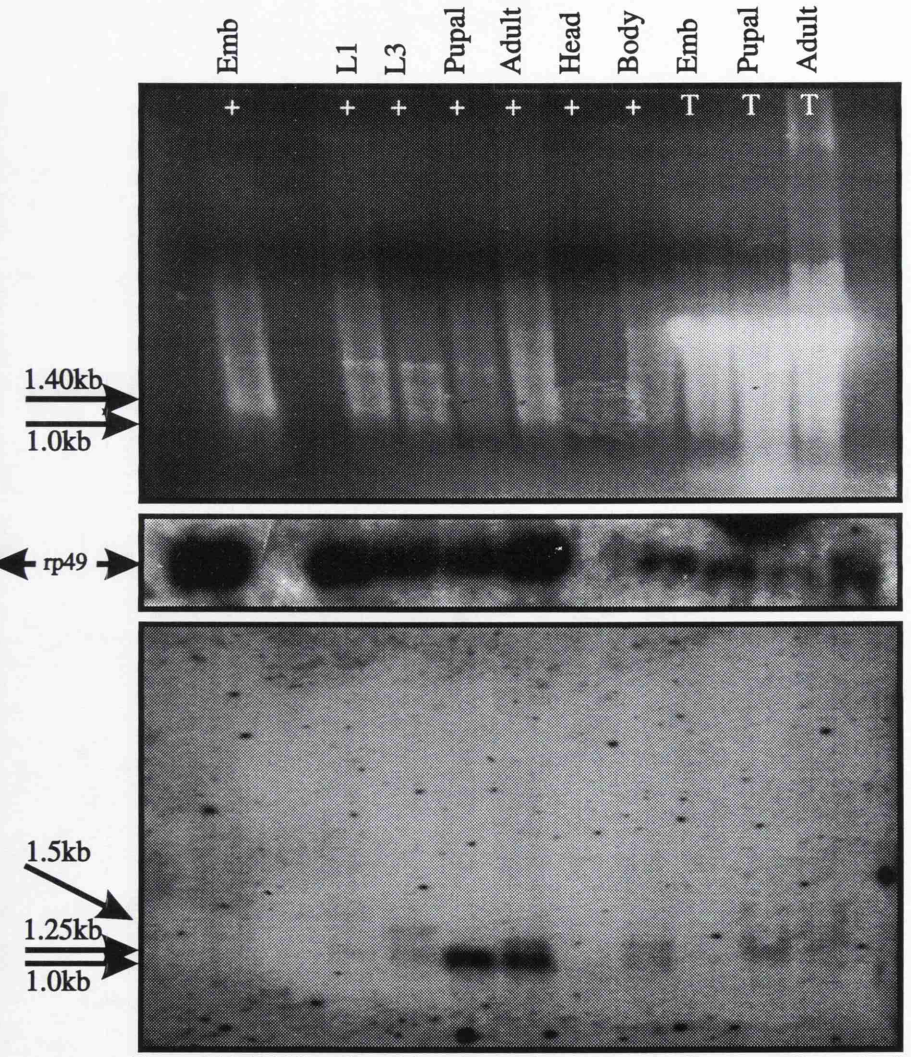


Figure 5.12-Ethidium bromide stained gel and corresponding autoradiograph of a developmental Northern blot, composed of various *Drosophila* staged RNAs probed with a 4.8kb *Eco* RI fragment from λ95208C9 and an *rp49* loading control. Three transcripts of 1.0kb - 1.5kb are visible in some of the tracks. See text (Section 5.4.4) for more details.

transcripts, with the addition of a signal at *ca.* 1.8kb which may or may not correspond to a real transcript as it co-migrates with the 18S rRNA. Therefore it may be spurious hybridisation to rRNA.

5.4.5 12.0kb Region from λ952-04C9.

This region is composed of four *Eco* RI fragments of 6.0kb, 2.5kb, 1.8kb and 1.7kb, all of which were shown to be single copy (Section 4.2.1). The 1.7kb and 1.8kb fragments co-migrate as a doublet and so will be called the 1.7/1.8kb fragment. The reverse Northern analysis suggested that the 6.0kb region contain sequences expressed in adult, L1 and pupal tissues. The 1.7/1.8kb fragment has sequences expressed in embryonic, L1, L3 and pupal and adult tissues. The analysis suggested that the 2.5kb fragment was expressed in embryonic, L1 and adult tissues. For the purpose of Northern analysis the region was divided into two. The 6.0kb fragment hybridised to a single 1.6kb transcript which was expressed only in adult heads (Figure 5.13.1). The 2.5kb and the more distal 1.7/1.8kb fragments were simultaneously hybridised to the same Northern blot and a single head specific transcript of 1.6kb was also detected (Figure 5.13.2). This transcript may lie in either or all of these fragments. No other signals were visible in any of the other staged RNAs present on the Northerns.

5.4.6 4.6kb Region from λ9915.

This region is composed of two *Sal* I fragments of 3.2kb and 1.4kb which were shown to be single copy (Section 4.2.1). The reverse Northern analysis suggested that one or both of these fragments have homology to transcripts expressed in pupae. Both fragments were radiolabelled and hybridised to a Northern blot composed of poly A^+ and total RNA (stages used are as indicated in figure). The regions hybridised to a 2.1kb transcript (Figure 5.14) which was expressed weakly in L3 tissue and strongly in pupal and adult tissues.

5.4.7 4.3kb *Eco* RI - *Eco* RI Fragment from λ9915.

This region was shown to be single copy (Section 4.2.1) and the reverse Northern analysis suggested it contained region(s) transcribed in L1 tissues. When hybridised to the developmental Northern blot used for the analysis of the 7.2kb *Eco* RI - *Eco* RI region from λAG0O2, (Section5.4.2), a 1.0kb transcript was observed (Figure 5.15) which appeared to be expressed approximately equally at all developmental stages.

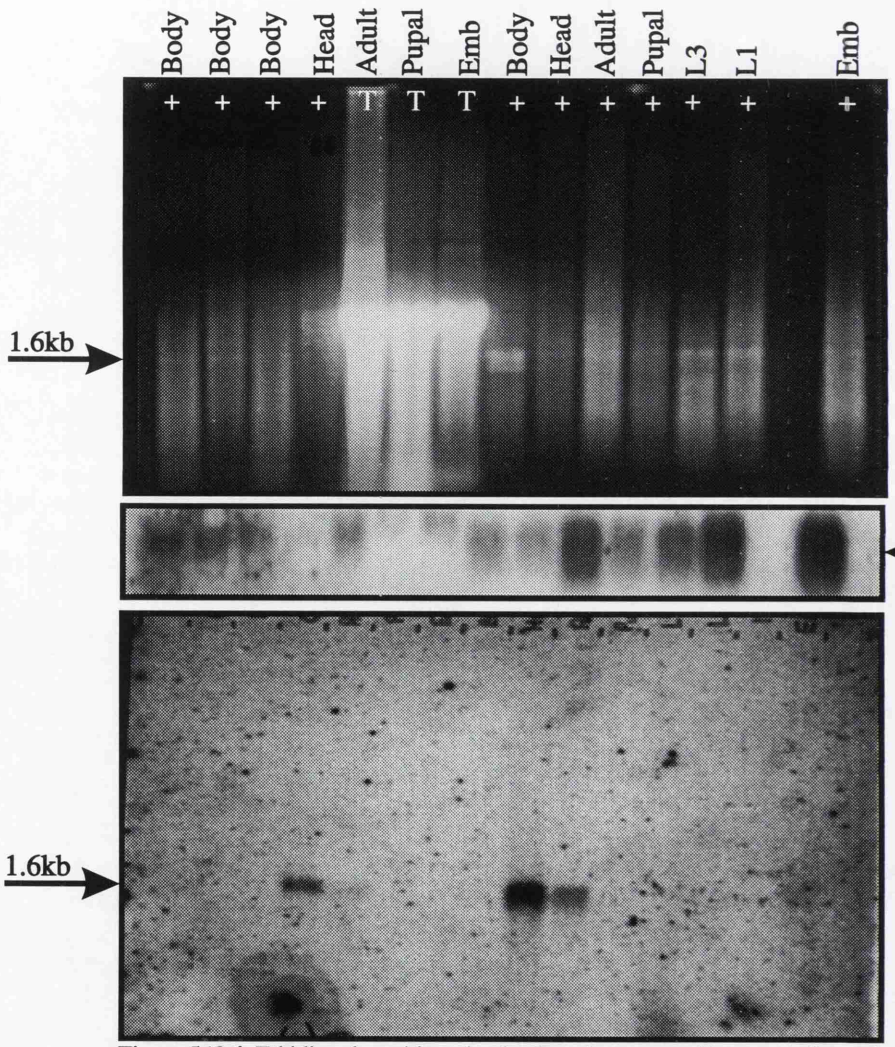


Figure 5.13.1 -Ethidium bromide stained gels and corresponding autoradiograph of a developmental Northern blot, composed of various *Drosophila* staged RNAs probed with a 6.0kb *Eco* RI fragment from λ95204C9, (and *rp49* loading control). A single head specific transcript of *ca.*1.6kb is visible. See text (Section 5.4.2) for more details.

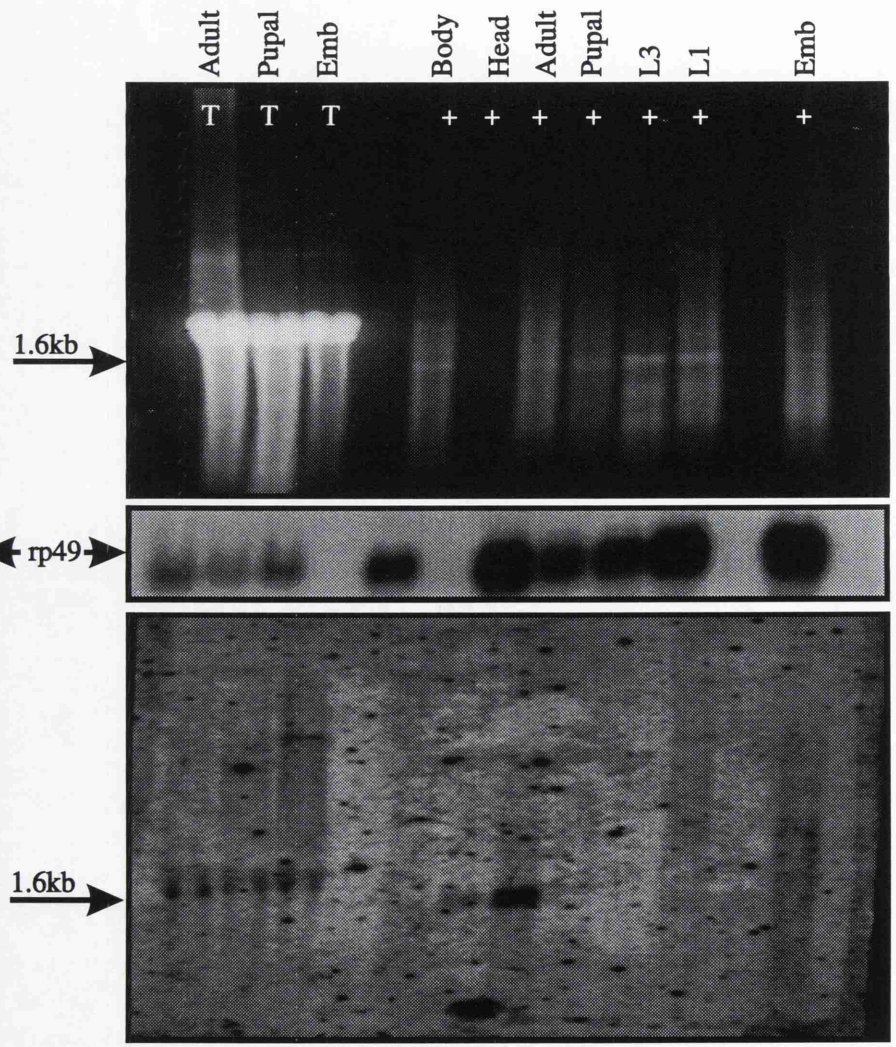


Figure 5.13.2 -Ethidium bromide stained gel and corresponding autoradiographs of a developmental Northern blot, composed of various *Drosophila* staged RNAs probed with a 1.7kb and 2.5kb *Eco* RI fragment from λ95204C9, (and *rp49* loading control). A single head specific transcript of *ca.*1.6kb is visible. See text (Section 5.4.5) for more details.

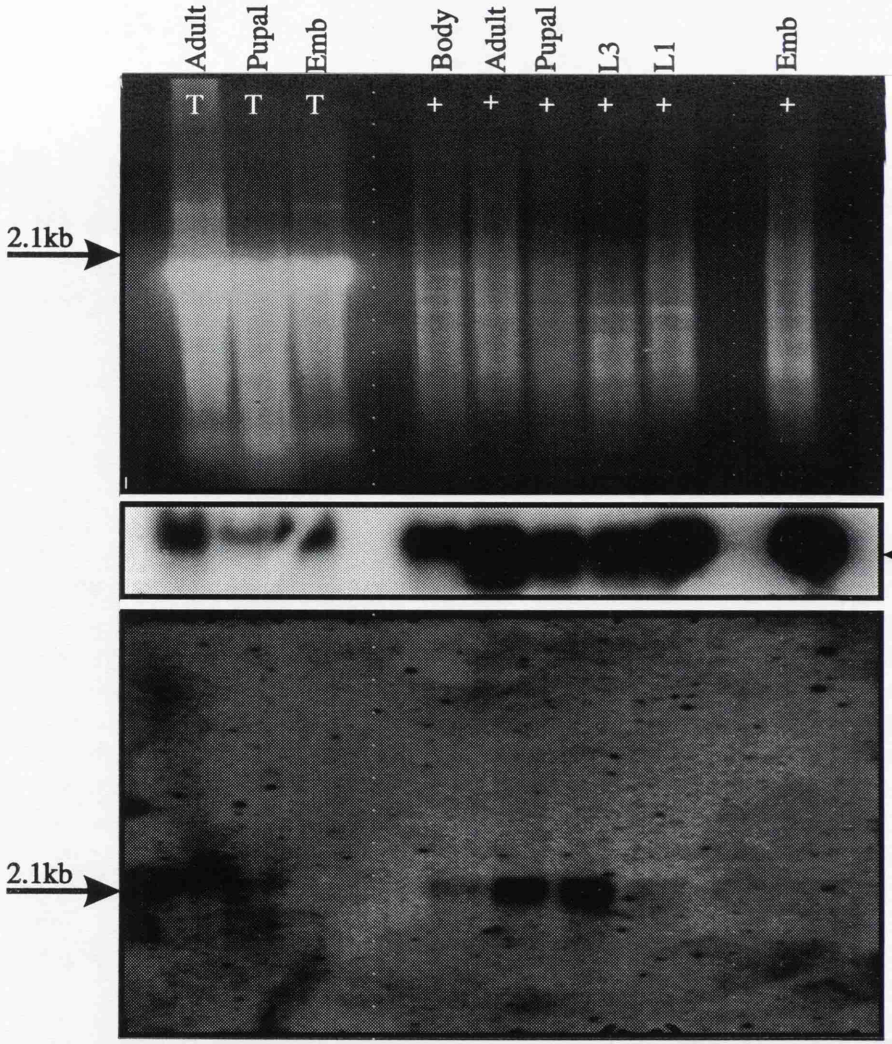


Figure 5.14 -Ethidium bromide stained gels and *rp49* loading control of a developmental Northern blot, composed of various *Drosophila* staged RNAs. A transcript of *ca.*2.1kb is visible in the L3, pupal and adult tracks. By comparing the adult and body tracks it is possible to say that the transcript is head enriched. See text (Section5.4.6) for more details. The probes were a 3.2kb and 1.4kb *Sal* I fragment from λ9915.

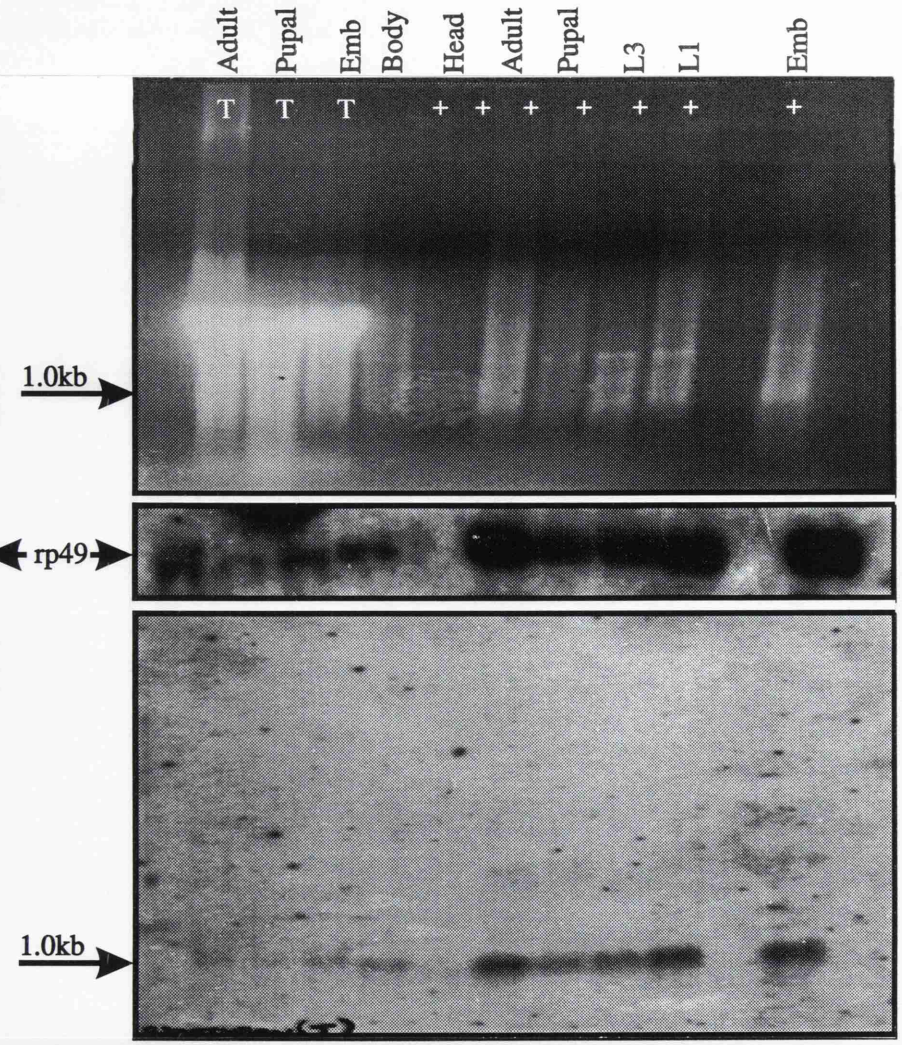


Figure 5.15-Ethidium bromide stained gels and *rp49* loading control of a developmental Northern blot, composed of various *Drosophila* staged RNAs. A single transcript of *ca.*1.0kb is visible in all stages of the lifecycle. See text (Section 5.4.7) for more details. The probe is a 4.3kb *Eco* RI fragment from λ9915.

5.4.8 10.5kb Region from λH962.

This region was shown to be single copy (Section 4.2.1). It is composed of a 2.5kb *Eco* RI - *Sal* I fragment from λM951 that the reverse Northerns suggests contains sequences expressed in pupae, adults and adult bodies. It also contains a *Hin* dIII fragment of 7.8kb which is expressed in embryos, first instar larval, third instar larval and adult tissues. Several cDNA clones have been isolated from this region (J. Davies, personal communication). No signal could be detected upon hybridisation of the genomic region to Northern blots. This was repeated three times and the same negative result, observed consistently. Hybridisation of m23, one of the cDNA clones isolated from this region, also produced no results. All positive controls were successful (Figure 5.16).

5.5 Hybridisation of *shak-B* cDNAs.

As discussed previously (Section 1.2.8.1) four transcripts have been isolated from a 17kb region delimited by the proximal endpoint of *Df(1)16-3-35* and the distal endpoint of *Df(1)LB6*. This region is known to contain a function, either regulatory or structural, that is necessary for a vital function in the fly (Section 1.2.5). The first cDNA clone to be isolated from this region was obtained from a screen of 100 000, 12 - 24 hour embryonic cDNA clones (J Davies personal communication; Crompton *et al.*, 1992) and was named KE2. This 1.8kb clone was radiolabelled to high specific activity ($1.5x10^9 cpmg^{-1}$) and hybridised to a developmental Northern blot composed of $polyA^+$ RNAs and total RNAs from different developmental stages (Figure 5.17C for details). The hybridisation itself was successful as shown by a positive control, a Southern blot of *Drosophila* genomic DNA which possessed all the bands expected (Figure 5.17A and B). To ensure the Northern filter used in the hybridisation was satisfactory, after three weeks of autoradiography to visualise any KE2 transcripts it was hybridised with the Actin-5C plasmid described earlier (Section 5.2.3). The transcripts detected using this probe (Figure 5.17C) were in good agreement with those expected (Fyrberg *et al.*, 1981; 1983). As the controls were successful this suggested that other factors were responsible for the lack of signal with KE2 and other associated transcripts (see Discussion 5.7.7).

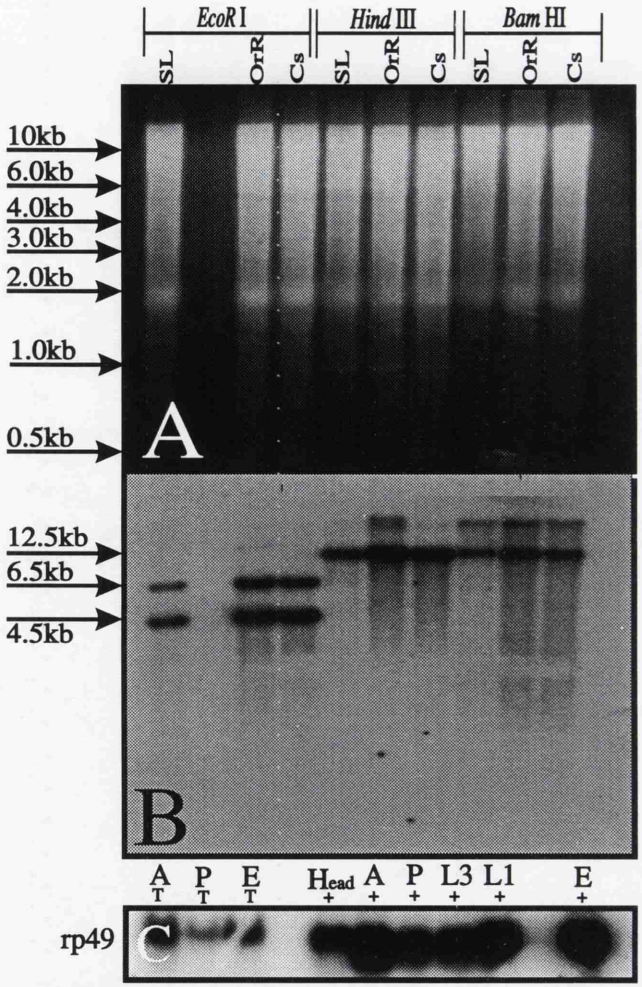


Figure 5.16-The results obtained when the m23 cDNA and region were simultaneously hybridised to developmental Northern and genomic blots. No signal was observed on the Northern after three weeks exposure. The positive controls for the hybridisation (B) and also for Northern (C) worked well. No reason can be given to explain this lack of signal.

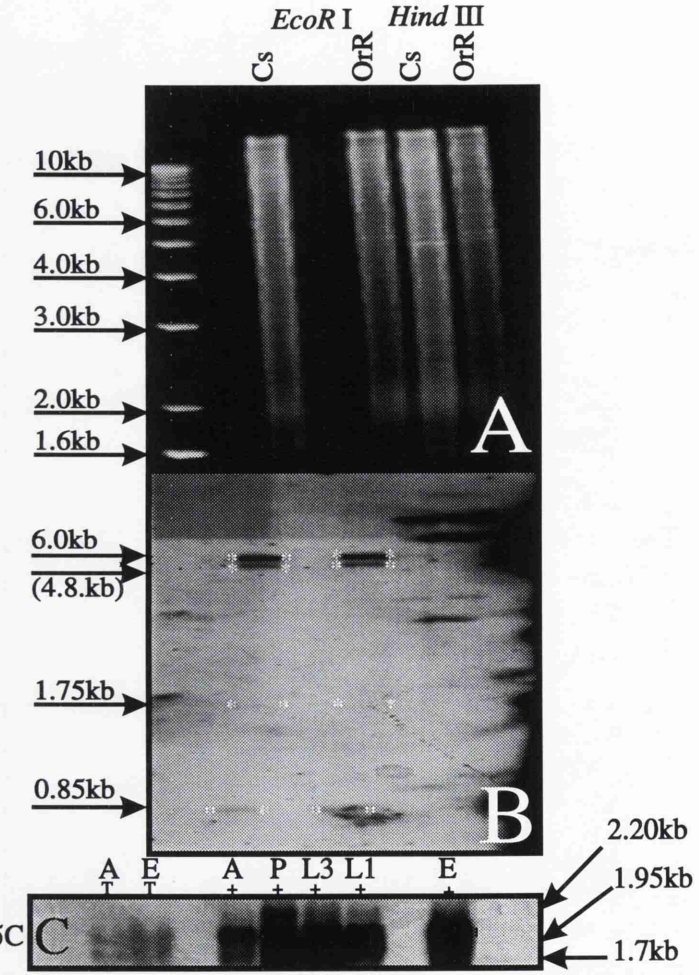


Figure 5.17-The results obtained when the cDNA KE2 was simultaneously hybridised to a genomic and developmental Northern blot. No signal was observed on the Northern after three weeks exposure. The positive controls for the hybridisation (B) and also for Northern (C) worked well. See discussion (Section 5.7.6) for further details. The 1.75 and 0.85kb bands are faint as they contain less overlap with the cDNA.

5.6 Further Analysis of the 12.0kb Region from λ952-04C9.

This region is situated distal to the *Df(1)LB6* distal breakpoint (Chapter 3). Northern analysis detected only a single 1.6kb head specific transcript originating from this area (Section 5.4.5). However, several other transcripts have now been isolated from this region, which have been identified as belonging to the *shak-B* locus (J. Davies, personal communication). Because of the rarity of these transcripts in cDNA libraries most were isolated using PCR based methods. The transcripts all have incomplete 3' ends as they were all internally primed. It is therefore possible that the remaining 3' ends of these transcripts lie in this 10.3kb region. It was decided to hybridise the whole 10.3kb region, plus another fragment of 1.7kb from the same region, to a Northern blot composed of adult polyA$^+$ RNA isolated from the *shak-B* alleles. This was done to determine if the 1.6kb head specific transcript was altered in any of the alleles so allowing it to be assigned to the *shak-B* locus. Although alterations in the 1.6kb transcript were not detected, the results were interesting as one of the alleles, had several extra bands (Figure 5.18) not visible in the wild type controls, or in the other alleles. Time constraints prevented more *shak-B*$^{17\text{-}360}$ flies being amplified and therefore the experiment could not be repeated. After allowing the signal to decay however, the filter was reprobed with radiolabelled 1kb ladder. This should detect any plasmid or phage sequences responsible for a spurious result. No hybridisation was detected. The filter was reprobed with the *rp49* clone described earlier and just the single 0.65kb transcript expected was detected. Thus the extra bands in the *shak-B*$^{17\text{-}3\text{-}60}$ allele RNA, were not caused by hybridisation of contaminating phage or plasmid sequences, and the Northern blot itself was shown to be to satisfactory. This strongly suggests that the extra transcripts are the result of an alteration caused by the *shak-B*$^{17\text{-}3\text{-}60}$ allele.

5.7 Discussion.

5.7.1 Identity of the Transcribed Regions.

The reverse Northern analysis described and carried out in this study has been used successfully by several other workers to identify transcribed regions (*e.g.* Hall *et al.*, 1983; Scott *et al.*, 1983; Knipple *et al.*, 1991). Thirty two different transcribed regions were identified of which fifteen were shown to contain non-repetitive sequences. Nine of these regions (from the 19E3 region) were characterised further by Northern analysis and / or partial sequence analysis (see Section 6.3). Although the primary aim

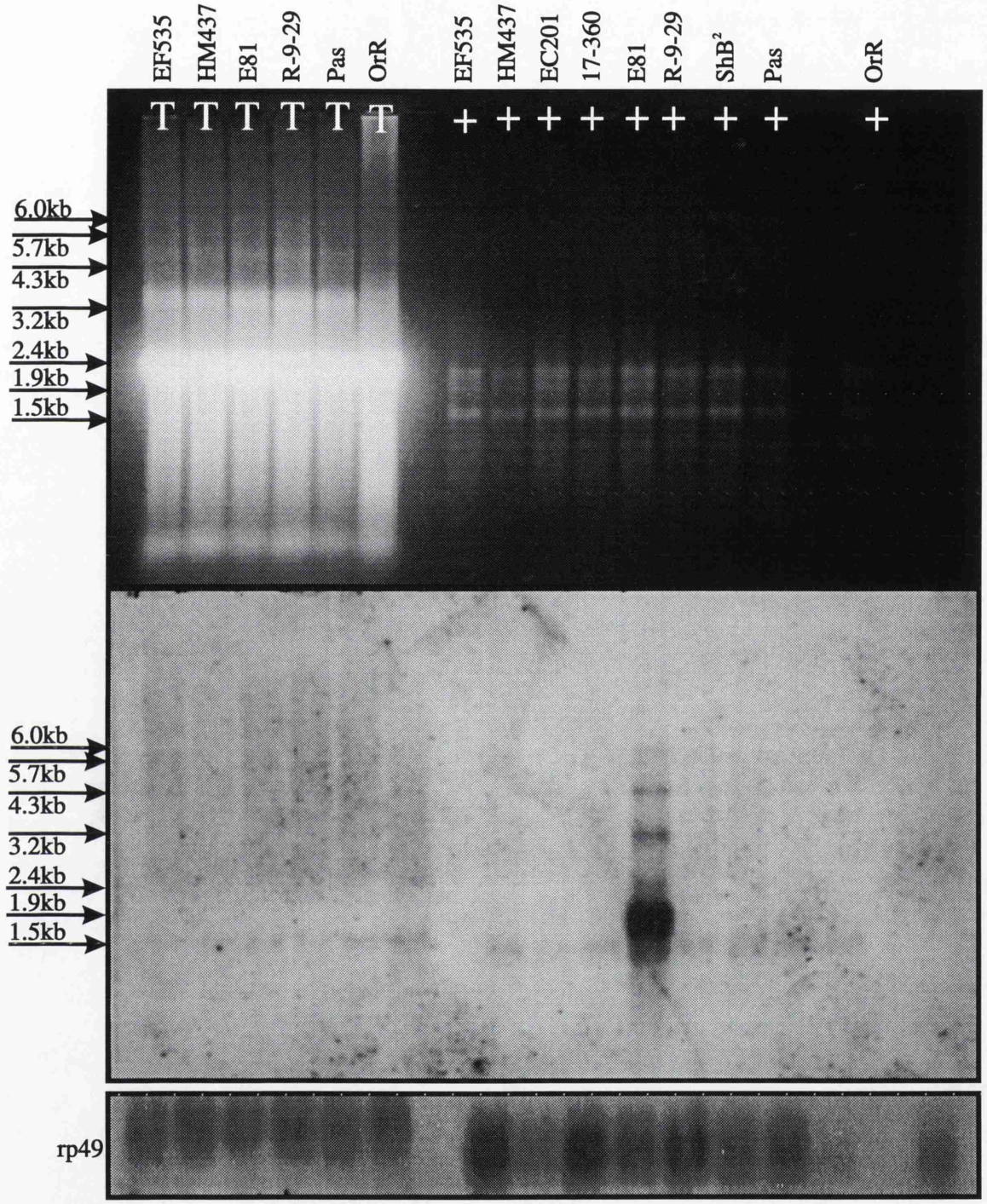


Figure 5.18-Ethidium bromide stained gel of a Northern blot (T= Total RNA; += Poly A^+ enriched RNA) composed of RNAs isolated from adult, *shak-B* mutant allele bearing flies, and the corresponding autoradiograph after hybridisation of the blot with 6.0, 1.7 and 2.5kb *Eco* RI fragments from λ95204C9. Extra bands are visible in the *shak-B*$^{17\text{-}360}$ RNA. See text (Section 5.6) for more details.

of this analysis was to identify *shak-B* transcripts, a secondary benefit was the identification of previously unidentified, and thus uncharacterised, transcribed regions that may be unknown genes. The analysis described established a developmental profile of the transcribed regions. However the analysis did not allow the grouping of transcribed regions into areas of related origin *i.e.* from the same locus but from a different location as differential splicing occurs. Nevertheless, as the regions detected will almost certainly represent the 3′ ends of mRNA products (as the oligo dT primer will nearly always prime off the poly A^+ tail) it is likely that most of the regions will correspond to novel 3′ ends.

It should be noted that other, non-transcribed regions may also be detected by reverse Northern analysis. Genomic features such as AT rich stretches, or short repetitive regions such as the *opa*, *ala* or *pen* repeats found in non-coding DNA that also have homologous sequences within mRNA, will hybridise to oligo dT primed cDNA. AT rich stretches in genomic DNA will hybridise to poly A tails whereas short repetitive sequences will hybridise with homologous sequences present within the mRNA. Other sequences, for example the remnants of transcribed transposable elements will also be detected by the reverse Northern analysis. All of these features will be indistinguishable from 'legitimate' transcribed regions.

5.7.2 Discrepancies Between the Northern and Reverse Northern Data.

As can be seen from Table 5.1, discrepancies exist between the results from the reverse Northern and Northern analysis. These probably represent probe differences. It was expected that the probes would not be identical for a variety of reasons. Firstly although the poly A^+ selection was judged to have removed over 95% of the rRNA, from the mRNA used as a template for the cDNA probes, (estimated from direct visualisation of the RNA on ethidium bromide stained gels), differences in the amount of contaminating rRNA between different RNA preparations will still occur. This will give rise to an over-estimate of the amount of radiolabel incorporated from priming of poly A^+ RNA as some of it will be incorporated into cDNA derived from misprimed rRNA transcripts. A second difference between the RNAs used for the synthesis of cDNA is the amount of degraded material that the mRNA samples contain *e.g.* pupal mRNA tends to be more degraded than RNA isolated from other stages of the fly's lifecycle due to the amounts of liposomes, RNases *etc.* that are present in the rapidly

differentiating tissue. This is demonstrated in Figure 5.02. The pupal cDNA probe is shorter than that prepared from the other mRNAs.

Besides these template variables there are other, intrinsic differences between the reverse Northern and Northern techniques that may account for some of the differences observed. With the reverse Northern technique secondary structure can interfere with the procession of the reverse transcriptase along the mRNA molecule so yielding short probes, which as noted previously (Section 5.2.2.2) do not hybridise well. This problem is avoided in Northerns as the RNA molecule is completely denatured and the single stranded DNA probe is normally far in excess of the RNA therefore all regions of the molecule are accessible for hybridisation.

Although these discrepancies between the Northern and reverse Northern data exist, all of the transcribed regions (with the exception of two regions which are discussed below, Section 5.7.7 and 5.7.8), which were hybridised to Northern blots detected transcripts. As a negative control for the transcribed region analysis, two regions that were not detected by the reverse Northern analysis (the λ952C5B region and a 4·6kb EcoRI fragment from λH973) were hybridised to Northern blots. The positive controls were successful but no signal was detected on the Northern blots after three weeks of autoradiography. This suggests that regions from across the 952 walk that the analysis suggested were not transcribed are indeed not transcribed. Therefore even though there is some discrepancy between the reverse Northern and Northern data the analysis was a success in that it could distinguish between transcribed and non-transcribed sequences and the regions detected by this study probably represent all (or at least most of) the transcribed regions across the walks.

The L1 probe appears to have given the most spurious hybridisation to phage sequences *i.e.* to lambda arms (Figure 5.04). This is possibly due to the length of the cDNA and amount of radiolabel incorporated as it is known that longer transcripts give higher levels of background hybridisation (Amersham Technical Reports, 1991). It would appear that all of the probes prepared for this study were satisfactory, but some of them with slightly better incorporation and longer transcript lengths, hybridised to some bands in a non-specific fashion. (Alternatively it may be that some regions gave spurious hybridisation for some reasons). Similar results to this have

been observed previously using similar protocols (S. Tomlinson, S. Russell, C. Milligan, personal communications).

5.7.3 Sensitivity of Reverse Northerns.

It has been calculated empirically that a reverse Northern analysis could detect transcripts to a level of 0.01% (Hall *et al.,* 1983). The conditions used by this study for the preparation of cDNA probes has been significantly modified since the Hall *et al.* study (Berger and Kimmel, 1987). For instance, the RTase enzyme used in this study was a commercially available, genetically engineered one, with much better processivity and much reduced exonuclease activity. Similarly, the chase reaction to increase probe length (Section 2.5.13.4) was not performed by Hall *et al.,* and the increased levels of radiolabel in the probe (four times the amount used by Hall *et al.,*) will also increase the levels of detection. This is supported by comparisons between the signal intensities of the Actin 5C and the walk genomic clones. The actin clones gave too strong a signal to be used for densitometry after just 12 hours (Figure 5.01), as the signal was out of range of the linear response curve of the preflashed film (Hahn, 1983). The faintest bands detected with the reverse Northern blots took over four weeks to appear and were only just visible after this time. This suggests that the level of transcripts which could be detected was better than 0.01% and probably around the same levels found to be detectable by Knipple *et al.,* (1991) in a similar study at polytene subdivision 26A. This was *ca.* 0.0003% of polyA$^+$ RNA.

5.7.4 The Number of Transcripts at 19E3.

The 19E3 region has one essential gene which encodes at least four transcripts (Crompton *et al.,* 1992;1995; Krishnan *et al.,* 1993;1995) with a further eleven transcribed regions surrounding it, six of which are known to encode at least eight transcripts (this study, Table 5.1). This large number of transcripts detected in this study appears to be in keeping with numbers observed in other studies both from distant, and from neighbouring regions of the genome (see Section 1.3.2).

5.7.5 Detection of Rare Transcripts.

One benefit of a reverse Northern analysis is that if several rare messages that, due to their rarity, are not visible individually on a Northern blot have their 3′ ends in a common genomic region, this region may be visualised as the collective signal from

the radiolabelled 3′ ends of several messages. It is thought likely that this is the most probable explanation for the results obtained with the 12.0kb genomic region from the clone λ952-04C9. The Northern analysis detected a single head specific transcript that was not visible in any other developmental stage. The reverse Northern analysis, however, suggests that this area has a complex developmental profile with different genomic fragments giving rise to transcripts expressed in embryos, larvae, pupae and adults. Spurious hybridisation to genomic features could not explain these results. This is because the pattern of hybridisation is inconsistent. If the hybridisation was due to small transcribed repetitive regions, (such as the *opa* repeat), it would be expected that every probe used would hybridise to this region (and they would also be detectable in the repetitive DNA analysis). The regions would also hybridise to all of the cDNA probes if hybridisation occurred to genomic features such as AT rich stretches *i.e.* every probe (as all are primed with oligo dT) should hybridise as would be expected if the hybridisation was due to a short transcribed, repetitive sequences in the region. At least four transcripts are now known to originate from this location, most of which have been isolated using PCR based strategies because of their rarity in oligo-dT primed cDNA libraries. It is possible that the 3′ ends of most of these transcripts originate from this 12.0kb region. This scenario would fit the explanation given above. The size of the (incomplete) *shak-B* transcripts to date range from 2.2kb (neural transcripts) to 2.4kb (vital transcripts). The 1.6kb transcript is therefore either a novel transcript from *shak-B* or it may originate from a separate locus that is very close to (or even interdigitated with) the *shak-B* locus. It will have to await the isolation and characterisation of this 1.6kb transcript before this question can be answered.

5.7.6 The Absence of Signal With KE2 and Related Transcripts.

KE2 (and other transcripts from this region) have been used several times as probes for Northern blotting experiments, and all have failed to detect any transcripts that are known to be present in this region (*e.g.* Figure 5.17). There are several possible explanations for this. The first and most obvious is that these transcripts are rare. That the *shak-B neural* transcript (Krishnan *et al.*, 1995) is rare is supported by the codon preference result for the neural protein. It stays consistently below normal (<1.0) throughout its entire length. The *shak-B vital* and *ogre* messages however exhibit a typical *Drosophila* codon preference throughout their open reading frame. In highly

expressed genes, codon usage is biased towards prevalent tRNAs, but in less expressed genes, alternative codons are used more frequently (Ikemura, 1985). Due to the rarity of at least some of the transcripts from the *shak-B* region an analysis is currently being performed using PCR based approaches (D. Crompton, 1995). This however, cannot be the whole explanation as it is felt that the level of expression seen during *in situ* hybridisation of *shak-B* exons to embryonic and also to pupal tissues (Crompton *et al.,* 1995) should have been detectable on Northern blots used in this study. This is because the embryos were collected in three hour developmental 'windows' and although the pupae were collected in longer, 48 hour 'windows', there should have been enough RNA present for signal to have been visible on a Northern blot. Other factors therefore may be causing the lack of Northern signals. One possibility is that at least some transcripts from this region (including the more common ones) may be non-polyadenylated. Two main lines of evidence would seem to support this; i) the lack of signals on hybridisation to poly A^+ RNA relative to the amount of tissue expression; ii) the internal priming of all cDNAs isolated from this region to date. If it is unpolyadenylated the enrichment of polyA^+ RNAs will have little effect upon signal intensity, although one would have to postulate that there are relatively large amounts of AT rich sequences from which binding to oligo dT can occur for it to be within the cDNA libraries at all, (though this is the case as shown by sequence analysis) and as evidenced by the internal priming. This would also explain why the region is detectable with reverse Northerns, the same internal priming off AT rich stretches may be occurring. A third possibility is that the transcripts may be very unstable. mRNA stability is now recognised as an important post-transcriptional control process (Belasco and Brawerman, 1993). The problem with this explanation is that the half life of the transcripts would have to be prohibitively short for none at all to appear in isolated poly A^+ RNA. Another possibility is that the signal detected during *in situ* hybridisations may not be due to hybridisation to just one transcript. As no negative controls were performed by Crompton *et al.,* (1995) it is impossible to rule this out, and, bearing in mind the large number of OPUS homologues *in C. elegans* (ten to date; Starich *et al.,* 1995) there are presumably more to be found in *Drosophila.*

5.7.7 The 10.6kb Region from λH962.

One possible explanation for the absence of signals with the 10.6kb region from λH962 that was thought possible was that the region is a transposable element that is absent (or practically so) in the *OrR* strain used for the RNA isolation. However the positive control used for the m23 hybridisation (Figure 5.16) contained DNA from three different *Drosophila melanogaster* strains and no differences (apart from a few faint bands that are almost certainly due to polymorphic restriction sites) were visible. The fact that the cDNAs (obtained from an adult cDNA library) from this region have introns shows that they are genuine transcripts. The positive control for the hybridisation worked and the Northern blot itself was successful, as evidenced by the *rp49* control. No explanation is therefore readily apparent for the lack of signal, although it is possible that it is similar to the *shak-B* region, in that transcripts from this area may be rare and so not be detectable with a Northern blot. The results from the reverse Northern analysis would go some way to support this as the pattern of expression differs quite markedly between the proximal 8.0kb λH962 fragment and the distal 2.5kb λM291 fragment. It may be therefore that this area has two genes situated within it both of which are expressed in adults neither of which is strong enough to give a signal on a Northern but which together are strong enough to give a signal on a reverse Northern.

Chapter Six

Chapter Six - Analysis Of Several Transcribed Regions From The 952 (19E3) Region.

6.1 Introduction.

In this study four chromosomal walks were localised to specific regions of the *Drosophila* X chromosome (Chapter 3). Twenty eight different transcribed regions within these walks were identified and characterised to at least some degree, using reverse Northern and Northern blots (Chapter 5). Analysis of the occurrence and type of repetitive sequences within the walks was also performed (Chapter 4) which in conjunction with the transcript detection study allowed single copy, transcribed regions to be identified. It was thought likely that these regions would contain gene sequences. Several of these transcribed regions were chosen for further analysis.

Two criteria were used for choosing the regions to be studied. The main one was the location of the transcribed region. Regions lying near to deficiency and / or duplication endpoints that were known from previous work to affect the *shak-B* locus, were of more interest to this study than lying outside of the *shak-B* region. The second criteria was the developmental profile of an area *i.e.* where and when in development messages from a region were expressed. This study was principally interested in genes involved with the nervous system. Such genes are likely to be expressed at the embryonic and pupal stages of the fly's life-cycle, when the nervous system is developing and / or undergoing extensive metamorphosis. In adult flies the head is principally composed of neural tissues (brain and eyes) and thus head specific genes are also likely to be involved with the nervous system.

The regions chosen were analysed in a variety of ways. Some were used to screen cDNA libraries, in an attempt to isolate clones corresponding to the message(s) originating from the regions. Genomic sequence was obtained from other transcribed regions and used to search the Genbank (Release 84 [8/94]) and EMBL (Release 39 [6/94]) databases. This was to search for homologies, so giving clues as to the type of message(s) originating from a region. The final approach was to perform *in situ* hybridisation experiments to *Drosophila* head sections using transcribed genomic

regions as probes. In this manner it was possible to determine the mRNA expression patterns within this tissue.

In this part of the study I describe how several regions were analysed in more detail than described previously (Chapters 3 - 5) and what conclusions can be drawn from the results obtained.

6.2 Results.

6.2.1 cDNA Library Screening.

Five regions were chosen for this part of the study (plus the 1.0kb λAGO02 fragment - see Section 4.2.4)and several different cDNA libraries were screened with various transcribed genomic regions as probes (see Table 6.1). The libraries were screened using conventional screening techniques with nitrocellulose being the support membrane. All screens were performed in duplicate. Some positives were detected with several of the clones, however all except one failed to appear in duplicate and were thus discarded after a primary round of purification. Fragments from the 12.0kb region from λ952-04C9 were used as probes and hybridised to a single clone in the *eya* head library which yielded a 0.65kb clone. This clone, and the 6.0kb region it originated from, were then used to screen a further 500,000 pfus from the same library. This further screen identified one other clone of 2.6kb. The 0.65kb and 2.6kb clones (called pI1 and pC10 respectively) were hybridised to a Southern blot of the 952 walk. Both were found to hybridise to the same 6.0kb *Eco* RI fragment from λ952-04C9 (Figure 6.01). The two clones and the 6.0kb region, were digested with several restriction enzymes and the resulting restriction maps compared. The maps of the genomic region and the two cDNA clones were in agreement (Figure 6.02) except for an *Xho* I site present in pC10, which was not present in the genome, and a *Hin* dIII site in the genome that was not present in pC10. The most proximal and distal terminal fragments from pC10 were isolated by restricting with *Xho* I / *Ava* I (to isolate the distal 3′ end) and *Xho* I / *Eco* RI (to isolate the proximal 5′ end), radiolabelled, and hybridised to Southern blots of the 6.0kb region which had been restricted with several different enzymes. The distal 0.20kb *Xho* I / *Ava* I fragment (common to both pI1 and pC10) hybridised as expected, however the most proximal 0.7kb *Xho* I / *Eco* RI fragment from pC10 failed to hybridise to the 952 walk (or to the

Transcribed Region	cDNA Library Screened	No of PFU Screened
1.0kb *Eco* R I fragment from λAGO02	*eya* head	60 000
	Stratagene Adult	60 000
	Pupal	60 000
1.7kb *Eco* R I fragment from λ952-C5B	Stratagene Adult	250 000
	Pupal	60, 000
2.0kb *Hin* dIII fragment from λ952-08C9	Stratagene Adult	250 000
	Pupal	60, 000
1.0kb *Eco* R I fragment from λ952-08	Stratagene Adult	60 000
	Pupal	60, 000
12.0kb region from λ952-04C9, composed of four *Eco* R I fragments.	Stratagene Adult	250 000
	eya head	700 000
	Male head	150 000
3.2kb *Sal* I region from λ9915	Stratagene Adult	60 000
	eya head	180,000
	Pupal	60, 000

Table 6.1 Table listing the cDNA libraries used, and the approximate number of plaque forming units (PFU) screened, in experiments to isolate cDNA clones from different transcribed regions across the 19E3 region. The **Stratagene Adult** library was prepared from whole adult flies and the insert size averaged around 1000bp, though it was later found to be heavily contaminated with genomic sequences. The **Pupal** library was prepared from mid - late staged pupae and known not to be completely representative, the insert sizes averaged around 1000bp. (S. Russell, personal communication). The ***eya* head** library was prepared from head tissue of *eyes absent* flies. The sequence representation was good and the insert sizes were known to be large (>1000bp) (C. Milligan, personal communication). The **male head** library was prepared from the heads of male *CS* flies. The clones were known to be short (average size 400bp) but the sequence representation was good (S. Russell, personal communication).

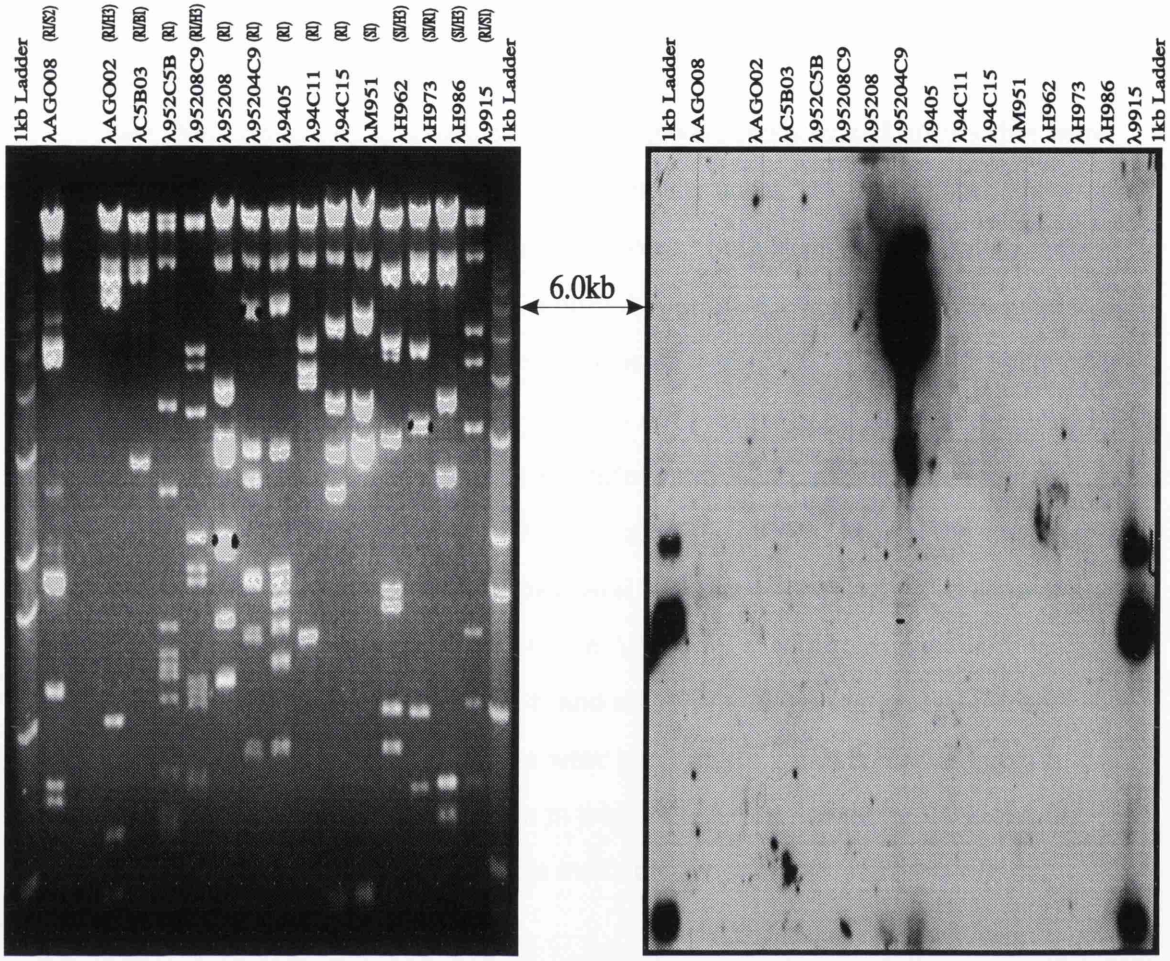


Figure 6.01-Hybridisation of the clones pC10 and pI1 to the 952 walk. Only a 6.0kb *Eco* RI fragment from the genomic clone λ952-04C9 was detected. The enzymes used to restrict the phages are shown in brackets. RI= *Eco* RI, H3 =*Hin* dIII, Sc2= *Sac* II, BI= *Bam* HI and SI= *Sal* I.

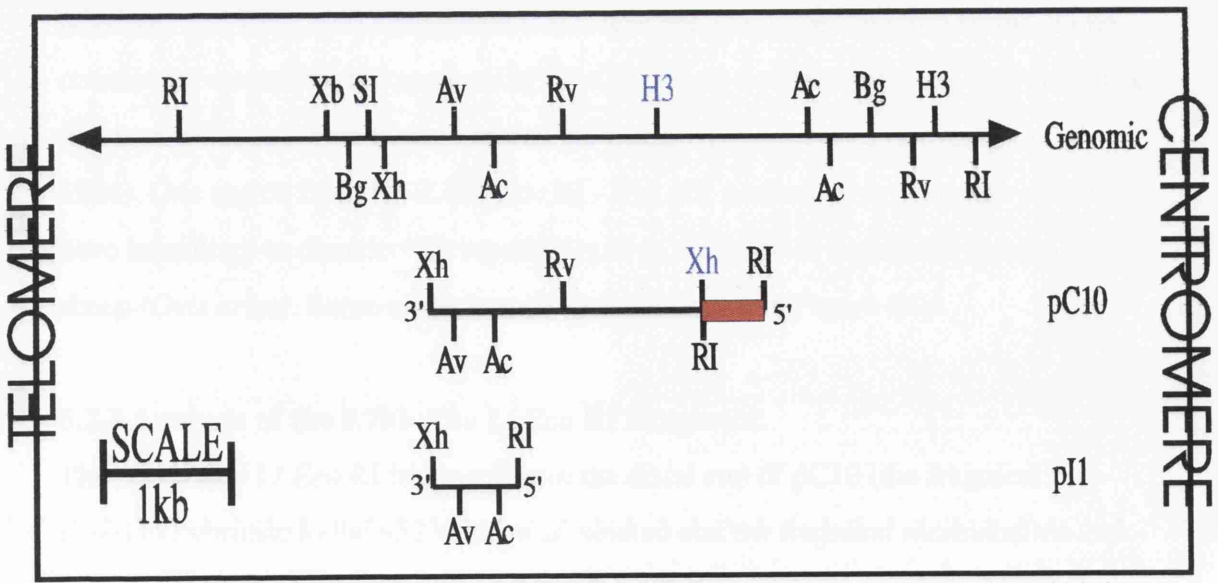


Figure 6.02-Comparison of the restriction maps of the clones pI1 and pC10 to the genomic region from which they originate. The *Hind* III and *Xho* I sites in blue are not present in both the genome and the clone pC10. The proximal fragment at the end of pC10 (in red) is not from the 952, or 896 walk (see Section 6.2.3).

896 walk). This fragment was later shown to be a separate cDNA that had become ligated onto the sequence that did hybridise to the 952 walk (see Section 6.2.3).

6.2.2 Sequence Analysis of pI1 and pC10.

The *eya* cDNA library from which pI1 and pC10 were isolated had inserts that were directionally cloned, with a 3′ *Xho* I restriction site and a 5′ *Eco* RI site. Sequencing of the 3′ ends of pC10 and pI1 showed both to have identical start sites. The 3′ ends differed in that pC10 had a poly A^+ tail of nineteen A residues whereas pI1 had one fewer with 18. No other discrepancies were detected across the sequences except for one G at position +9 in pI1 which is not present in pC10. Neither clone had the polyadenylation signal AAUAAA (Proudfoot and Whitelaw, 1988). Sequencing of the genomic region corresponding to the 3′ end of the clones showed the sequence AAAAAAAAAATCAAA, which aligned well with the poly A^+ tail. The sequencing rationale and resulting sequence is shown in Appendix 5. Subfragments were isolated and subcloned, from both pC10 and pI1, and also from the genomic region from which they originated. *Exo* III deletions were prepared of a 2.3kb *Hin* dIII genomic fragment, to obtain access to the region in which the 5′ end of the pC10 clone lay (Figure 6.03) as no restriction sites were available for subcloning in this region.

One open reading frame of 90 amino acids was identified spanning the *Eco RV* site in the middle of pC10, though the lack of a good translation consensus sequence (Cavener and Ray, 1991) suggested that it was not a real open reading frame. Time constraints meant that the sequencing of pC10 could not be completed. The sequences that were obtained were used to search the GenEMBL databases (Devereux *et al.*, 1984). One region from the 3.4kb *Eco* RI - *Hin* dIII genomic fragment was shown to have homology to dinucleotide repeats found in a variety of organisms *e.g.* man and sheep *(Ovis aries)*. Some of the homologies are shown in Figure 6.04.

6.2.3 Analysis of the 0.7kb *Xho* I / *Eco* RI Fragment.

The 0.7kb *Xho* I / *Eco* RI fragment from the distal end of pC10 (the fragment that failed to hybridise to the 952 walk) was isolated and the fragment radiolabelled and hybridised (Sections 2.5.10, 2.5.13.2 and 2.5.17) to a genomic Southern blot to determine if the fragment was repetitive. The genomic blot was composed of wild-

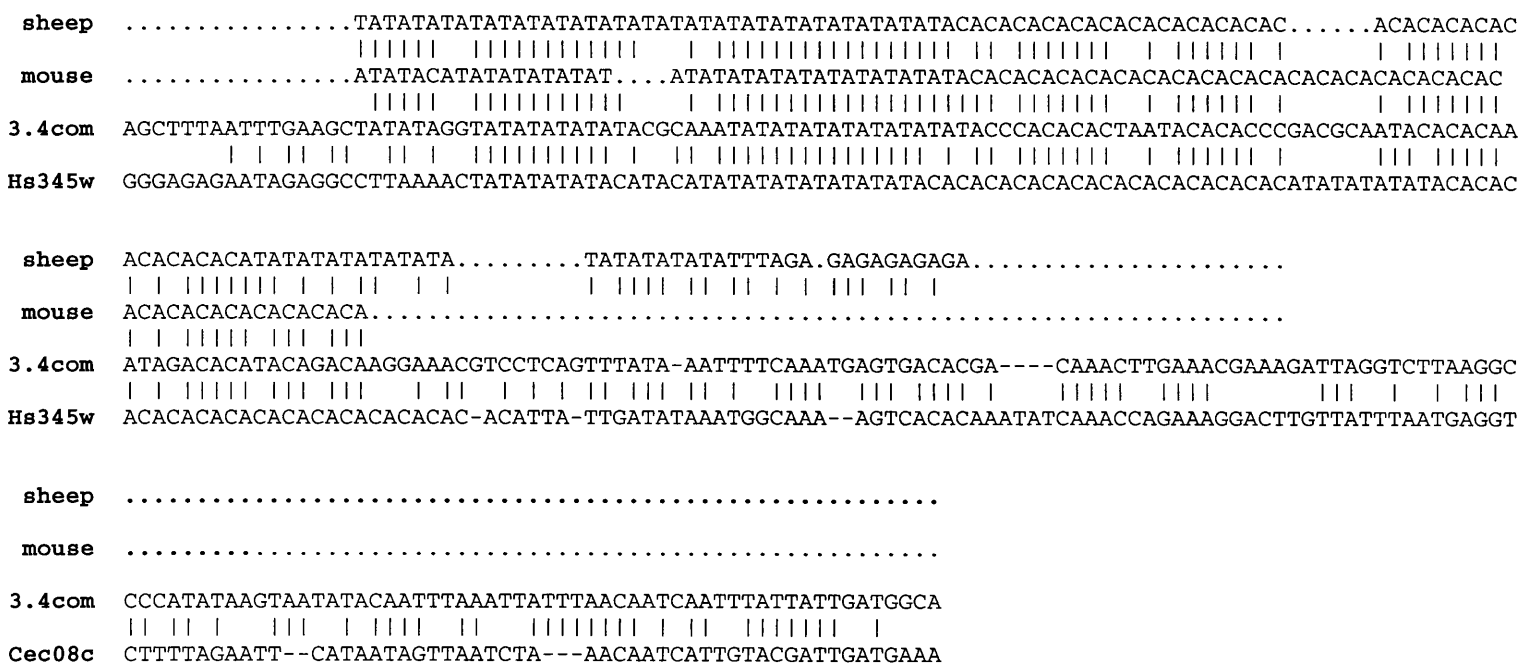


Figure 6.04-Homology of a sequence from the 3.4kb *EcoR* I - *Hind* III fragment from the genomic clone λ952-04C9 to several sequences from the database. **Cec08c** is from a *C. elegans* cosmid; the **mouse** sequence is from a gastric (H^+,K^+)-ATPase; the **sheep** sequence is from a dinucleotide repeat region, whereas **Hs345w** is from a human microsatellite sequence. **3.4com** is the sequence identified by this study. All of the homologies were to dinucleotide repeat sequences in mammals, no homologies were detected to *Drosophila* dinucleotide repeat sequences.

type *Sierra Leone* (SL) *Canton-S* (*CS*) and *Oregon R* (*OrR*) DNAs, restricted with *Eco* RI, *Hin* dIII and *Bam* HI. A single band was detected in all three strains and with all three restriction enzymes used (Figure 6.05A), thus suggesting the fragment contains only single copy sequences. The same fragment was then hybridised to a Northern blot composed of RNAs isolated from different developmental stages. A strong signal was observed that corresponded to a transcript of *ca.* 750bp which was present in all stages of the fly's life cycle (Figure 6.05B).

6.2.3.1 Sequence Analysis of the 0.7kb *Xho* I / *Eco* RI Fragment.

Dideoxy sequencing of both strands of the 0.7kb *Xho* I / *Eco* RI fragment was performed using T3 and T7 primers (Section 2.15). The cDNA insert is 707bp long of which 28bp is polyA$^+$ tail. It has a 3′ non-coding sequence of 79bp, a 5′ non-coding sequence of 48bp and a single open reading frame of 552bp. In the other two reading frames the sequence is interrupted by termination codons. The open reading frame begins at an AUG codon at position 47 and ends at a TAG termination codon at position 598. The polyadenylation signal, AAUAAA (Proudfoot and Whitelaw, 1988) is 30bp from the poly A$^+$ tail. The translation start is close to the *Drosophila* consensus sequence, *i.e.* TACATACCACA[**AUG**] vs (C/A),A,C, (A/C), (A/G), (A/G), (A/C/T),C,A,(A/C),(A/C) [**AUG**] (Cavener and Ray, 1991).

The resulting sequence was used to search the GenEMBL databases (Genbank, Release 84 [8/94]; EMBL, Release 39 [6/94]). The nucleotide sequence and the amino-acid sequence (Figure 6.06) showed strong homology to ribosomal proteins isolated from many different species and phyla. The best homologies at the amino-acid level were to ribosomal proteins in humans and rats, although other homologies almost as good were detected across four kingdoms. Homologies were detected to animals (the *rpL11* proteins of humans and rats); to plants (the *rpL5* protein of alfafa); to eubacteria (the *rpL21* protein of *Tetrahymena thermophila*), and to archebacterium (the *rpL5* protein of *Halobacterium marismortui*). These homologies are displayed in Figure 6.07. The homologies detected to the sequence showed that the complete coding region was contained in the clone.

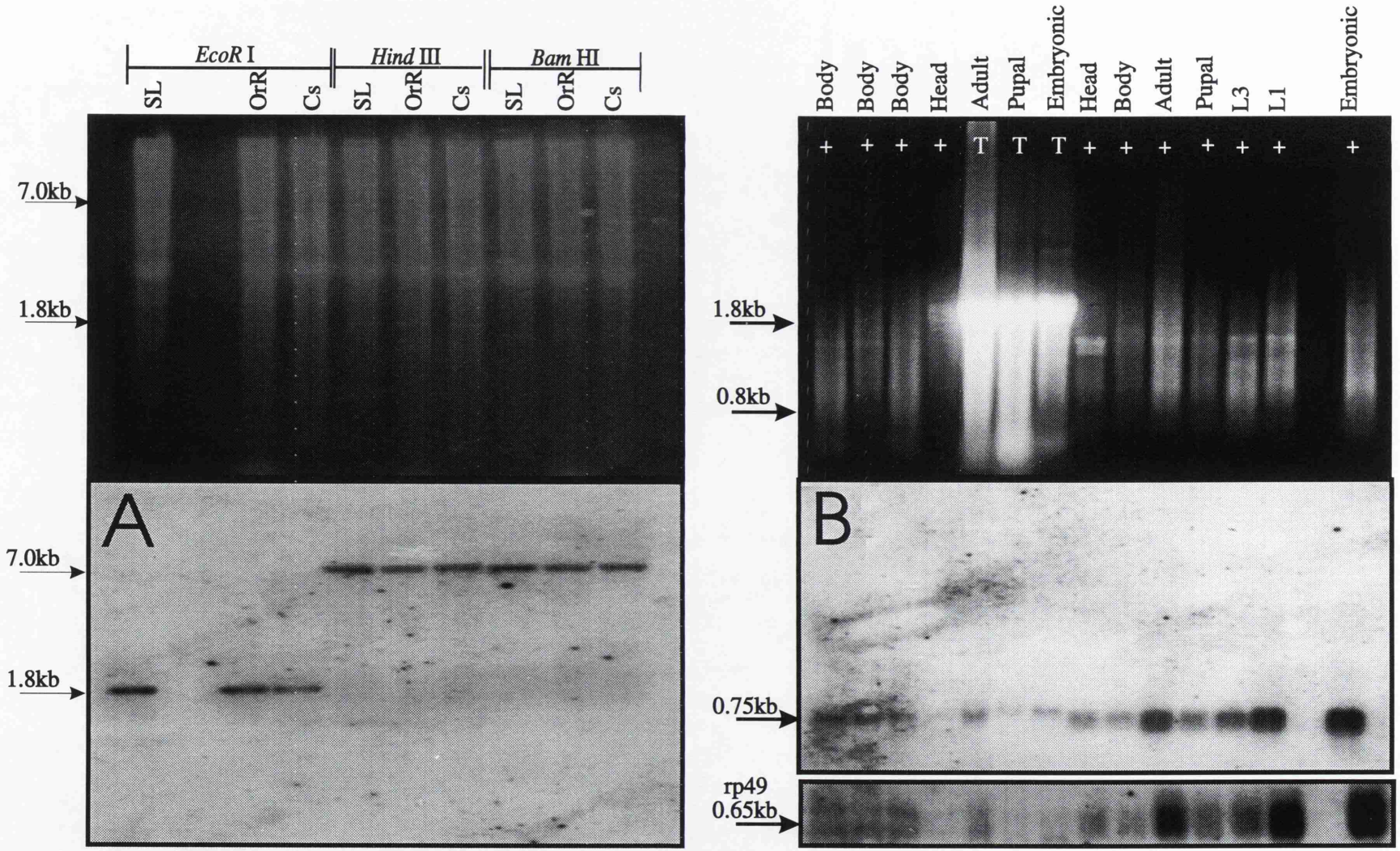


Figure 6.5 Hybridisation of the 0.7kb *Eco* RI / *Xho* I fragment from pC10 to, (A) a genomic Southern blot of wild type DNA from three *Drosophila* strains; (B) a Northern blot of polyA+ RNAs (labelled with +) and total RNAs (labelled with T) from different developmantal stages of *Drosophila.* The panel below B shows the same blot probed with *rp49* as a loading control, there is some residual hybridisation visible from the 0.7kb *Eco* RI / *Xho* I fragment hybridisation.

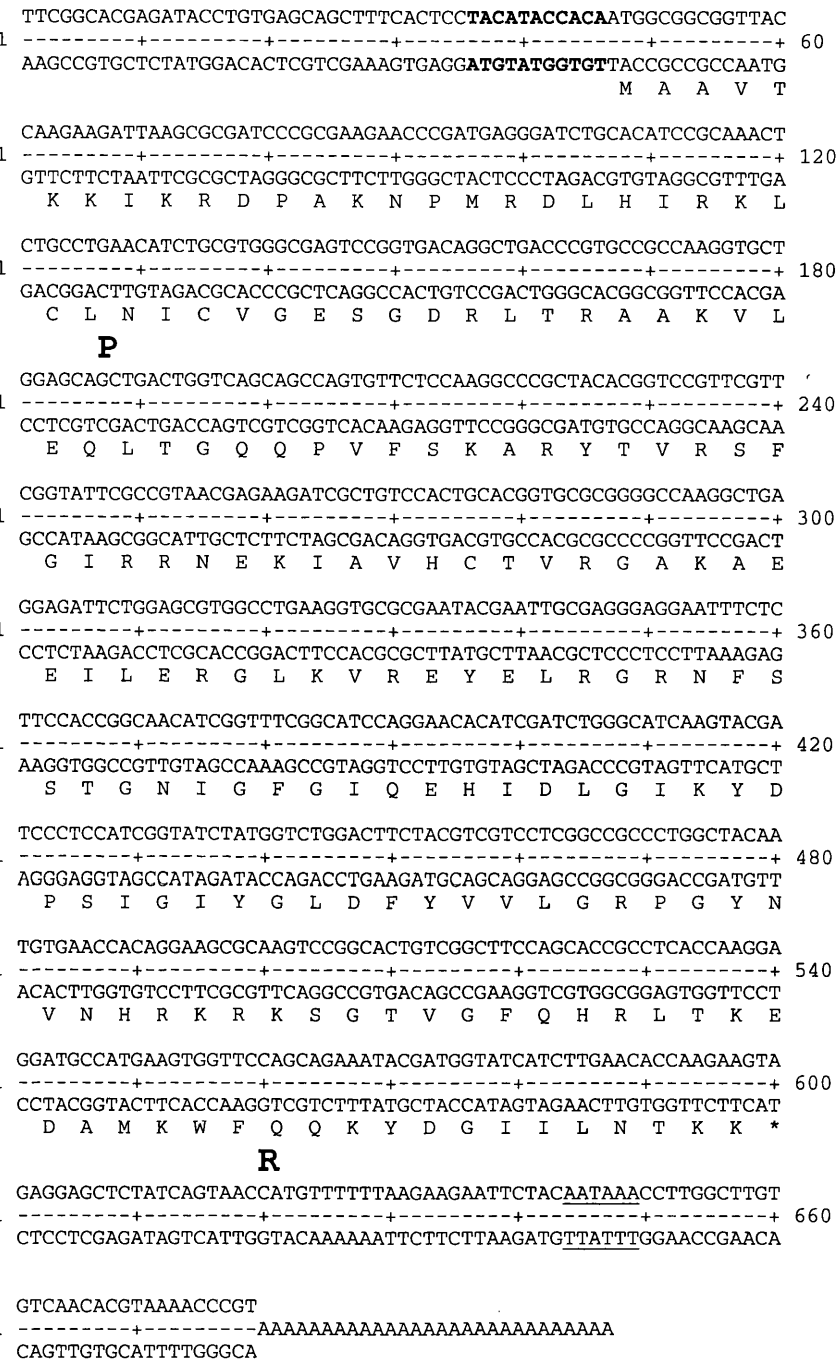


Figure 6.06--Sequence of the *Drosophila melanogaster*, *rpL11* gene, isolated from an *eya* head cDNA library. The polyadenylation signal is underlined at position 643 to 649. The translation start sequence is highlighted in bold at positions 36 to 46. **R** and **P** indicate the positions of *Eco* R I and *Pvu* II sites within the clone.

6.3 Analysis of Genomic Sequences from Transcribed Regions.

Three regions were chosen for this analysis plus the 1.0kb λAGO02 fragment -see Section 4.2.4 (Table 6.2). The main criteria by which these regions were selected was proximity to *shak-B i.e.* located in the 952 walk. Another criteria by which these regions were selected for study was their small size. This was to increase the likelihood of detecting the sequence(s) responsible for hybridisation of the region during the reverse Northern and Northern analysis. The regions chosen were isolated from the appropriate phages and ligated into pBluescript KS+ (Sections 2.5.6 to 2.5.7) and sequencing reactions performed (Section 2.15). The sequences obtained were used to search the GenEMBL databases at both the nucleotide level, using the Fasta and Blast commands, and at the amino-acid level after translation in all six frames, using the TFasta and TBlast commands. None of the sequences obtained (except the *opa* repeat in the 1.0kb fragment) had homology to repetitive sequences in the databases.

6.3.1 The 0.8kb *Eco* RI / *Bam* HI Fragment from λ95208C9.

No homology was detected at either the nucleotide or the amino acid level to sequences from the 0.8kb fragment from λ95208. The genomic sequences (from both ends of the fragment), together with all open reading frames are displayed in Appendix 5.

6.3.2 The 1.0kb *Eco* RI / *Eco* RI Fragment from λ95208.

The search with the sequence from the 1.0kb λ95208 fragment detected no homologies at the nucleotide level. The Frame 3 translation however was shown to be homologous (22% identity and 70% homology) to an uncharacterised human genomic sequence from chromosome 21 (Database Accession Number: D12986, Figure 6.08). The genomic sequence from both ends of the fragment, with all open reading frames is displayed in Appendix 5.

6.3.3 The 3.2kb *Sal* I / *Sal* I Fragment from λ9915.

Homology was detected using the nucleotide sequence from the 3.2kb *Sal* I fragment (65% identity over *ca.* 100bp, to an uncharacterised genomic sequence from the *Antennapeadia* complex (Database Accession Number: L32626) from *Drosophila melanogaster.* Other homologies almost as good were detected in a variety of other organisms, all had approximately 60% identity to around 100bp of the sequence. The homologies are unusual however in that they are spread across 260bp

Genomic Region	Stage detected in northern / reverse northern blot.	Size(s) of transcripts seen	ORFs present	Amount of sequence	Homologies detected in database and section reference.
1.0kb *Eco*RI λ95208	**(RN)** -A **(N)** -N/D	N/D	Frame 3	198bp	Homology to an unidentified human sequence using Frame 3 search (Id 22%; Hom 70%).
0.8kb *Eco*RI-*Bam*HI λ95208	**(RN)** -A+P **(N)** -A(Head)+P	1.4kb	Frame 1 Frame 3 Frame 5	258bp	No homology detected
1.0kb *Eco* RI λAGO02	**(RN)** -E, L1, L3, P, A, **(N)** -Smears	Individual bands not visible.	Frame	852bp	No homology detected except to *opa* repeat although one o.r.f. (135 amino acids) extends past end of current sequence
3.2kb *Sal*I λ9915	**(RN)** -P **(N)** -L3+P+A	2.15kb	Frame 1	287bp	65% homology at nt. level to unknown genomic sequence from *Antennapaedia*. Frame 1 homology to mammalian tyrosine kinases (Id 29%; Hom 64%)

Table 6.2-Information and results about and from the transcribed regions and the sequences obtained from them which were used in the sequencing analysis (Section 6.3).

```
           1                                                        50
humanl11   ..........  .......... MAQDQGEKEN PMRELRIRKL CLNICVGESG
  ratl11   ..........  .......... MAQDQGEKEN PMRELRIRKL CLNICVGESG
 DmrpL11   SARDTCEQLS LLHTTMAAVT KKIKRDPAKN PMRDLHIRKL CLNICVGESG
   alf15   ..........  .......... .MASEKKLSN PMREIKVQKL VLNISVGESG
yeastl16   ..........  .......... ...MSAKAQN PMRDLKIEKL VLNISVGQSG
   hma15   ..........  .........MS SESESGGDFH EMREPRIEKV VVHMGIGHGG
     tet   KLKRKQN*IQ KEKKDNLIQV LAEMTDKKEN KMREVKIAKL VINCCVGESG
consenseus                      -- ---------N pMRelrirKl vlnicvGesG

           51                                                       100
humanl11   DRLTGAAKVL EQLTGQTPVF SKARYTVRSF GIRRNEKIAV HCTVRGAKAE
  ratl11   DRLTRAAKVL EQLTGQTPVF SKARYTVRSF GIRRNEKIAV HCTVRGAKAE
 DmrpL11   DRLTRAAKVL EQLTGQQPVF SKARYTVRSF GIRRNEKIAV HCTVRGAKAE
   alf15   DRLTRAAKVL EQLSGQTPVF SKARYTVRSF GIRRNEKIAC YVTVRGDKAM
yeastl16   DRLTRASKVL EQLSGQTPVQ SKARYTVRTF GIRRNEKIAV HVTVRGPKAE
   hma15   RDLANAEDIL GEITGQMPVR TKAKRTVGEF DIREGDPIGA KVTLRDEMAE
     tet   DKLTKAAKVL KDLSGQEPVF SRARYTIRSF GIKRNEKMAV HVTIRGDKAR
consenseus drLtrAakvL eqltGQtPvf sKAryTVrsF gIRrnekIav hvTvRgakAe

           101                                                      150
humanl11   GI........  ..........  ..........  ..........  ..........
  ratl11   EILEKGLKVR EYELRKNNFS DTGNFGFGIQ EHIDL.GIKY DPSIGIYGLD
 DmrpL11   EILERGLKVR EYELRGRNFS STGNIGFGIQ EHIDL.GIKY DPSIGIYGLD
   alf15   QLLESGLKVK EYELLRRNFS DTGCFGFGIQ EHIDL.GIKY DPSTGIYGMD
yeastl16   EILERGLKVK EYQLRDRNFS ATGNFGFGID EHIDL.GIKY DPSIGIFGMD
   hma15   EFLQTALPLA E..LATSQFD DTGNFSFGVE EHTEFPSQEY DPSIGIYGLD
     tet   DILTRGLKVK EMELRKKNFS NTGNFGFGIQ EHIDL.GMKY DPSTGIFGMD
consenseus eiLe-gLkv- EYeLr-rnFs dTGnfgFGiq EHidl-gikY DPSIGIYGlD

           151                                                      200
humanl11   ..........  ..........  ..........  ..........  ..........
  ratl11   FYVVLGRPGF SIADKKRRTG CIGAKHRISK EEAMRWFQQK YDGIILPGK*
 DmrpL11   FYVVLGRPGY NVNHRKRKSG TVGFQHRLTK EDAMKWFQQK YDGIILNTKK
   alf15   FFVVLERPGY RVGRRRRCKA RVGIQHRVTK DDAMKWFQVK YEGVILNKSQ
yeastl16   FYVVMNRPGA RVTRRKRCKG TVGNSHKTTK EDTVSWFKQK YDADVLDK*.
   hma15   VTVNLVRPGY RVAKRDKASR SIPTKHRLNP ADAVAFIEST YDVEVSE*..
     tet   FYVVLERPGT RVARRRRATS RVGNNQMISK EECINWFKTE FEGNVY*EIK
consenseus fyVvl-RPGy rv--rkr--- -vg--Hr-tk edam-wfqqk Ydg-il----

           201                        227
humanl11   ..........  ..........  .......
  ratl11   ..........  ..........  .......
 DmrpL11   *RSSISNHVF *EEFYNKPWL VSTRKTR
   alf15   AIV*......  ..........  .......
yeastl16   ..........  ..........  .......
   hma15   ..........  ..........  .......
     tet   NTYTINNIII *NYNIYFFSK LLNKRGL
consenseus ----------  ----------  -------
```

Figure 6.07-An alignment of proteins related to *Drosophila* L11 protein. The sequence alignments were performed using the program, PILEUP, from the GCG sequence analysis software package (Devereux *et al.*, 1984). In the **consensus** sequence, invariant amino acids are in capital letters, small case letters indicate conservative changes. The sequences it is being compared to are **human** and **rat** *rpL11*(Chan, 1992), *Medicago sativa rpL5* (**alfafa**-Asemoto *et al.*, Unpublished); **yeast** *rpL16* (**yeastl16**-Leer *et al.*, 1984); *Halobacterium marismortui rpL5*, (**Hmal5**-Scholzena and Arndt, 1991); *Tetrahymena thermophila rpL5* (**tet**-Rosendahl *et al.*, 1991), and the *Drosophila rpL11* homologue identified by this study (**DmrpL11**).

HUME116 D12986 Human genomic DNA, (sequence-tagged *Not* I)
(Identity 22%, Homology 75% in 56 aa overlap)

```
1fr3    HRXTRAGGGVKVKTMTRKLAIIVAPAPVRGGRSDLHTNQTKRNIYQNNSDHWGEIRLQIGWQIDI
          :.: :..    ..  .  ........ ..   .:.....:..... :. :.: : :::
HUME11  RLGWRGGKGLRRVRSSPGRPRYLQRSAIPAPWRRVHGARARRRLFPRVSXTWSEGRKQIGKLQRP
```

Figure 6.08-The amino acid homology between the 1.0kb *EcoR* I fragment sequences from λ95208 and the uncharacterised human, *Not* I tagged, sequence from the database. The levels of identity and homology are as indicated; (:) identical amino acids; (.) homologous amino acids. See text (Sections 6.3 and 6.5.1 for more details).

of the 287bp of sequence *i.e.* the casein A homology is from 1 - 135bp, whereas the *Euglena* sequence has homology from 179bp to 260bp.

The 3.2kb sequences (from both ends) were translated in all six frames and these putative amino acid sequences were then used to search the Swissprot protein database (Release 29; 8/94) using the TFasta program (Devereux *et al.,* 1984). There is only one open reading frame that remains open throughout the entire 287bp (Frame 1). A second frame remains open for 83 codons. No homology was detected to the frame 4 sequence however the frame 1 translation suggested some homology to tyrosine-kinases. The degree of relatedness ranged from 22-28% identity and *ca.* 64% homology over the first 70 amino acids of the 95 residues coded for by the sequence (Figure 6.10). The genomic sequence from both ends of the fragment, with all open reading frames is displayed in Appendix 5.

6.4 *In situ* Hybridisation of Genomic Fragments to *Drosophila* Head Sections.

A 12.0kb region, composed of four contiguous *Eco* RI fragments from λ952-04C9 were isolated from agarose gels and the individual fragments (6.0kb + 2.5kb + 1.8kb + 1.7kb) were then labelled with digeoxygenin (Section 2.5.10 and 2.17.3). The 12.0kb region was known to hybridise to a head specific transcript of 1.6kb (Section 5.4.5), although several other transcripts, undetected in the Northern analysis (presumably because of their rarity), also originated from this region (Section 5.6). The hybridisation conditions were carried out as described in Section 2.17.

After visualisation of the colour reaction (Section 2.17.5) the fly head sections were photographed. The results are presented in Figure 6.11. Signal was detected in the lamina of the eye and a more general staining was also detected in the central brain. This general staining was only visible when high probe concentrations were used (*i.e.* two or three probes per hybridisation reaction) and it is felt this was probably a probe artefact and did not represent actual hybridisation. The experiment was repeated and the hybridisation to the lamina of the eye was again observed.

KSYK_Pig q00655 Tyrosine-protein kinase *syk* (Identity 23.8%, Homology 68% in 80 aa overlap

```
3fr1    SFDAEDFGLSAFSLLNALLANCIPHFPAIFHFPPTLSGHFPTGHRCWQRTVRLRW--DSYSR-VVNYNNFGLNLGKWGFG-HVWECNSQDFFHQE
             .:  ....  :.::.  ..:.. : .:  .::  .........  .. .:  . :..  .X: .:. .  . :.:X  .::
KSYK_P YLEECNFVHRDLAARNVLLV--TQHYAKISDF--GLSKALRADENYYKAQTHGKWPVKWYAPECINYYKFSSKSDVWSFGVLMWEAFSYGQKPYR
      477                                                                                  567
```

A47333_Human A47333 T-cell-specific tyrosine kinase (Identity 22.4% Homology 66% in 67 aa overlap

```
        1                                                                 93
3fr1    SFDAEDFGLSAFSLLNALLANCIPHFPAIFHFPPTLS-GHFPTGHRCWQRTVRLRWDSYSRV-VNYNNF
         ....:::.. : : ..  ..  ..:X. .  : ..: .....    :  .: :.:. .:   . :.:
A47333 VIKVSDFGMTRFVLDDQYTSSTGTKFPVKWASPEVFSFSRYSSKSDVWSFGV-LMWEVFSEGKIPYENR
      494                                                                        561
```

KATK_Mouse q06187 Tyrosine-protein kinase *atk* (Identity 28.6%, Homology 64% in 63 aa overlap

```
3fr1    SFDAEDFGLSAFSLLNALLANCIPHFPAIFHFPPTLS-GHFPTGHRCWQRTVRLRWDSYSRV-VNYNNFGLN
          ::::: . : ..  ..  ..::. .  : .X  ..:..    :. .: :.:. ::   . :..:.
KATK_H VVKVSDFGLSRYVLDDEYTSSVGSKFPVRWSPPEVLMYSKFSSKSDIWAFGV-LMWEIYSLGKMPYERFTNS
      534                                                                          597
  561
```

MYR3_sinal p29092 Myrosinase mb3 precursor (Identity 21.4%, Homology 70% in 84 aa overlap

```
3fr1,   SFDAEDFGLSAF-SLLNALLANCIPHFPAIFH--FPPTLSGHFPTGHRCWQRTVRLRWDSYSRVVNYNNFGLNLGKW-GFGHVWECNSQDFFHQEST
                 ::. . .:..X::.. :..: ..:X  .:.:: .... .     .: .    ...:. .   ...:: .. .: .......   ....
MYR3_S SRGVDQAGLDYYHNLIDALLEKNITPFVTLFHWDLPQTLQDEYEG---FLDRQIIQDFKDYADLC-FKEFGGKVKNWITINQLYTVPTRGYALGTDA
```

Figure 6.10-Amino acid comparisons between several tyrosine kinases and the 3.2kb *Sal* I fragment sequences from λ9915. Levels of identity and homology are as indicated. The lower comparison is to the Myrosinase enzyme (*aka* thioglucosidase) from *Brassica napus*, (Rape) which is responsible for glucosinolate degradation. See text (Sections 6.3 and 6.5.1.3) for more details and references.

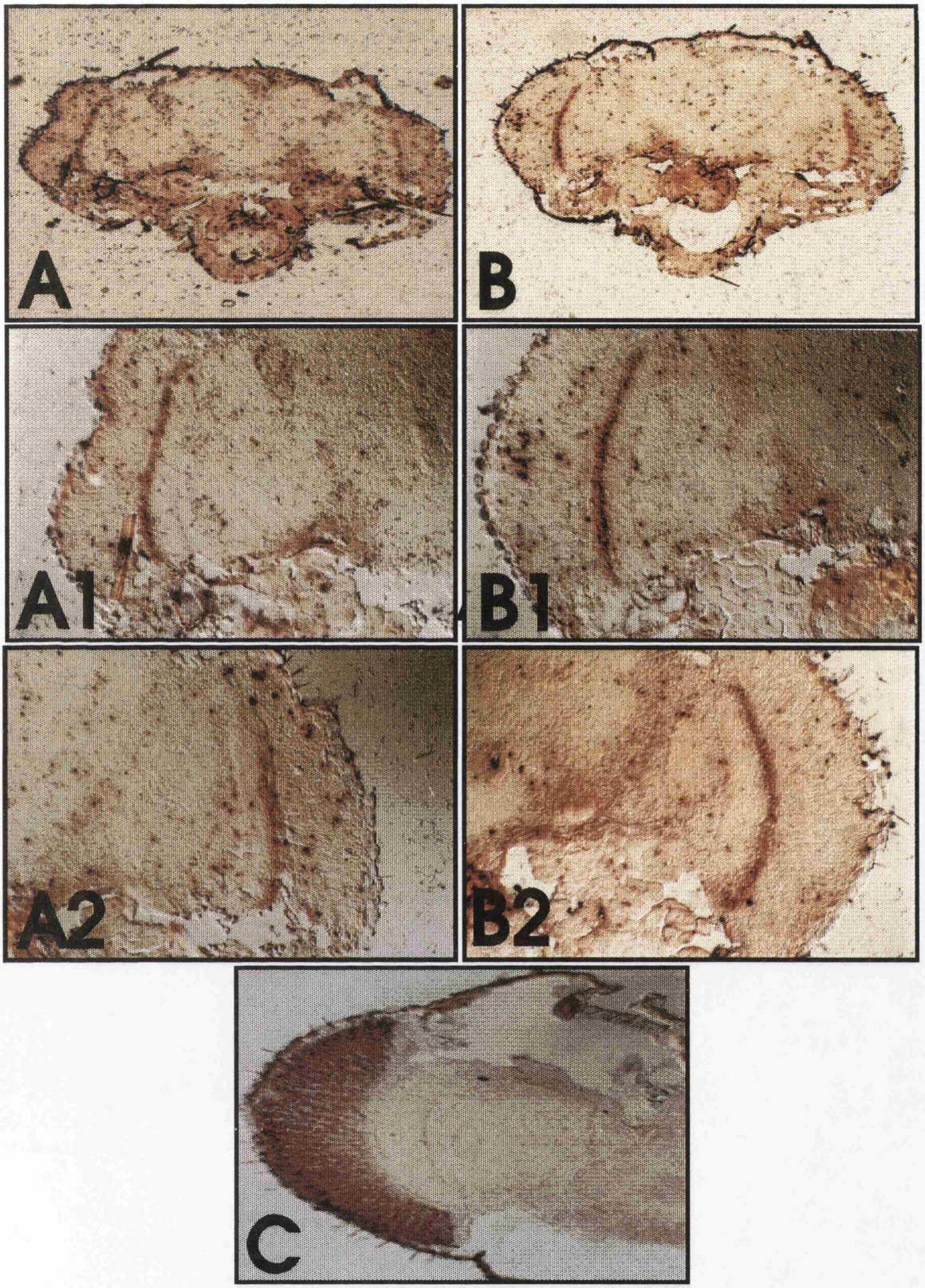


Figure 6.11-*In situ* hybridisation of genomic fragments to *Drosophila* 10μm thick head sections. A and B were hybridised with the 11.7kb region from the genomic clone λ952-04C9. Expression is detected only in the lamina of the eye. C is a positive control for the hybridisation, which uses the opsin clone, *ninaE* (O'Tousa *et al.*, 1985) as a probe. As expected, strong expression was only observed in the eye.

6.5 Discussion.

6.5.1 The cDNA Library Screens.

The cDNA library screens were performed under standard conditions with positive and negative controls which were successful for each probe. Bearing in mind the large number of pfus that were screened it is perhaps unfortunate that more clones were not isolated. However the screens were successful in that two clones were obtained from the *eya* head library. It is unfortunate that the *in situ* hybridisation results were obtained only after the majority of the screening was completed. It is now felt that a wild type adult head library (*i.e.* prepared from flies with eyes) would have been the best library to have used.

6.5.1.1 The cDNA Clones pC10 and pI1.

The region from which these two clones originate is known to contain a 1.6kb head specific transcript which may have a function at the *shak-B* locus. Besides this transcript several others, undetectable using Northern analysis, are known to terminate in this region and are known to be part of the *shak-B* locus (Section 1.2.8 and Sections 5.6). The region has also been shown to be a source of transcripts expressed in the lamina of the eye (Section 6.4), but how the pI1 and pC10 clones relate to these other transcripts is unknown. pC10 is too big to be the 1.6kb transcript and Northern blot analysis showed no cross hybridisation to this, or any other transcript. This suggests two possibilities. The first is that pC10 and pI1 are members of the group of cDNAs that have their 3′ ends located in the 6.0kb region but which are undetectable using Northern analysis. The second possibility is that both clones are genomic contaminants of the library which both fortuitously, primed internally off the same region of contaminating genomic DNA.

Several workers have used the *eya* head cDNA library for extensive cDNA screens (C. Milligan, S. Thomlinson, S. Goodwin, personal communications) and have not reported finding any genomic sequences. The library was primed using an *Xho* I - oligo dT primer with mmLV as the reverse transcriptase. The second strand was prepared using standard RNaseH / DNA polymerase techniques. The 3′ and 5′ ends were then blunt ended with Klenow and *Eco* RI adapters ligated to both ends. The analogue 5-methyl dCTP was used in the first strand nucleotide mixture. This hemi-methylation was used to protect *Xho* I sites within the cDNA as clones were then

restricted with *Xho* I. In theory the methyl group protects the cDNA from restriction enzymes used in subsequent steps. So the only sites that should be restricted are those occurring in the *Xho* I - oligo dT primer. The cDNA inserts were then directionally cloned into the phage arms with *Xho* I 3′ and *Eco* RI 5′. This method would prevent the creation of a 5′ *Xho* I site in the clones unless, a) it existed within a genomic contaminant, (which must have stayed double stranded throughout the whole procedure and did not incorporate methylated dCTP as if it was single stranded it would have incorporated methylated dCTP and hence the *Xho* I site would not have been restrictable); b) the hemi-methylation was not a 100% successful so allowing some internal sites to be restricted (C. Milligan; Stratagene Technical Support, personal communications). No *Xho* I site is present in the genome at the position of the 5′ end of pC10, this would seem to preclude the first explanation, and suggest that pC10 and pI1 are genuine cDNAs that are both internally primed from the same A rich region. As well as this the hemi-methylation was known not to have been 100% successful as *bona fide* cDNAs have been obtained from the library as one of a double insert with the clone having a 3′ and a 5′ *Xho* I site (C. Milligan and C. Hutchinson, personal communications). In summary therefore it is known that pC10 has a double insert, with one of the fragments being a transcript from a ribosomal protein gene. The remaining 2kb of the insert is from the 952 walk but has a 3′ and a 5′ *Xho* I sites which should have been impossible using the cloning technique described.

The two discrepancies between the pC10 clone and the genome, namely the 5′ *Xho* I site present in pC10 but not in the genome, and a *Hin* dIII site in the genome but not present in pC10, remain unresolved. Sequence was obtained for both regions but not in both cases *i.e.* the sequence around the *Hin* dIII site was obtained from the genome but not from pC10 whilst the sequence surrounding the *XhoI* site was obtained from pC10 but not from the genome. However, in both cases sequence very close to the restriction sites was obtained, and these were found to be identical in both pC10 and the genome when compared using the BestFit program (Devereux *et al.*, 1984). It is therefore possible that both discrepancies are due to nucleotide polymorphisms between the two strains as they do not appear to be caused by the crossing of intron / exon boundaries.

It was hoped to establish the identity of pC10 and pI1 by sequencing them and searching the database with the resulting sequences. The sequences were used to search the GenEMBL databases (Genbank, Release 84 [8/94]; EMBL, Release 39 [6/94]). One region of homology was detected to a repetitive dinucleotide sequence. This result was unexpected in that $(TA)_n$ dinucleotide repeats are rare in *Drosophila* relative to $(CA)_n$ repeats and also because dinucleotide repeats are structural and are rarely if ever transcribed. It is known that β-heterochromatic regions of most of the *Drosophila melanogaster* chromosome arms, and especially chromosome 4, are deficient in long tracts of mono- and dinucleotide sequences such as $(C)_n$ and $(CA)_n$ (Pardue *et al.,* 1987; Lowenhaupt *et al.,* 1989). However, few studies have looked at the frequency of short dinucleotide repeats, as demonstrated by the lack of database homologies found to *Drosophila* sequences. During the repetitive DNA analysis (Chapter 4, Figure 4.01a) this region was shown to hybridise very faintly to nick translated probes. The homology detected in the database search therefore supports the supposition that the faint hybridisation observed in this and other bands, may in part be accounted for by hybridisation to short repetitive sequences (Section 4.3.4). Most of the database 'hits' were to sequences from man and from the mouse. Dinucleotide repeats occur about once every 60kb in the human and mouse genome. Although their origin and function in the genome remains unclear, in humans and mice CA dinucleotides have been implicated in the formation of Z-DNA and also as 'recombination hotspots' to promote the evolution of the genome (*e.g.* Hellman, 1993); this idea however remains controversial. No other homology was detected to the pC10 clone or the genomic sequence obtained from this region.

pC10 and pI1 are quite AT rich, a characteristic of 3' ends of eukaryotic mRNAs (Belasco and Brawerman, 1993). In addition to this, with the exception of a very small open reading frame of 90 residues around pC10's *Eco* RV site, the sequences contain no open reading frame. The fact that the pC10 clone is *ca.* 2.0kb does not rule out the possibility that it is a real 3' end, as extensive 3' untranslated sequences are not uncommon in developmentally regulated genes where they are thought to be involved in post-translational regulation (Hultmark *et al.,* 1986). Both clones would have to be internally primed, which appears to be quite common in this region. None of the transcripts identified from the *shak-B* locus (five to date) have a convincing poly A tail, *i.e.* they are all internally primed.

It is not possible to state categorically whether pI1 and pC10 are genuine transcripts or whether they are genomic contaminants. One possible way to resolve this question would be to carry out *in situ* hybridisation experiments using pC10 as a probe. Time constraints however meant this experiment could not be performed.

6.5.2 The Genomic Sequences.

All of the genomic sequences isolated by this study were put through *Drosophila* codon preference programs (Ikemura, 1985) as well as other programs developed to determine partial open reading frames in eukaryotic sequences. This was done to determine the likelihood of them containing coding DNA. (These programs would not indicate if the sequences are from the 3' ends of the sequences *i.e.* in a legitimate transcribed region but not coding which is a likely as all of these regions were identified by a reverse Northern analysis). The results obtained were inconclusive in that the programs could not determine if the sequences were coding or if they were non-coding sequences that happened to contain open reading frame. In some cases this was because the optimum sequence length for analysis was in excess of the sequences obtained (*e.g.* the codon preference program (Staden *et al.*, 1982; Gribskov *et al.*, 1984) looks at the sequence using a 200bp window and so it has difficulty coping with sequences smaller than 400bp.

6.5.2.1 The 0.8kb *Eco* RI / *Bam* HI Fragment from λ95208C9.

This fragment is part of a 1.7kb *Eco* RI / *Eco* RI fragment that was shown to be expressed in adult and pupal tissues. It is known that both *Eco* RI / *Bam* HI fragments from the larger fragment contain transcribed sequences and so it was decided to sequence the middle of the fragment *i.e.* from the *Bam* HI site. Three open reading frames were identified that extended for the full 258bp of the sequence obtained next to the *Bam* HI site (sequence and open reading frames shown in Appendix 5). Although over half of the fragment (including the 163bp of sequence from next to the outer *Eco* RI site) had been sequenced and three open reading frames identified, no homology was detected to sequences from the databases searched. If one of the open reading frames identified is actually part of the transcript detected by the Northern analysis then the transcript has no homologues in the databases at present.

6.5.2.2 The 1.0kb *Eco* RI / *Eco* RI Fragment from λ95208.

This fragment was not analysed by Northern analysis. Reverse Northern analysis suggests it was expressed in adults only. One open reading frame (Frame 2, see Appendix 5) extended throughout the 198bp of sequence obtained, whereas one other, Frame 3, extended for 51 codons. This latter frame had homology to a human genomic sequence that was isolated using *Not* I sequence tagging. This technique involves restricting genomic DNA with *Not* I and another more common enzyme to produce smaller DNA fragments. These fragments are then ligated into a relevant vector from which the insert can be sequenced. As *Not* I sites tend to occur within CpG islands at the 5′ promoter regions of human genes (Haines, 1993), this approach has been adopted as part of the Human Genome Mapping Project to increase the probability of detecting transcribed genomic sequences. The high level of homology (70%) would tend to support the idea that a transcript derived from this region is either structurally or evolutionary related to a gene in man. Little else can be said about this homology as both sequences are currently uncharacterised.

6.5.2.3 The 3.2kb *Sal* I / *Sal* I Fragment from λ9915.

The homology detected at the nucleotide level was to several different organisms (see Figure 6.10) and was consistent, *i.e.* approximately 65% over *ca.* 100bp. Its location in the sequences to which it had homology were in coding regions (with the exception of the *E. gracilis* gene in which it is located in an intron). The reason for its conservation is unclear. To try and explain the somewhat confusing results the sequences obtained (from both ends of the fragment) were translated in all six frames and used to search the protein data bases. It is felt that homologies detected at the amino acid level are more credible than those detected at the nucleotide level. Nucleotide similarities may arise due to chance as apparently similar DNA sequences can code for different proteins depending upon the frame used to produce the protein.

The amino acid translation detected homology to a variety of tyrosine kinases and also to Myrosinase (*i.e.* Thioglucosidase), an enzyme from *Brassica napus* (Rape) which is responsible for glucosinolate degradation. The largest known protein family (now termed a superfamily) is made up of protein kinases which have been identified largely from eukaryotic sources. These enzymes use the γ phosphate of ATP (or GTP) to generate phosphate monoesters utilising protein alcohol groups (on serine and

threonine) and / or protein phenolic groups (on tyrosine) as phosphate group acceptors. They are related by virtue of their homologous kinase domains also known as catalytic domains which consist of 250-300 amino acids residues (Hanks *et al.,* 1988; Hanks and Quinn, 1991; Hardie and Hanks, 1995). Phosphorylation of tyrosines on proteins plays a central role in the regulation of cell growth and development, and a large number of tyrosine kinases have been identified in *Drosophila* (Perrimon *et al.,* 1993), in other invertebrates (*e.g.* Chan *et al.,* 1994), and also in vertebrates (Sudol *et al.,* 1993; Ruiz and Robertson, 1994).

The homologies in the tyrosine kinases are all in frame, and within coding regions, but away from the SH2 and SH3 domains (which are amongst the most conserved parts of the proteins, Waksman *et al.,* 1992). The top scores were to regions that included the kinase domains of the proteins. The kinase domains of eukaryotic protein kinases impart the catalytic activity and three roles can be ascribed to the kinase domains; 1)binding and orientation of the ATP (GTP) phosphate donor; 2)binding and orientation of the protein (or peptide) substrate; 3) transfer of the γ phosphate from ATP (GTP) to the acceptor hydroxyl residue on the serine, threonine or tyrosine, of the protein substrate. The kinase domains can be sub-divided into twelve smaller subdomains (indicated by Roman numerals), defined as regions never interrupted by large amino acid insertions and containing characteristic patterns of conserved residues. They tend to be invariant (or nearly so) throughout the superfamily and are hence strongly implicated as playing essential roles in the function of the enzyme.

The homologies to the tyrosine kinases was detected to the same regions in all of the hits, the region encompassing subdomains VII, VIII and IX. Subdomain VII folds into a β-strand-loop-β strand structure and contains a highly conserved DFG triplet. Subdomain VIII folds into a complex chain that faces the cleft in the protein (Knighton *et al.,* 1991). It contains a highly conserved APE motif. This domain appears to play a major role in the recognition of peptide substrates. Subdomain IX contains a nearly invariant aspartic acid residue. It corresponds to a large α-helix on the largest loop of the proteins.

The 3.2Fr1 sequence contains the DFG triplet plus a further two residues (L and S) that are also quite highly conserved across the superfamily (the G of the DFG triplet

corresponds to residue 186 of the catalytic domain). Neither of the conserved sequences in subdomains VIII or IX (the APE sequence at 206 or the D residue at 220) are present. This suggests that the sequence, if it really is translated, is not encoding a kinase. It is known that this region is expressed in pupae and also in adult heads which suggests that it may have a function in the nervous system. Without further study it is not possible to say just what the region is encoding however, or why the homology, if not the identity, to the tyrosine kinase superfamily is so high *i.e.* 60-70%.

The homology between the sequence and the thioglucosidase open reading frame is surprising. No explanation could be found for it in the literature and there is no relationship between the kinase superfamily and thioglucosidases.

6.5.3 The Ribosomal Protein.

6.5.3.1 Ribosomal Proteins in *Drosophila*.

The ribosomes (and their subunits) from *Drosophila* are typical of eukaryotes (Chooi, 1981; Scofield and Chooi, 1982). Many of the ribosomal proteins have been biochemically analysed (Chooi, 1980, 1981; Chooi *et al.,* 1980, 1982) and in *Drosophila* a large number of the genes encoding them have been cloned (*e.g. L1* Rafti *et al.,* 1988; *L17A* Noselli and Vincent, 1992; *S2* Barrio *et al.,* 1993; *S3* Wilson *et al.,* 1993; *S4* Yokokura *et al.,* 1993; *S6* Spencer and Maki, 1993; *S14* Andersson *et al.,* 1994; *S17* Maki *et al.,* 1989; *S19* Baumgartner *et al.,* 1993; *S26* Itoh *et al.,* 1989b *S31* Itoh *et al.,* 1989a *rpA1* Qian *et al.,* 1987; *rpA2* Olsen *et al.,* 1993; *rp17* McNabb and Ashburner, 1993; *rp21C* Wigboldus, 1987; *rp49* Kongsuwan *et al.,* 1985; see Ashburner, 1992 for review). As a group they are of note, not only because of their primary role in protein translation, but also as a model system for studying translational discrimination (*e.g.* Hongo and Jacobs-Lorena, 1991; Patel and Jacobs-Lorena, 1992). As part of the translational machinery, they are of interest because of the effects they have on other processes, *e.g.* tumorigenicity (Stewart and Denell, 1993) and oogenesis (Cramton and Laski, 1994).

6.5.3.2 The Size of the rp11 Protein.

The *rp11* homologue identified by this study is a member of a highly conserved group of proteins, the members of which are found in four different kingdoms (*e.g.* see Figure 6.07). The protein DmrpL11 encodes has an excess of basic residues in

common with other members of this group (*e.g.* the rat rpL11 protein, identified by Chan *et al.,* 1992). The predicted size of the polypeptide is also very similar. This estimated size assumes that the start methionine is situated within an open reading frame at nucleotide 47 (Figure 6.06). Two lines of evidence suggest this residue to be the correct start site. Firstly, it sits in a local consensus sequence context aligning well with other *Drosophila* translation start sites (see Section 6.2.3). Secondly, the estimated amino acid length from this methionine (184 residues) is close to that of the rat L11 protein to which it shows homology (178 amino acids Chan *et al.,* 1992). The only alternative start site is situated some 17 residues downstream; this has no good translation consensus sequence and is situated within a highly conserved region. It is therefore likely to have an important role within the protein. Unless the *Drosophila* rpL11 protein is much larger than those of other organisms, (which is unlikely bearing in mind the degree of conservation of ribosomal proteins), the methionine indicated must be the correct start. It is therefore possible to conclude that the cDNA contains the complete coding region of the *Drosophila melanogaster rpL11* homologue (*DmrpL11*). It would be interesting to carry out a chromosomal location of the *DmrpL11* cDNA to see if it hybridises to a region known to contain a *Minute* mutation (see below).

6.5.3.3 Ribosomal Gene Mutations in *Drosophila*.

Minutes are a group of over 50 mutations scattered throughout the genome which are associated with similar dominant visible phenotypes. When homozygous or hemizygous, *Minute* mutants are lethal, and the fly dies at the late embryonic or early hatching stage of development. Heterozygous *Minute* mutants have short thin bristles and a prolonged larval development time that can vary from a few hours to several days in extreme *Minutes*. Other associated phenotypes include reduced viability, roughened eye structure, small body size and etched tergites (Lindsley and Zimm, 1992).

The large number of *Minute* loci, and the observation that the phenotype of flies heterozygous for more than one *Minute* is not additive, led investigators to suggest that all *Minutes* loci encode ribosomal proteins (Vaslet *et al.,* 1980; Kay and Jacobs-Lorena, 1988). Although this is likely to be so in the majority of cases, it has now

been demonstrated that all *Minutes* do not in fact encode ribosomal proteins (see below).

Several ribosomal protein genes (Burns *et al.,* 1984; Brown *et al.,* 1988; Noselli and Vincent, 1992; Kongsuwan *et al.* 1985; Andersson *et al.,* 1994) have been mapped to chromosomal locations known to contain *Minute* mutations. Qian *et al.,* (1988) and Patel and Jacobs-Lorena (1992) have also demonstrated that anti-sense interference of *rpA1* gene expression can mimic *Minute* phenotypes. However, no direct evidence is available for a general correlation between ribosomal proteins and *Minutes.* To date, only two *Minutes* have been shown to encode ribosomal proteins, these being *rp49* (Kongsuwan *et al.* 1985) and *rpS3* (Andesson *et al.,* 1994). Several workers have in fact shown that some ribosomal proteins that were previously thought to encode Minutes proteins due to their cytological location, do not actually do so (*e.g.* Kay *et al.,* 1988; Qian *et al.,* 1988). It seems much more likely that *Minutes* are genes involved in a common cellular process and probably not solely ribosomal proteins. This is supported by observations that the mutations *bobbed* and *mini*, caused by mutations in other genes involved in translation (rRNA and 5S RNA respectively), have phenotypes similar to that of *Minute* genes (Ritossa, 1976; Procunier and Dunn, 1978).

6.5.3.4 Possible Roles for the rpL11 Protein in the Ribosome.

The results showing high levels of amino acid homology across the very diverse organisms (displayed in Figure 6.07), reflect the fundamental importance of these ribosomal proteins, and suggests that rpL11 was a component of the early ribonucleoprotein ribosome. The function of this protein within the fly's ribosome (and those of other eukaryotes) remains unclear. It is known that in eubacteria, three proteins bind to 5S rRNA; in *E.coli* they are L5, L18 and L25 (Huber and Wool, 1984). In eukaryotes only one protein, L5 has been shown to bind to the nucleic acid (see Huber and Wool, 1986). Some anomaly exists however, as the eubacterial L5 proteins have homology with eukaryotic L11 proteins (*e.g.* this study), but eukaryotic L11 proteins are not known to associate with 5S rRNA. To complicate this situation further the eubacterial L5 proteins do not share homology with the eukaryotic L5 proteins. Further study is needed to clarify this confusion.

6.5.4 The *in situ* Hybridisation Results.

6.5.4.1 The 12.0 kb Region Fragment from λ95204C9.

As P2.4 and other transcripts proximal to this 12.0kb region are all internally primed (Section 1.2.8.1) it is therefore possible that the 3′ end of P2.4 and other cDNAs, such as N52 and S2.2, may originate from this 12.0kb genomic region. Other *shak-B* neural transcripts (Crompton *et al.,* 1995) are also expressed in the giant fibres at least two days post eclosion. The lamina expression is interesting as it may have a correlation with a phenotype from *shak-B* mutants. As discussed in Section 1.2.4, *shak-B* mutant flies have a defective electroretinogram (ERG). The type of ERG defect present in *shak-B* flies, an alteration in the transient component, is known to be caused by defects in the neurones of the lamina of the eye (*e.g.* Coombe, 1986). It is therefore possible that the defect in the ERG, is caused by some alteration of a transcript expressed in the very tissue that the 12.0kb region hybridised to. It is unknown whether the 1.6kb transcript is responsible for this expression in the lamina. However, as this is the only transcript detected upon Northern analysis of the region, and as the amount of eye expression would seem to correlate with the amount of expression on the Northern blots *i.e.* both visible, it would seem logical for this to be the case. Results using antibodies to the protein produced from the S2.2 cDNA (Section 1.2.8.4) have shown expression of S2.2 protein in the lamina of the eye (M. Wilkin personal communication). The S2.2 clone is 2.6kb in length and therefore it cannot be the 1.6kb transcript. It is possible that this may be a more common splice variant from the locus. More study is needed to clarify this situation.

The expression patterns observed in the head indicate that transcripts detected in the *in situ* hybridisations using the 12.0kb region may have a role in the *shak-B* locus *i.e.* transcripts from this region are expressed in the lamina of the eye, where mutations in at least one of the transcripts could be causing the defects associated with the ERG. Alternatively, there may be two different loci in this region, it being merely fortuitous that the more common gene is also expressed in the lamina of the eye, as is the *shak-B* neural transcript(s). However as no negative controls were included in this paper it is not possible to eliminate the possibility that the probes used were cross hybridising with other members of the OPUS family (see Section 1.2.8.3).

Chapter Seven

Chapter Seven - Discussion.

This project had two main aims. Firstly to delineate the region in which the *shak-B* gene lies by chromosomal walking and by deficiency and duplication breakpoint analysis. The second aim was primarily to attempt to determine the actual molecular location of the *shak-B* transcripts (and any other unrelated transcripts contained in the cloned walk genomic DNA), but also to examine the molecular organisation of the region, lying as it does in an cytologically interesting region that is a transition zone from euchromatic to heterochromatic DNA.

7.1 Chromosomal Walking and the Chromosome Aberration Breakpoints.

Two of the four chromosomal walks used in this work were localised by this study to specific polytene subdivisions using deficiency and duplication breakpoint mapping. The 798 walk was localised to 19E1 (Section 3.4) and shown to contain the *mell* locus (Section 3.); the *runt* walk was known to be located at 19E2 as it was started using a clone from the chromosomal walk that cloned the *runt* gene (Gergen and Weischaus, 1986). The 952 walk was localised to 19E3 (Sections 3.5 - 3.8) whereas the 896 walk was localised to 19E4 by the finding of the distal breakpoint of *Df(1)T2-14A* (C. Taylor and J. Davies personal communications). These four walks therefore cover a large part of the distal half of the 19E3 subdivision. This is a fascinating region lying as it does on the border between the euchromatic and heterochromatic regions of the base of the X chromosome in *Drosophila melanogaster.*

Taken together these four walks constitute one of the largest areas analysed at the molecular level in *Drosophila* as the combined amount of cloned DNA from all four walks is in excess of 450kb. None of the walks however have yet joined up which is unfortunate for a number of reasons. Firstly, being able to analyse the whole region from a cytogenetic viewpoint would be very informative. For example being able to analyse the way gene incidence, repetitive sequence occurrence, chromosomal abnormality frequency *etc.* alters across the region would be made much simpler if the whole region was cloned as it would be possible to get a true picture of how the genomic landscape varies as the centromere is approached. It would also be interesting to find out just how large the band subdivisions are at the molecular level. The average amount of DNA per band, plus adjacent interbands is 21.6kb with a range

from 5 - 150kb (Ashburner, 1989). This figure agrees well with that calculated from a 315kb chromosome walk in the *rosy - Ace* region (Spierer *et al.,* 1983) which identified 15 bands with a size range of 3.7kb - 160kb. It would however imply that *shak-B* is an unusual locus in that its size is around 200kb. As the 896 and 952 walks have yet to be joined, it is probably more than this. It may be that large band sizes are a feature at the base of the X chromosome as other large band areas have been identified in this region *e.g.* at 20A (Yammamoto *et al.,* 1990).

7.2 The *shak-B* Locus.

Thirty two different transcribed regions were identified by this study of which fifteen were shown to contain non-repetitive sequences. Nine of these regions (from the 19E3 region) were characterised further using Northern analysis, sequence analysis and *in situ* analysis to *Drosophila* head sections (see Chapters 5 and 6). It is felt that a much more detailed analysis of these regions, together with the other regions identified but not characterised, from the 19E3 region would be of value. This would preferably be by isolating transcripts that originate from these regions.

7.2.1 The size of the *shak-B* Locus.

The size of the *shak-B* locus is of note. Transcripts from the locus extend over a region of 30kb, (although this is probably greater as no complete 3′ ends have yet been identified). The genetic interactions that occur between the *shak-B* locus and deficiencies and duplication that impinge upon it however are known to extend much further than this. 120kb separates the region from which the *shak-B* transcripts originate to the distal breakpoint of *Df(1)17-351.* It should also be noted that the distal breakpoint of *Df(1)T2-14A* (which also interacts with the locus) is 64kb proximal to the distal end of the 896 walk. Therefore the interactions stretch at least 184kb (120kb in the 952 walk plus 64kb in the 896 walk). It almost certainly stretches further as more DNA, (containing the distal breakpoints of *Df(1)A118, HC279 and 26B*), lie between the proximal end of the 952 walk used in this study and the distal end of the 896 walk. If this interaction is found to be due to regulatory elements, (as has been suggested, Baird *et al.,* 1993), this would make *shak-B* one of the largest loci yet identified in *Drosophila,* although its size is not unique.

Other loci in *Drosophila e.g.* the *Antennapaedia Complex* and the *Bithorax Complex* (Scott *et al.,* 1983; Duncan, 1987; Kaufman *et al.,* 1990) and *dunce* (Chen *et al.,* 1987; Furia *et al.,* 1991; 1993; Qiu and Davis, 1993) are relatively large and posses a complex organisation compared to the majority of other loci. The *dunce* locus stretches for 148kb and has several non-related genes in its introns. This is a situation that may well exist in the *shak-B* locus *i.e. R-9-28* (and possibly several other genes - see Chapter 5) lie between the locus at 19E3 and the region that interacts genetically with it at 19E4 / 5. Further study will be needed to determine if other parallels between *shak-B* and *dunce* exist. For example *dunce* has a very complex pattern of transcripts, at least eight arise from five transcribed regions. *shak-B* has only one transcribed region identified to date, however at least five transcripts originate from it, all having common 3′ but slightly different 5′ exons (Krishnan *et al.,* 1995).

The *Antennapaedia* and *Bithorax* genes span upto a 100kb and are expressed in complex spatial and temporal patterns. Many of the regulatory sequences have been identified by mutations in *cis*-regulating elements which cause expression of wild-type proteins in subsets of the normal patterns. Some of these mutations are insertions or small deletions that map as far as 100kb away from their promoter. Such long distance interactions (between *cis*-regulatory elements and promoters) have also been observed in many higher eukaryotes including man. It is thought that many of the homeo-box genes *trans*-acting proteins *e.g. zeste*, *brahma* and *trithorax* (see Kennison, 1995 for references) which influence homeotic gene activity facilitate interactions between *cis*-regulatory elements and a promoter by bringing together distant chromosomal elements (Kennison, 1995). This type of *cis*-interaction has been postulated to explain the long distance interactions of the 19E3 region with that at 19E5/6 (Baird *et al.,* 1990).

7.2.2 Isolation of Transcripts from the 19E3 Region.

Although conventional cDNA library screening has been used for many years it has several disadvantages. The main one is that searching for a rare transcript is tedious and time consuming due to the large number of clones and therefore filters that must be screened. Most of the alternatives to conventional library screening utilise PCR. RACE RT PCR (Frohman, 1990) is probably one of the most commonly used techniques for isolating 3′ and 5′ ends of cDNAs. Alternatively by using a small

amount of known sequence (inferred from either protein sequence, homologies with other genes (*e.g.* heterologous screening) or from genomic sequence), it is possible to isolate a novel transcript using RT PCR or direct PCR screening of a cDNA library using a primer to the transcribed sequence and another primer to the 3′ or 5′ ends of the vector.

In at least three (and probably four cases) it may be possible, using PCR to isolate transcripts identified in this study as sequence data is available from regions that are transcribed (Section 6.3). Although there is no definite proof that these sequences are located in a transcribed area it would be worth while attempting using primers designed to the sequence obtained for 3′ and 5′ RACE PCR on either RNA or cDNA libraries. PCR does have some disadvantages (the main one being sequence and length bias- *i.e.* GC rich regions tend to cause stalling of the polymerase whilst long transcripts are often not isolated due to problems with the processivity of the enzyme, its advantages; amplification of rare sequences and its speed (experiments can be completed in a day) means an approach utilising PCR would be by far the easiest option for isolating transcripts from the 19E3 region.

It is common in studies using *Drosophila* to have a large genomic region (*e.g.* the 952 walk) which has one well characterised region (normally the gene of interest to the study) but the surrounding cloned DNA has normally only been partially characterised. Due to the makeup of the *Drosophila* genome (See Chapter four) a particular walk normally has only a few repetitive regions, with several uncharacterised transcribed regions contained within it. A technique termed "direct selection" would allow the isolation of the majority of these transcripts in one experiment. Direct selection was developed initially for screening large areas of human genomic DNA (YAC clones, cosmid contigs *etc.*) but should also work well in *Drosophila*. Direct cDNA selection is a procedure that has been utilised to 'fish out' transcripts originating from a genomic region from a complex mixture of cDNAs (Lovett *et al.*, 1994); Parisimoo *et al.*, 1991; Lovett *et al.*, 1995). Many of the problems associated with the technique in humans simply do not occur in *Drosophila e.g.* very frequently occurring repetitive DNA, pseudogenes *etc.*

The procedure involves hybridising genomic DNA (either immobilised on a filter or alternatively, labelled with biotin which is attached to magnetic beads coated with streptavidin which are immobilised on a magnet), to a solution containing a mixture of cDNAs amplified from a cDNA library by PCR with some of the surrounding vector sequences included. After hybridisation, the filters (or beads) are washed to remove non-specifically bound cDNAs. This washed filter is then stripped by boiling to remove the hybridised cDNAs which are then amplified using primers specific for the flanking vector sequences. This approach has been used successfully in man; in a 900kb cosmid contig 81 cDNA clones were isolated of which 54 mapped back to the contig (Morgan *et al.,* 1992). This approach should work well in the isolation of transcripts from the 19E3 region. Besides allowing the isolation of transcripts from these uncharacterised regions it may also allow the identification of more *shak-B* transcripts as it is not a technique that is dependent upon sequence data. If any transcripts from the region exist that do not cross hybridise to the cDNAs isolated to date and therefore not included in the PCR selection (Crompton, 1995), it may be possible to 'fish' the cDNAs out of the cDNA mixture.

Once transcripts from several regions have been identified it would then be possible to determine which (if any) of the regions are related and also determine which (if any) are associated with the *shak-B* locus.

7.2.3 Isolation of *shak-B* Transcripts if Non-Polyadenylated.

Several explanations have been proposed (Section 5.7.6) to explain this studies lack of success in identifying and characterising the expression of *shak-B* transcripts. Of these suggestions it is felt that that either the rarity of, or non-polyadenylated state of the *shak-B* transcripts is the most likely explanation. There are very few alternatives that can be used if it is rarity of the transcripts that is stopping detection of the transcripts as only PCR based approaches and not conventional RNA based techniques will be successful (Crompton *et al.,* 1995). The main disadvantage of using a PCR based approach however is, as mentioned above, that it is sequence specific. That is, those transcripts not containing the priming sites will fail to be isolated even if other regions of the transcript are common. It would therefore be better to use whole transcripts as probes in conventional library screens (or in a similar screen to mentioned earlier Section 7.2.2) so any regions of cross-homology will hybridise and therefore the

transcripts may be detected. If the *shak-B* transcripts are non-polyadenylated however then these approaches are very difficult if not impossible as the lack of a 3′ poly A tract will mean the transcripts (or at least their 3′ ends) are absent from poly A^+ enriched libraries.

If this is the case (transcripts from *shak-B* are not polyadenylated) then it is also the case that they are too rare to be detected with upto 100μg of total RNA. If it is non-polyadenylated then this would be potentially very exciting as very few loci with poly A^- transcripts have been identified to date, even though over a third of *Drosophila* transcripts have been estimated to be poly A^- (Zimmerman *et al.*, 1982). The reasons very few unpolyadenylated transcripts have been identified to date is probably because very few people have looked for them. Also most of the screening techniques will exclude them *e.g.* screening of poly A^+ libraries *etc.* Those genes from which poly A^- transcripts have been identified, tend to be because the gene in question is extremely well characterised and an identified transcript *e.g.* identified from genomic sequence analysis, is shown to be missing from the poly A^+ fraction of the RNA. This was the case for one of the *Bithorax* transcripts. It is possible to perform RT PCR on total RNA which will contain both poly A^+ and poly A^- transcripts amongst the tRNA and rRNA, however, the same problems as mentioned above will be encountered, that of selecting only those transcripts from the locus with the priming sequence. Conventional library screening of poly A^- cDNA libraries (or direct selection of transcripts using poly A^- cDNAs) would be the best option, however, it is therefore necessary to obtain poly A^- RNA. One way of obtaining this is to isolate cytoplasmic mRNA which is composed of poly A^- or poly A^+ RNA (Berger and Kimmel, 1987; Mechler, 1987).

Proteins of eukaryotes are synthesised on polysomes that are either free in the cytoplasm or associated with the endoplasmic reticulum. Soluble cytosolic proteins are synthesised on free polysomes whereas secreted or integral membrane proteins are synthesised on membrane bound ribosomes. In most cells the membrane bound ribosomes represent 5 -15% of the total cell ribosomal population. To isolate microsomes RNases are firstly removed or inactivated and then an isopycnic centrifugation in a discontinuous sucrose gradient is performed. The discontinuous gradient allows the membrane vesicles to float whereas the free ribosomal particles

are partially sedimented. It is therefore possible to isolate one, or both, of the types of polysomes by removing either one, or both, fractions. The microsomal fraction contains membrane bound ribosomes and some 80S and 60S subunits as well as mitochondrial ribosomes. The free ribosomal fraction contains the free polysomes, the inactive 80S ribosomes and the recycling ribosomal subunits as well as untranslated mRNP particles. It is therefore possible to extract RNA from these polysomal fractions and this RNA will contain poly A^- and poly A^+ transcripts. Reverse transcription using random hexamers will allow the RNA to be converted to cDNA and hence cloned to make a microsomal cDNA library. This cDNA library will be enriched for all mRNAs (*i.e.* poly A^- and poly A^+) and so if *shak-B* is non-polyadenylated, it should be possible to isolate poly A^- transcripts from the locus, by screening this library.

7.3 The *opa* Repeats.

As shown in Chapter 4 (Section 4.2.5) a large number of the repetitive transcribed fragments were shown to contain *opa* repeats. *Opa* repeats are known to occur in a large number of developmentally regulated genes in *Drosophila* many of which have important functions (see Section 1.4.5.1) for examples and references). One problem that may be encountered using direct selection is that although cDNAs containing *opa* repeats can be isolated it may be difficult to ascertain whether the clone is a spurious contaminant or was selected by hybridisation to a stretch of single copy sequence (*e.g.* Figure 5.09). Bearing in mind that a large number of the transcribed regions identified by this study, contained *opa* repeats (Section 4.2.5). it would therefore probably be worthwhile to isolate flanking sequences outwith the *opa* repeat sequences or block the repetitive sequence in some way prior to hybridising these cDNAs to Northern. Blocking of the *opa* repeat can be achieved by preparing single stranded DNA containing the *opa* repeat sequence which when used in excess in a hybridisation, will competitively bind to and so block hybridisation to other *opa* repeats. This approach is used routinely in human cDNA screens to block *Alu* sequences, many of which are contained in other genes as well as being transcribed themselves (*e.g.* Lovett *et al.*, 1991; Parimoo *et al.*, 1991).

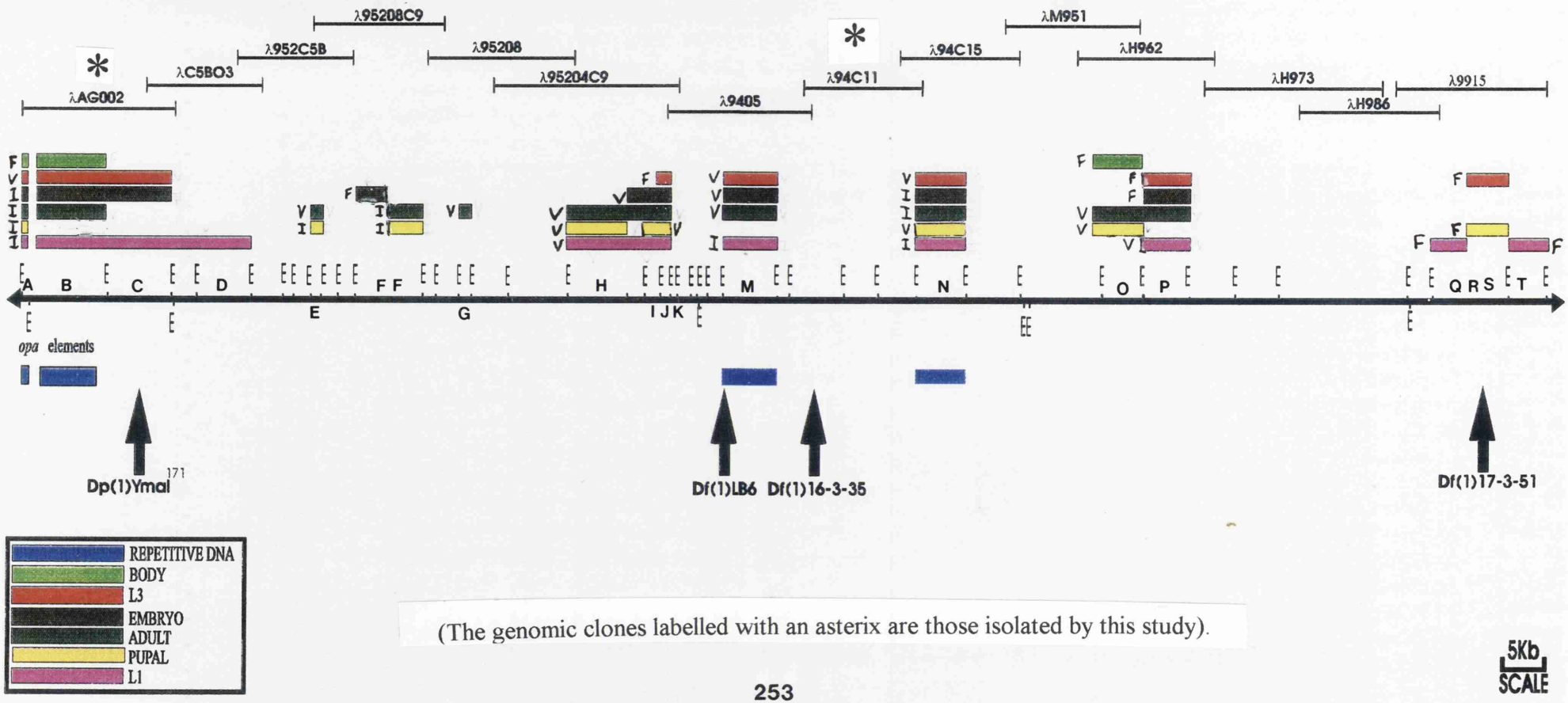
This approach would be useful for allowing the identification of transcribed *opa* repeats for further analysis from regions of interest to this study; it would however

also allow the number of transcribed versus non-transcribed *opa* repeats to be calculated, which has not been done for any large genomic region to date.

APPENDIX 1.

Schematic diagram of the 952 walks transcribed and repetitive regions and other results referred to in this study. Overlapping lambda clones are shown at the top of the diagram, only *EcoR* I sites are displayed on the map of the region.

Each band is labelled with the amount of signal observed on the autoradiograph: (F) = Faint band; (V) = Visible band; (I) = Intense band. It is therefore assumed that the more intense bands represent more common transcripts.

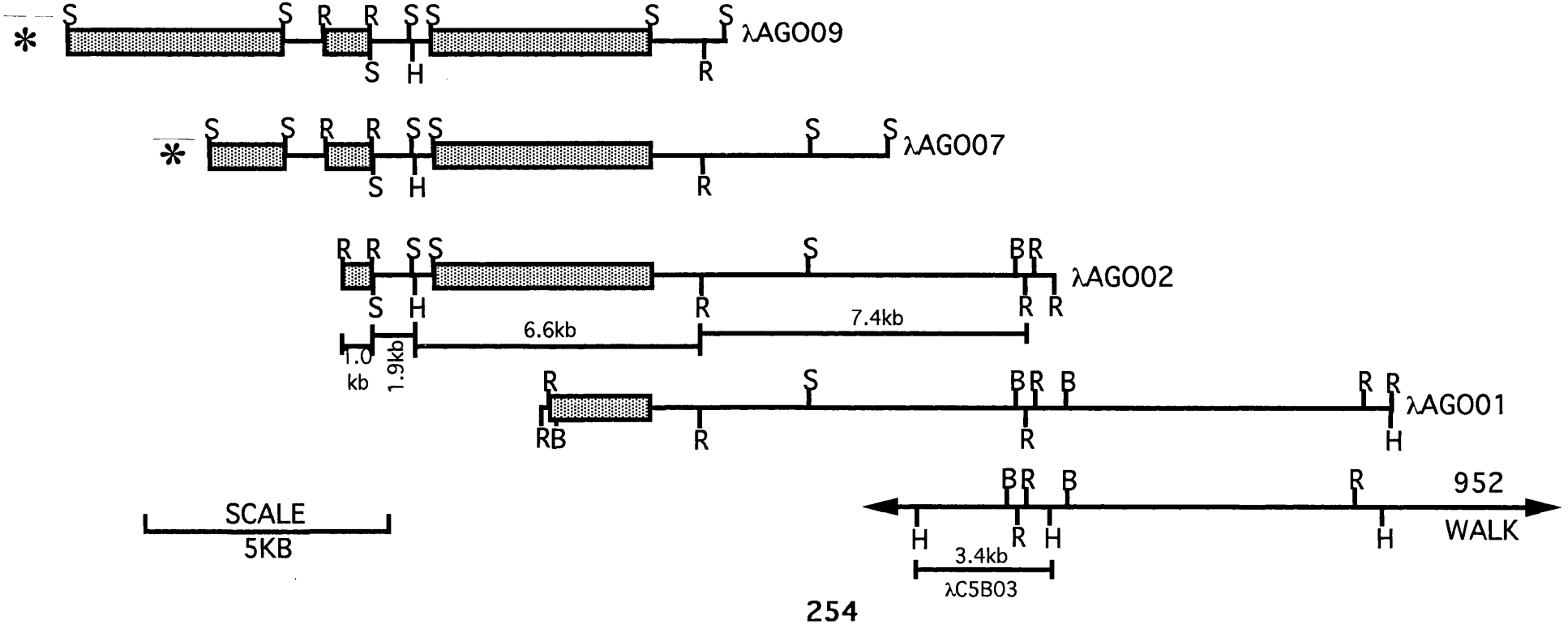


(The genomic clones labelled with an asterix are those isolated by this study).

Appendix 1.1

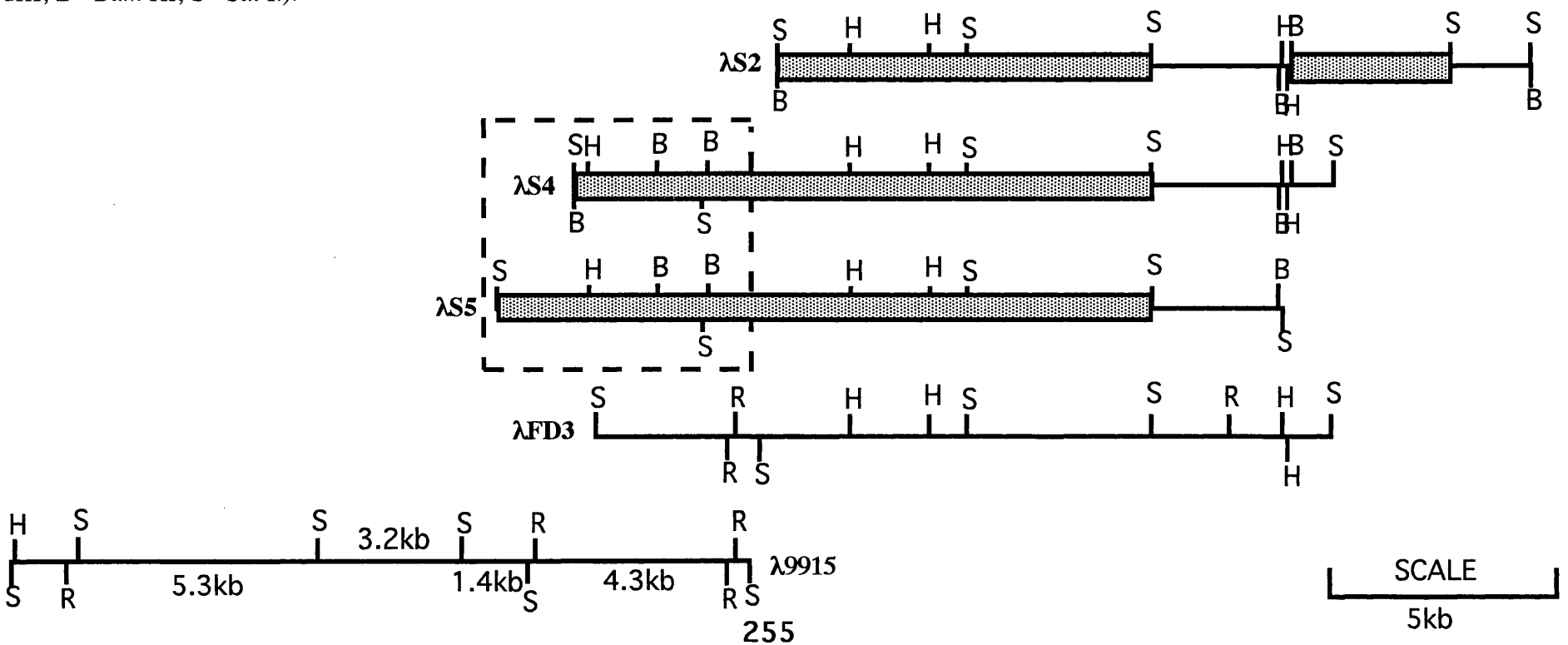
Maps of the genomic lambda clones isolated during the course of this study from the distal end of the 952 walk (at 19E3). See Section 3.2.3 for more details. The fragments labelled with their sizes were used as probes for parts of this study and are described in the text of Chapter Three. Those regions marked as ▒▒ contain repetitive sequences. (R=*Eco* RI; H= *Hin* dIII; B= *Bam* HI; S= *Sac* I.).

(The genomic clones labelled with an asterix are those isolated by this study).



Appendix 1.2.

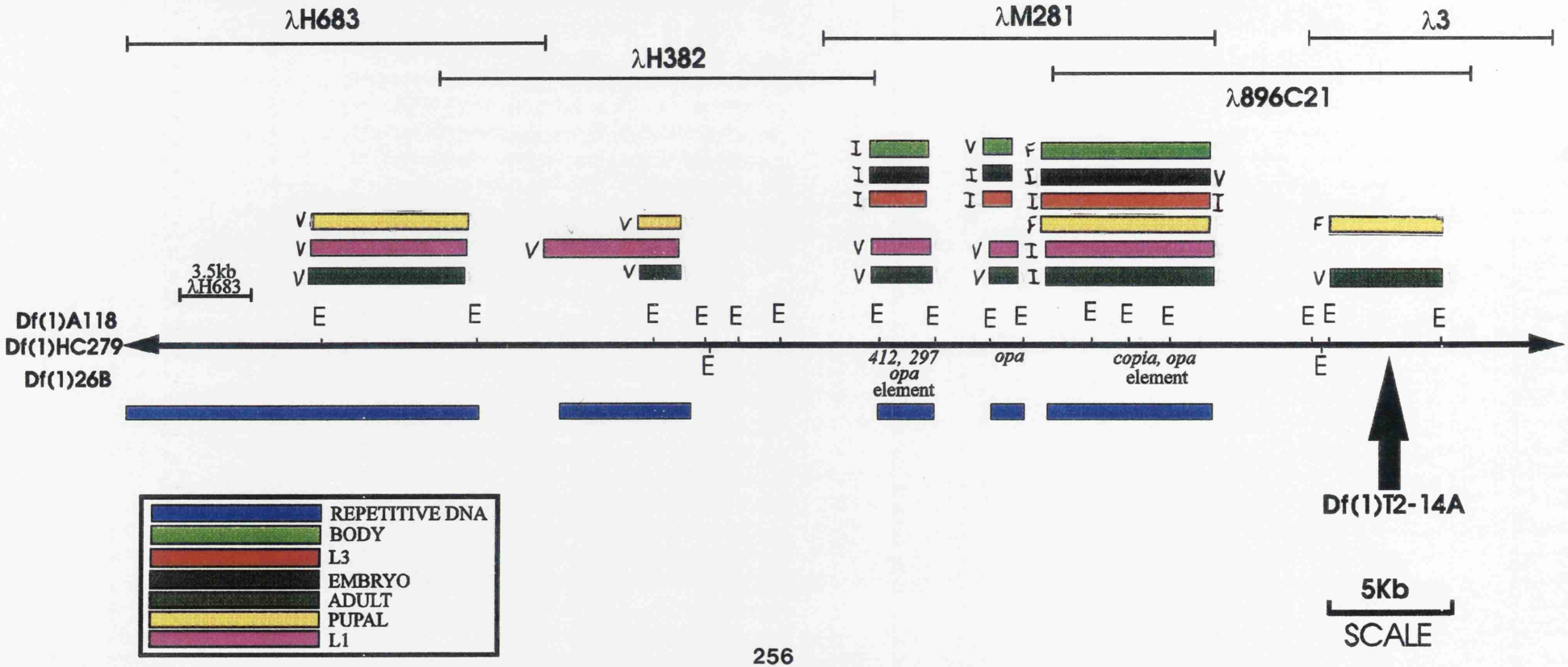
Schematic diagram of the genomic lambda clones from the proximal end of the 952 walk. Those regions marked as ▒ contain repetitive sequences. The region in the hatched box is not contained in the 952 walk and maybe a transposable element. See text (Section 3.2.2 and 3.11.1 for more details. The areas labelled with their sizes have been used as probes throughout this study and are referred to in the text. (R=*Eco* RI; H= *Hin* dIII; B= *Bam* HI; S= *Sal* I.).



APPENDIX 2.

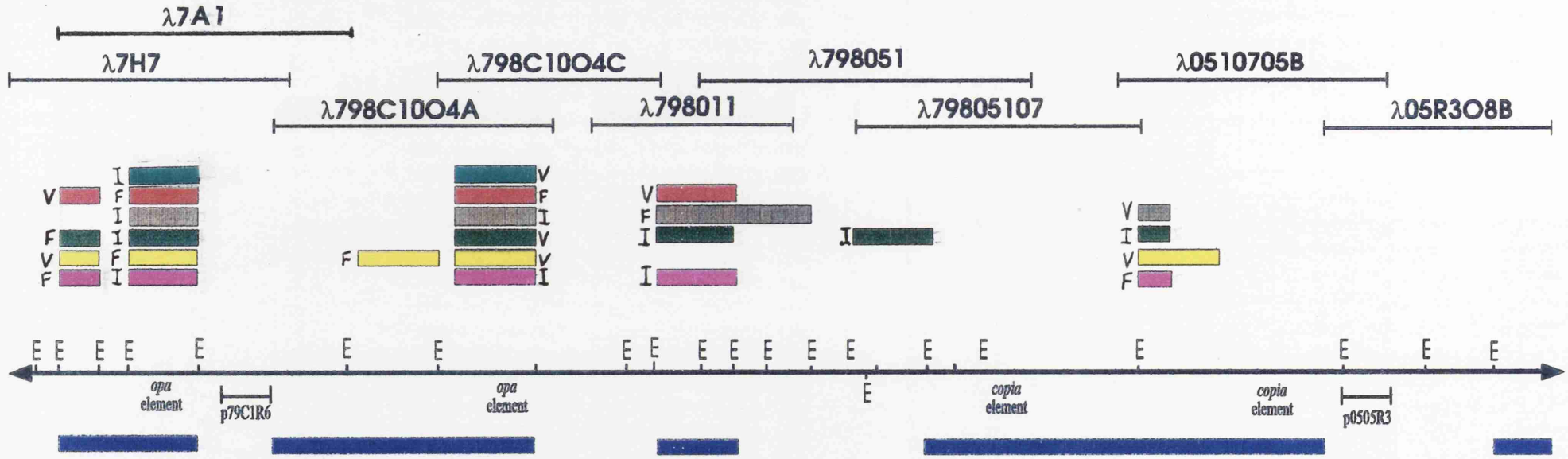
Schematic diagram of the 896 walks transcribed and repetitive regions and other results referred to in this study. Overlapping lambda clones are shown at the top of the diagram, only *EcoR* I sites are displayed on the map of the region.

Each band is labelled with the amount of signal observed on the autoradiograph: (F) = Faint band; (V) = Visible band; (I) = Intense band. It is therefore assumed that the more intense bands represent more common transcripts.

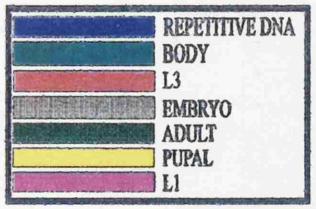


APPENDIX 3.

Schematic diagram of the 798 walks transcribed and repetitive regions and other results referred to in this study. Overlapping lambda clones are shown at the top of the diagram, only *EcoR* I sites are displayed on the map of the region.



Each band is labelled with the amount of signal observed on the autoradiograph: (F) = Faint band; (V) = Visible band; (I) = Intense band. It is therefore assumed that the more intense bands represent more common transcripts.

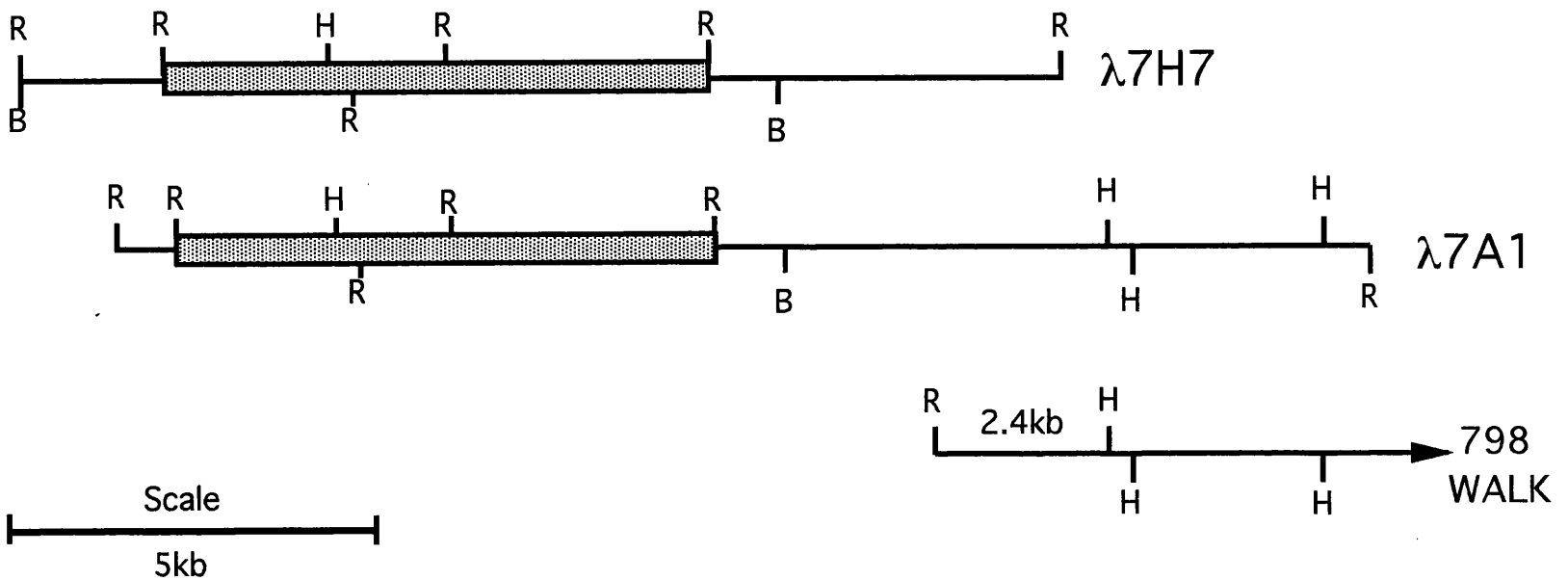


(The genomic clones labelled with an asterix are those isolated by this study).

5Kb
SCALE

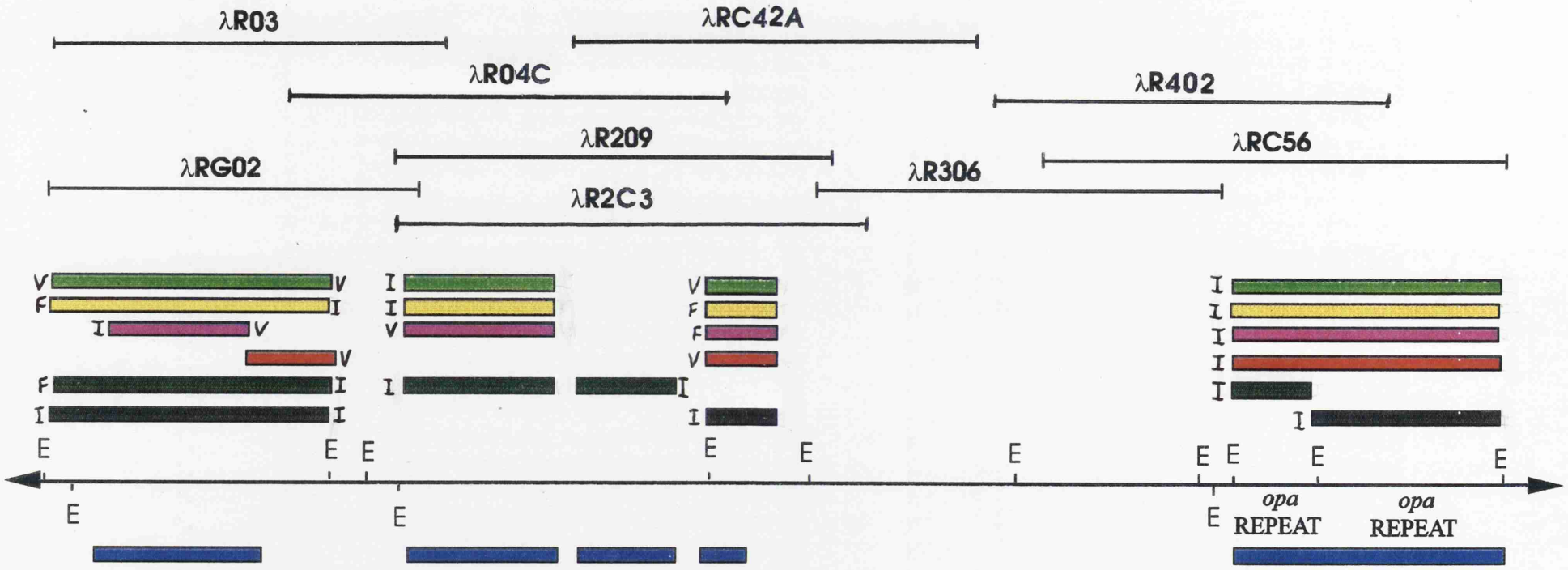
Appendix 3.1

Maps of the genomic lambda clones isolated during the course of this study from the distal end of the 798 walk (at 19E1). See Section 3.3 and 3.11.3 for more details. The restriction fragment labeled 2.4kb was used as a probe to isolate the lambdas. Those regions marked as ▒ contain repetitive sequences. (R=*Eco* RI; H=*Hin* dIII; B=*Bam* HI).



APPENDIX 4.

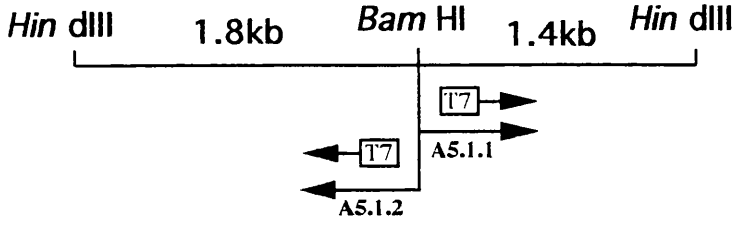
Schematic diagram of the transcribed and repetitive regions from the Runt walk and other results referred to in this study. Overlapping lambda clones are shown at the top of the diagram, only *EcoR* I sites are displayed on the map of the region.



Each band is labelled with the amount of signal observed on the autoradiograph: (F) = Faint band; (V) = Visible band; (I) = Intense band. It is therefore assumed that the more intense bands represent more common transcripts.

Appendix Five

A5.1. Sequencing rationale for the repetitive region adjacent to the *Bam* HI site within the 3.2kb *Hin* dIII-*Hin* dIII fragment from λR209. See Section 4.2.4 and 4.3.4 for details.



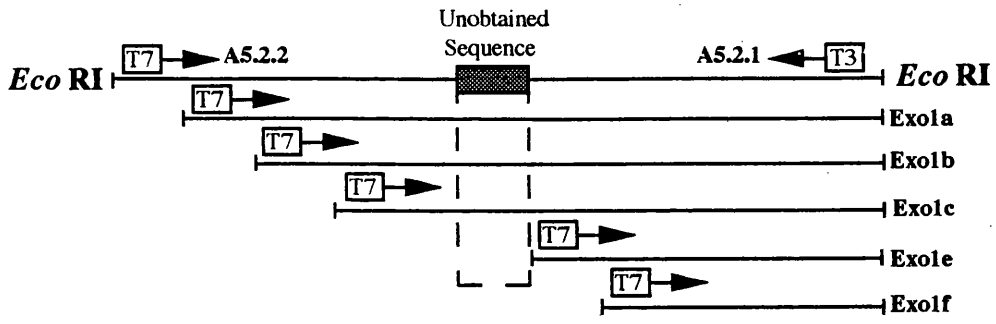
A5.1.1 Sequence and open reading frames found from the T7 (*Bam* HI) end of the 1.4kb *Hin* dIII-*Bam* HI fragment from λR209. There are no O.R.F.s that extend throughout the length of the sequence, although the one that continues the longest is shown. The frame the O.R.F. originates from is indicated down the left hand side of the page.

```
         CCAAGCAATTCATACTGGAAAATTTTCGATTTATGAGAGTGTATAAAAAATTGAAAAAAC
       1 ---------+---------+---------+---------+---------+---------+ 60
         GGTTCGTTAAGTATGACCTTTTAAAAGCTAAATACTCTCACATATTTTTTAACTTTTTTG
FRAME 3    K  Q  F  I  L  E  N  F  R  F  M  R  V  Y  K  K  L  K  K  R

         GTAGAGCCTGGCAAGACTTTGCCATTTGTTTGTATTTTGTTTTTTTTTTTTATGTGCTCA
      61 ---------+---------+---------+---------+---------+---------+ 120
         CATCTCGGACCGTTCTGAAACGGTAAACAAACATAAAACAAAAAAAAAAAAATACACGAGT
FRAME 3    R  A  W  Q  D  F  A  I  C  L  Y  F  V  F  F  F  Y  V  L  S

         GTGCCCCTTTACTGATTTTCCTGTTGTGCAATTTACTTTCATCAGCCGTCTGTCTTGAAA
     121 ---------+---------+---------+---------+---------+---------+ 180
         CACGGGGAAATGACTAAAAGGACAACACGTTAAATGAAAGTAGTCGGCAGACAGAACTTT
FRAME 3    A  P  L  L  I  F  L  L  C  N  L  L  S  S  A  V  C  L  E  N

         ATATCAAAACAGTGTAAGGGAAA
     181 ---------+---------+--- 203
         TATAGTTTTGTCACATTCCCTTT
FRAME 3    I  K  T  V  *
```

A5.1.2 Sequence and open reading frames found from the T7 (*Bam* HI) end of the 1.8kb *Hin* dIII-*Bam* HI fragment from λR209. There are no O.R.F.s that extend throughout the length of the sequence.

```
         CCACCGTACACCAAGAGAAATTTCAAATATATTTTAAGTTTTTAAGTTGAGAGTCTCTTG
       1 ---------+---------+---------+---------+---------+---------+ 60
         GGTGGCATGTGGTTCTCTTTAAAGTTTATATAAAATTCAAAAATTCAACTCTCAGAGAAC

         AATGCCACTTAAGAAATGCGAATAACTGAGGTGTAGAATATAAAGTAATCCATATTAAAA
      61 ---------+---------+---------+---------+---------+---------+ 120
         TTACGGTGAATTCTTTACGCTTATTGACTCCACATCTTATATTTCATTAGGTATAATTTT

         ATCTATTCTAATAAATGCAATAGTGTTTTAAGTAAGTGAATTTACAACTTCGACGTATAG
     121 ---------+---------+---------+---------+---------+---------+ 180
         TAGATAAGATTATTTACGTTATCACAAAATTCATTCACTTAAATGTTGAAGCTGCATATC

         AGCATTGCCCATTTATAATTCTTGAATGAAGTCCTGATTTTTGTATTTTACGTTAAATAC
     181 ---------+---------+---------+---------+---------+---------+ 240
         TCGTAACGGGTAAATATTAAGAACTTACTTCAGGACTAAAAACATAAAATGCAATTTATG

         CTTGGCTCAGGCCTATCTTATTTCTCGTCGTGTACTCATCGTGTGAGTTTAATATTTCAA
     241 ---------+---------+---------+---------+---------+---------+ 300
         GAACCGAGTCCGGATAGAATAAAGAGCAGCACATGAGTAGCACACTCAAATTATAAAGTT

         TTAATTCGG
     301 --------- 309
         AATTAAGCC
```

A5. 2 Sequencing rationale for the 1.0kb *Eco* RI fragment from λAGOO2.



Due to 'chewing back' of the *exo* III nuclease one of the *exo* III deletions could not be sequenced. The region this unobtained sequence lies in is shown above.

A5. 2.1 Sequence and open reading frames found from the T3 (distal) end of the 1.0kb *Eco* RI fragment from λAGOO2. There are no O.R.F.s that extend throughout the length of the sequence, although the longest (Frame 4) is shown as it starts before, and extends through, the *opa* repeat, (the sequence of which is shown in bold text). See Sections 1.4.5.1, 4.3.4 and 4.3.4.1 for more details.

```
        AAACGGTGGGAATGGACTGTAATTGTGCCGCTTAAAAGGCAGGTGATTCGGCAACAAGTG
      1 ---------+---------+---------+---------+---------+---------+ 60
        TTTGCCACCCTTACCTGACATTAACACGGCGAATTTTCCGTCCACTAAGCCGTTGTTCAC
FRAME 4                                  *  F  A  P  S  E  A  V  L  -

        AGACTACCGCATAACCCACAGCGACCACAAAGGCCACACGGTGCGGCATTAATAAATTTG
     61 ---------+---------+---------+---------+---------+---------+ 120
        TCTGATGGCGTATTGGGTGTCGCTGGTGTTTCCGGTGTGCCACGCCGTAATTATTTAAAC
FRAME 4 S  V  V  A  Y  G  V  A  V  V  F  A  V  R  H  P  M  L  L  N  -

        CCTCTGGCAAAAAGCTGAAAAATATTGAAAAAGCGATCTCTACACAATTGTCCATAAAGT
    121 ---------+---------+---------+---------+---------+---------+ 180
        GGAGACCGTTTTTCGACTTTTTATAACTTTTTCGCTAGAGATGTGTTAACAGGTATTTCA
FRAME 4 A  E  P  L  F  S  F  F  I  S  F  A  I  E  V  C  N  D  M  F  -

        GGACACTTTTTGCATTACTTTTCATTTTATTTTTGCCTTTGCAATGCTGTTGCTGTTGCT
    181 ---------+---------+---------+---------+---------+---------+ 240
        CCTGTGAAAAACGTAATGAAAAGTAAAATAAAAACGGAAACGTTACGACAACGACAACGA
FRAME 4 H  V  S  K  A  N  S  K  M  K  N  K  G  K  C  H  Q  Q  Q  Q  -

        GTTGCTGTTGCTTCTGCTGTTGCTGCTGCTGCTGCTGCTGCAGTTGCAGTTGTATGC
    241 ---------+---------+---------+---------+---------+---------+ 300
        CAACGACAACGAAGACGACAACGACGACGACGACGACGACGACGTCAACGTCAACATACG
FRAME 4 Q  Q  Q  Q  K  Q  Q  Q  Q  Q  Q  Q  Q  Q  Q  L  Q  L  Q  I  -

        TCCGGGGTGTCCTTATTTTTGCCGGGCAACCACAGCGAGGATTAAAGCGGACAGTAAAGC
    301 ---------+---------+---------+---------+---------+---------+ 360
        AGGCCCCACAGGAATAAAAACGGCCCGTTGGTGTCGCTCCTAATTTCGCCTGTCATTTCG
FRAME 4 S  R  P  T  R  I  K  A  P  C  G  C  R  P  N  F  R  V  T  F  -

        GGACCGNTGCCATTTGTTGTGACAATGTGGCTAACAAAAGCGAACAGCAACACATGACTC
    361 ---------+---------+---------+---------+---------+---------+ 420
        CCTGGCNACGGTAAACAACACTGTTACACCGATTGTTTTCGCTTGTCGTTGTGTACTGAG
FRAME 4 R  V  ?  A  M  Q  Q  S  L  T  A  L  L  L  S  C  C  C  M  V  -

        GTTGATATGACAACTTTTT
    421 ---------+--------- 439
        CAACTATACTGTTGAAAAA
FRAME 4 R  Q  Y  S  L  K  K  -
```

A5.2.2 Sequence obtained from the T7 (proximal) end of the 1.0kb *Eco* RI fragment from λAGO02. There are no O.R.F.s that extend throughout the length of the sequence, although the longest (Frame 6) is shown as it extends almost all of the way through, the sequence.

```
          CTTGGGAACATCAATCAAAGATATCTCTTTTAAAATATATATTTAAAATTATAAATTAAA
        1 ---------+---------+---------+---------+---------+---------+ 60
          GAACCCTTGTAGTTAGTTTCTATAGAGAAAATTTTATATATAAATTTTAATATTTAATTT

          AAATAGGTTTTGTAATTAGTATATACTTAACTTCTTAACTATNTGAGCTTNGTTGACAAA
       61 ---------+---------+---------+---------+---------+---------+ 120
          TTTATCCAAAACATTAATCATATATGAATTGAAGAATTGATANACTCGAANCAACTGTTT
FRAME 6                                    *  S  ?  S  S  ?  N  V  F -

          CACTTTTTGTTGCCCAAAAAAAAAAAAAGGGCAACCAAATATTTGTGGAGCCGAAGACCAG
      121 ---------+---------+---------+---------+---------+---------+ 180
          GTGAAAAACAACGGGTTTTTTTTTTTTCCCGTTGGTTTATAAACACCTCGGCTTCTGGTC
FRAME 6     V  K  Q  Q  G  F  F  F  F  P  L  W  I  N  T  S  G  F  V  L -

          TTGAATCCTTNCGACAGCAATTTTTCCGAGTCTTTTGTTCATTG.CAGGCGCTTGAGTAT
      181 ---------+---------+---------+---------+---------+---------+ 240
          AACTTAGGAANGCTGTCGTTAAAAAGGCTCAGAAAACAAGTAAC.GTCCGCGAACTCATA
FRAME 6     Q  I  R  ?  V  A  I  K  G  L  R  K  N  M  ?  L  R  K  L  I -

          TCTGATTGAGTGACGGATGAAAGGGTTCGTGATTGAGGTGGGACTTTGATTGGGGCCAGG
      241 ---------+---------+---------+---------+---------+---------+ 300
          AGACTAACTCACTGCCTACTTTCCCAAGCACTAACTCCACCCTGAAACTAACCCCGGTCC
FRAME 6     R  I  S  H  R  I  F  P  N  T  I  S  T  P  S  Q  N  P  G  P -

          GAATTTGGGGGNGGGGCTAGCCGTTGTCC
      301 ---------+---------+--------- 329
          CTTAAACCCCCNCCCCGATCGGCAACAGG
FRAME 6     F  K  P  ?  P  S  A  T  T  -
```

A5.3 Both ends of the 3.2kb *Eco* RI-*Bam*HI fragment from the distal end of λAGO01 were sequenced using T3 and T7 primers.

A5.3.1 Sequence and open reading frames found at the T3 end of the 3.2kb *Eco* RI-*Bam*HI restriction fragment from the proximal end of λAGO01. There are two O.R.Fs that extend throughout the length of the sequence.

```
          AACCAAACCAAACCGAACCGTACGGAACGGAACCACCCGGTTGAAGTGGTTGGTTGTACG
        1 ---------+---------+---------+---------+---------+---------+ 60
          TTGGTTTGGTTTGGCTTGGCATGCCTTGCCTTGGTGGGCCAACTTCACCAACCAACATGC
FRAME1    N  Q  T  K  P  N  R  T  E  R  N  H  P  V  E  V  V  G  C  T  -
FRAME2     T  K  P  N  R  T  V  R  N  G  T  T  R  L  K  W  L  V  V  R -

          ATTTGGATTTGGAATTTGCGATTTCAATGCCACTCAACTGGCCGCTGGGAAAATCGCCTT
       61 ---------+---------+---------+---------+---------+---------+ 120
          TAAACCTAAACCTTAAACGCTAAAGTTACGGTGAGTTGACCGGCGACCCTTTTAGCGGAA
FRAME1    I  W  I  W  N  L  R  F  Q  C  H  S  T  G  R  W  E  N  R  L  -
FRAME2     F  G  F  G  I  C  D  F  N  A  T  Q  L  A  A  G  K  I  A  L -

          AAAAAGTCCTTTACCCCGCATCCGCGCATCCCTATCCGCAAATGTATCTGTAGCTGTAGT
      121 ---------+---------+---------+---------+---------+---------+ 180
          TTTTTCAGGAAATGGGGCGTAGGCGCGTAGGGATAGGCGTTTACATAGACATCGACATCA
FRAME1    K  K  S  F  T  P  H  P  R  I  P  I  R  K  C  I  C  S  C  S  -
FRAME2     K  S  P  L  P  R  I  R  A  S  L  S  A  N  V  S  V  A  V  V -

          TGTAGCTGTAGCTGTATCTGTATCGTATCTGTATGCGCGCAGAGATACATTGTATATGGC
      181 ---------+---------+---------+---------+---------+---------+ 240
          ACATCGACATCGACATAGACATAGCATAGACATACGCGCGTCTCTATGTAACATATACCG
FRAME1    C  S  C  S  C  I  C  I  V  S  V  C  A  Q  R  Y  I  V  Y  G  -
FRAME2     V  A  V  A  V  S  V  S  Y  L  Y  A  R  R  D  T  L  Y  M  A -
```

```
          AGGAAGAAGGTCGAGCATGCCTCGTGGTGCT
      241 ---------+---------+---------+- 271
          TCCTTCTTCCAGCTCGTACGGAGCACCACGA
FRAME1    R  K  K  V  E  H  A  S  W  C    -
FRAME2     G  R  R  S  S  M  P  R  G  A   -
```

A5.3.2 Sequence and open reading frames found at the T7 (distal) end of the 3.2kb *Eco* RI-*Bam* HI restriction fragment from the distal end of λAGO01. The *opa* repeat sequence is shown in bold text. (**B**= *Bam* HI).

```
            B
          CCCCGGATCCACCTTTGCTTTTTTTTTTGTCCCCTGTTGTTTTTGGCCAGGCTTTTTTCC
        1 ---------+---------+---------+---------+---------+---------+ 60
          GGGGCCTAGGTGGAAACGAAAAAAAAAACAGGGGACAACAAAAACCGGTCCGAAAAAAGG
FRAME 2    P  G  S  T  F  A  F  F  F  V  P  C  C  F  W  P  G  F  F  P  -
FRAME 3      P  D  P  P  L  L  F  F  L  S  P  V  V  F  G  Q  A  F  F  Q -
FRAME 4       G  S  G  G  K  S  K  K  K  D  G  T  T  K  P  W  A  K  K   -

          AGGGCGGTTGGCTGACAGTCACGGCGGCACGTGTCACTTGGACAGTTGTACGAGCGGCAC
       61 ---------+---------+---------+---------+---------+---------+ 120
          TCCCGCCAACCGACTGTCAGTGCCGCCGTGCACAGTGAACCTGTCAACATGCTCGCCGTG
FRAME 2    G  R  L  A  D  S  H  G  G  T  C  H  L  D  S  C  T  S  G  T  -
FRAME 3      G  G  W  L  T  V  T  A  A  R  V  T  W  T  V  V  R  A  A  R -
FRAME 4    W  P  P  Q  S  V  T  V  A  A  R  T  V  Q  V  T  T  R  A  A   -

          GAGCAACAGCAGCAGCAACAACAACAGCAACAATGAACAGCAACAGCAGCAGAAGCATCA
      121 ---------+---------+---------+---------+---------+---------+ 180
          CTCGTTGTCGTCGTCGTTGTTGTTGTCGTTGTTACTTGTCGTTGTCGTCGTCTTCGTAGT
FRAME 2    S  N  S  S  S  N  N  N  S  N  N  E  Q  Q  Q  Q  Q  K  H  Q  -
FRAME 3      A  T  A  A  A  T  T  T  A  T  M  N  S  N  S  S  R  S  I  N -
FRAME 4    R  A  V  A  A  A  V  V  V  A  V  I  F  L  L  L  L  L  M   -

          ACTGCACGGCAGCAACATCAACGGAGTGAAGCGAGAAACACTAGAGAACAAAGCGT
      181 ---------+---------+---------+---------+---------+------ 236
          TGACGTGCCGTCGTTGTAGTTGCCTCACTTCGCTCTTTGTGATCTCTTGTTTCGCA
FRAME 2    L  H  G  S  N  I  N  G  V  K  R  E  T  L  E  N  K  A    -
FRAME 3      C  T  A  A  T  S  T  E  *                              -
FRAME 4
```

A5.4 Both ends of the 0.8kb *Eco* RI-*Bam*HI fragment from λ95208 were sequenced using T3 and T7 primers.

A5.4.1 Sequence and O.R.Fs obtained from the T3 end of the 0.8kb *Eco* RI-*Bam*HI fragment from λ95208. There are three O.R.Fs that extend throughout the length of the sequence.

```
          AGTCCGGGCAGGCGCGAGAGCCTCAAGATGCCCACGATTTCATCGAACACGTTCGCGAGG
        1 ---------+---------+---------+---------+---------+---------+ 60
          TCAGGCCCGTCCGCGCTCTCGGAGTTCTACGGGTGCTAAAGTAGCTTGTGCAAGCGCTCC
FRAME 1    S  P  G  R  R  E  S  L  K  M  P  T  I  S  S  N  T  F  A  R   -
FRAME 3      S  G  Q  A  R  E  P  Q  D  A  H  D  F  I  E  H  V  R  E  G -
FRAME 5       G  P  L  R  S  L  R  L  I  G  V  I  E  D  F  V  N  A  L  -

          GTGAAATCGAAGGAGTTTACTGGACGGGCGAGGGCAAACGGACAGCCACCCAAACGCCAC
       61 ---------+---------+---------+---------+---------+---------+ 120
          CACTTTAGCTTCCTCAAATGACCTGCCCGCTCCCGTTTGCCTGTCGGTGGGTTTGCGGTG
FRAME 1    V  K  S  K  E  F  T  G  R  A  R  A  N  G  Q  P  P  K  R  H   -
FRAME 3      E  I  E  G  V  Y  W  T  G  E  G  K  R  T  A  T  Q  T  P  Q -
FRAME 5      T  F  D  F  S  N  V  P  R  A  L  A  F  P  C  G  G  L  R  W  -

          AGCGTAAGACGGGGCATTCGGGATTCATCGATGGCCATGGCGAAGAGATTCCCCATGAGG
      121 ---------+---------+---------+---------+---------+---------+ 180
          TCGCATTCTGCCCCGTAAGCCCTAAGTAGCTACCGGTACCGCTTCTCTAAGGGGTACTCC
FRAME 1    S  V  R  R  G  I  R  D  S  S  M  A  M  A  K  R  F  P  M  R   -
FRAME 3      R  K  T  G  H  S  G  F  I  D  G  H  G  E  E  I  P  H  E  D -
FRAME 5      L  T  L  R  P  M  R  S  E  D  I  A  M  A  F  L  N  G  M  L -
```

```
          ACTCCAAACAAGAGGTGGCTCCCTATCACCCTGATACCTCGGATCCCAGTTCGCAGGTCC
      181 ---------+---------+---------+---------+---------+---------+ 240
          TGAGGTTTGTTCTCCACCGAGGGATAGTGGGACTATGGAGCCTAGGGTCAAGCGTCCAGG
FRAME 1    T  P  N  K  R  W  L  P  I  T  L  I  P  R  I  P  V  R  R  S  -
FRAME 3     S  K  Q  E  V  A  P  Y  H  P  D  P  S  D  P  S  S  Q  V  Q -
FRAME 5    V  G  F  L  L  H  S  G  I  V  R  I  G  R  I  G  T  R  L  D  -

          AACTCCAACTTTAATTTT
      241 ---------+--------- 258
          TTGAGGTTGAAATTAAAA
FRAME 1    N  S  N  F  N  F   -
FRAME 3     L  Q  L  *        -
FRAME 5    L  E  L  K  L  K  -
```

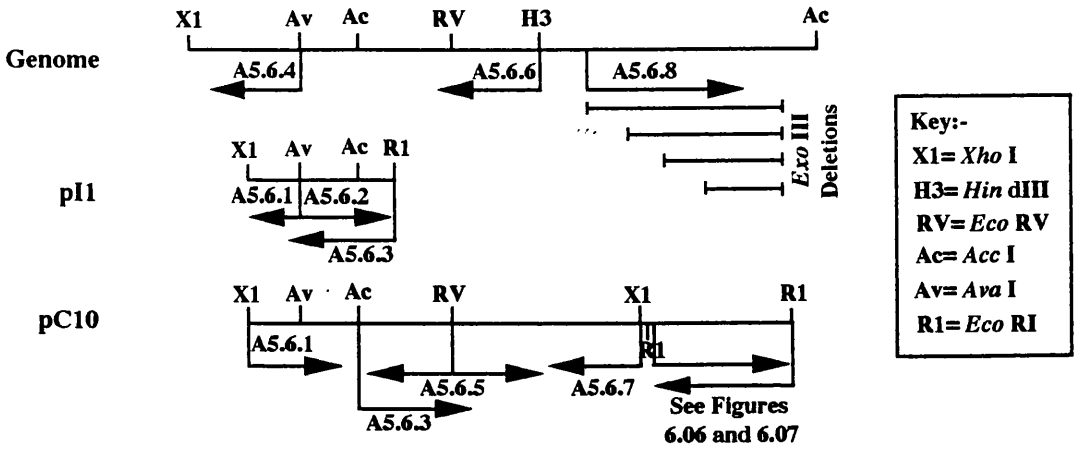
A5.4.2 Sequence obtained from the T7 end of the 0.8kb *Eco* RI-*Bam*HI fragment from λ95208. There are no O.R.F.s that extend throughout the length of the sequence.

```
          CGCCAAAAGCCAAAAAAAGGATTAGGGCTTGTTAAAAAAGGATTTTAGAGGACTGCAAAA
        1 ---------+---------+---------+---------+---------+---------+ 60
          GCGGTTTTCGGTTTTTTTCCTAATCCCGAACAATTTTTTCCTAAAATCTCCTGACGTTTT

          AAGAGTGAAAGTTGCCCACCGNCCGCGTGTGGGAAGCGTGTGTTGTGACCGCCCAGGTAA
       61 ---------+---------+---------+---------+---------+---------+ 120
          TTCTCACTTTCAACGGGTGGCNGGCGCACACCCTTCGCACACAACACTGGCGGGTCCATT

          ATCGTCCTTGCACTCCTGGTCATTGCATAAATTTGCTAAGCGG
      121 ---------+---------+---------+---------+--- 163
          TAGCAGGAACGTGAGGACCAGTAACGTATTTAAACGATTCGCC
```

A5.5 Sequence obtained from the T7 end of the 1.0kb *Eco* RI fragment from λ95208. Only the T7 end was sequenced. There are no O.R.F.s that extend throughout the length of the sequence, although the longest (Frame 3) is shown as it has homology to an uncharacterised sequence in the database, see Section 6.3.2 for details. (RV=*Eco* RV restriction site).

```
          TTCACAGGTAAACACGCGCCGGGGGAGGTGTAAAAGTCAAAACAATGACAAGAAAATTAG
        1 ---------+---------+---------+---------+---------+---------+ 60
          AAGTGTCCATTTGTGCGCGGCCCCCTCCACATTTTCAGTTTTGTTACTGTTCTTTTAATC
FRAME 3     H  R  *                                    M  T  R  K  L  A -

          CTATAATTGTGGCCCCCGCCCCGGTGCGGGGGGGAAGAAGCGACTTACACACGAACCAAA
       61 ---------+---------+---------+---------+---------+---------+ 120
          GATATTAACACCGGGGGCGGGGCCACGCCCCCCCTTCTTCGCTGAATGTGTGCTTGGTTT
FRAME 3     I  I  V  A  P  A  P  V  R  G  G  R  S  D  L  H  T  N  Q  T -

          CAAAGAGAAATATATATCAGAACAACTCAGATCATTGGGGGGAAATCCGACTTCAGATCG
      121 ---------+---------+---------+---------+---------+---------+ 180
          GTTTCTCTTTATATATAGTCTTGTTGAGTCTAGTAACCCCCCTTTAGGCTGAAGTCTAGC
FRAME 3       K  R  N  I  Y  Q  N  N  S  D  H  W  G  E  I  R  L  Q  I  G -

                    RV
          GTTGGCAAATCGATATCT
      181 ---------+--------- 198
          CAACCGTTTAGCTATAGA
FRAME 3       W  Q  I  D  I    -
```

A5.6 Sequencing rationale for the cDNA clones pI1 and pC10 and the genomic region to which they hybridise. See Sections 6.2.2, 6.2.3, 6.5.1.1 and 6.5.3 for more details.



A5.6.1 Sequence obtained from the *Xho* I to *Ava* I subclone from the 3' end of the cDNA clones pI1 and pC10. There are no O.R.F.s that extend throughout the length of the sequence. The poly A^+ tail is shown in bold text.

```
    TTTTTTTTTTTTTTTTTTGTGTTTTTCGTTTTTTTCGTTTCGTATTTTTAATCTTGTCTG
  1 ---------+---------+---------+---------+---------+---------+ 60
    AAAAAAAAAAAAAAAAAACACAAAAAGCAAAAAAAGCAAAGCATAAAAATTAGAACAGAC

    TTTTTGGTTTTTACGTTTGTTGCTTGCTTTAATAATAATTTATAATTATAATAATTTATA
 61 ---------+---------+---------+---------+---------+---------+ 120
    AAAAACCAAAAATGCAAACAACGAACGAAATTATTATTAAATATTAATATTATTAAATAT

    TAATAATATTTTATAATTTTACTTGGTAATTATTAATTCCCATCAACGTCCATGGCCTTT
121 ---------+---------+---------+---------+---------+---------+ 180
    ATTATTATAAAATATTAAAATGAACCATTAATAATTAAGGGTAGTTGCAGGTACCGGAAA

                                          Ava I
    TGTTTTTTGTTGGAAGTAACGAAGGGCGGGGAGTTG
181 ---------+---------+---------+------ 216
    ACAAAAAACAACCTTCATTGCTTCCCGCCCCTCAAC
```

A5.6.2 Sequence obtained from the 3' end of the *Ava* I to *Eco* RI subclone from the cDNA clones pI1. There are no O.R.F.s that extend throughout the length of the sequence. The primer used was the T3 primer. (A=*Acc* I site) .

```
    TTTCTTCCGACTGGGCTCTTTTGCTTTGCTTTTTCATTTGCTTTTGCCTTTGCCTCTGCC
  1 ---------+---------+---------+---------+---------+---------+ 60
    AAAGAAGGCTGACCCGAGAAAACGAAACGAAAAAGTAAACGAAAACGGAAACGGAGACGG

    TTTGTTTGTCTTTGCCTTTACTTTACTTTTAATGCATATGATTTTTGGTTTTATTTCTTC
 61 ---------+---------+---------+---------+---------+---------+ 120
    AAACAAACAGAAACGGAAATGAAATGAAAATTACGTATACTAAAAACCAAAATAAAGAAG

    GTTCTTCTTCTTTGCCGCGTTTTAGTTGTATTTCATTTTATGTAATGCGCCTGCTGCAAT
121 ---------+---------+---------+---------+---------+---------+ 180
    CAAGAAGAAGAAACGGCGCAAAATCAACATAAAGTAAAATACATTACGCGGACGACGTTA

    AATTTTAGTTTGTAAAATTAACTAATCGTAAAATGTGTGTATGTGGGCATGTGTGAGTGG
181 ---------+---------+---------+---------+---------+---------+ 240
    TTAAAATCAAACATTTTAATTGATTAGCATTTTACACACATACACCCGTACACACTCACC

          A
    GTGTATGTATACTGATGTGACTGTCTGCGGTCTGTAGGTGTGTTGGCTGTGTGTATGCGT
241 ---------+---------+---------+---------+---------+---------+ 300
    CACATACATATGACTACACTGACAGACGCCAGACATCCACACAACCGACACACATACGCA

    GTTATTTACAAGTCGATTTTT
301 ---------+----------+- 321
    CAATAAATGTTCAGCTAAAAA
```

A5.6.3 Sequence obtained from the 5' end of the *Ava* I to *Eco* RI subclone from the cDNA clones pI1. There are no O.R.F.s that extend throughout the length of the sequence. The primer used was the T7 primer. The region in lower case is polylinker sequence. (A=*Acc* I site; R = *Eco* RI site).

```
                              R
    gtggatccccgggctgcagGAATTCGGCACGAGACTAACGGAAACGGAAGTGGAGGCGA
  1 ---------+---------+---------+---------+---------+---------+ 60
    cacctaggggcccgacgtcCTTAAGCCGTGCTCTGTTGCCTTTGCCTTCACCTCCGCT


    GGACACCTTACATTTAAGACACTGAACAAAAAATCGACTTGTAAATAACACGCATACACA
 61 ---------+---------+---------+---------+---------+---------+ 120
    CCTGTGGAATGTAAATTCTGTGACTTGTTTTTTAGCTGAACATTTATTGTGCGTATGTGT

                                              A
    CAGCCAACACACCTACAGACCGCAGACAGTCACATCAGTATACATACACCCACTCACACA
121 ---------+---------+---------+---------+---------+---------+ 180
    GTCGGTTGTGTGGATGTCTGGCGTCTGTCAGTGTAGTCATATGTATGTGGGTGAGTGTGT

    TGCCCACATACACACATTTTACGATTAGTTAATTTTACAAACTAAAATTATTGCAGCAGG
181 ---------+---------+---------+---------+---------+---------+ 240
    ACGGGTGTATGTGTGTAAAATGCTAATCAATTAAAATGTTTGATTTTAATAACGTCGTCC

    CGCATTACATAAAATGAAATACAACTAAAACGCGGCAAAGAAGAAGAACGAAGAAATAAA
241 ---------+---------+---------+---------+---------+---------+ 300
    GCGTAATGTATTTTACTTTATGTTGATTTTGCGCCGTTTCTTCTTCTTGCTTCTTTATTT

    ACCAAAAATCATATGCATTAAAAGTAAAGTAAAGGCAAAGACAAACAAAGGCAGAGGCAA
301 ---------+---------+---------+---------+---------+---------+ 360
    TGGTTTTTAGTATACGTAATTTTCATTTCATTTCCGTTTCTGTTTGTTTCCGTCTCCGTT

    AGGCAAAAGCAAATGAAAAAGCAAAGCAAAAGAGCCCAGTCGGAAGAAAGGGGCGGGTGG
361 ---------+---------+---------+---------+---------+---------+ 420
    TCCGTTTTCGTTTACTTTTTCGTTTCGTTTTCTCGGGTCAGCCTTCTTTCCCCGCCCACC

    TACACACACCCGA
421 ---------+--- 433
    ATGTGTGTGGGCT
```

A5.6.4 Sequence obtained from the end of the 0.6kb *Ava* I to *Xho* I genomic subclone that contains the 3'end of both the pI1 and pC10 cDNA clones. The sequence highlighted bold is where the 3' end of these clones finish (*i.e.* it is where their poly A tails hybridise). The primer used was the T7 primer.

```
    GTTACTTCCAACAAAAAACAAAAGGCCATGGACGTTGATGGGAATTAATAATTACCAAGT
  1 ---------+---------+---------+---------+---------+---------+ 60
    CAATGAAGGTTGTTTTTTGTTTTCCGGTACCTGCAACTACCCTTAATTATTAATGGTTCA

    AAAATTATAAAATATTATTATATAAATTATTATAATTATAAATTATTATTAAAGCAAGCA
 61 ---------+---------+---------+---------+---------+---------+ 120
    TTTTAATATTTTATAATAATATATTTAATAATATTAATATTTAATAATAATTTCGTTCGT

    ACAAACGTAAAAACCAAAAACAGACAAGATTAAAAATAGCAAACGAAAAAAACGAAAAAC
121 ---------+---------+---------+---------+---------+---------+ 180
    TGTTTGCATTTTTGGTTTTTGTCTGTTCTAATTTTTATCGTTTGCTTTTTTTGCTTTTTG

                    Sequence after this point not in clones.
    ACAAAAAAAAAATCAAACAAAACCAAACATAAACCAAAAACACATTATTATAATGCGATA
181 ---------+---------+---------+---------+---------+---------+ 240
    TGTTTTTTTTTTAGTTTGTTTTGGTTTGTATTTGGTTTTTGTGTAATAATATTACGCTAT

    ATTGAGTAAGCAACAAAACGAATTGACAACGTTGTATGATTTTGATCAGAAGTAATAAAA
241 ---------+---------+---------+---------+---------+---------+ 300
    TAACTCATTCGTTGTTTTGCTTAACTGTTGCAACATACTAAAACTAGTCTTCATTATTTT

    CAAGGGG
301 ------- 307
    GTTCCCC
```

A5.6.5 Sequence obtained from the adjacent regions (proximal and distal) to the *Eco* RV site in the pC10 clone.

```
      TAGACACATACAGACAAGGAAACGTCCTCAGTTTATAAATTTTCAAATGAGTGACACGAC
    1 ---------+---------+---------+---------+---------+---------+ 60
      ATCTGTGTATGTCTGTTCCTTTGCAGGAGTCAAATATTTAAAAGTTTACTCACTGTGCTG

      AAACTTGAAACGAAACGAACGAAACGAAACACGTAGGCAGTACGGATTAGGTCTTAAGGC
   61 ---------+---------+---------+---------+---------+---------+ 120
      TTTGAACTTTGCTTTGCTTGCTTTGCTTTGTGCATCCGTCATGCCTAATCCAGAATTCCG-

      CCCATATAAGTAATATACAATTTAAATTATTTAACAATCAATTTATTATTGATGGCAAAC
  121 ---------+---------+---------+---------+---------+---------+ 180
      GGGTATATTCATTATATGTTAAATTTAATAAATTGTTAGTTAAATAATAACTACCGTTTG

      GGAAGAGAAAGCGGAACGATGGCGATAGCGATAGTGATGACGATAACGACAGCGAGAGGG
  181 ---------+---------+---------+---------+---------+---------+ 240
      CCTTCTCTTTCGCCTTGCTACCGCTATCGCTATCACTACTGCTATTGCTGTCGCTCTCCC

                                                              RV
      GAGAGCCAGGCAGTATAGATCGGAGCCGCGAACATATACATGTGATCATATATCCGATAT
  241 ---------+---------+---------+---------+---------+---------+ 300
      CTCTCGGTCCGTCATATCTAGCCTCGGCGCTTGTATATGTACACTAGTATATAGGCTATA

      CCGTCGATATATATCACGTCCGATATATGCGCACGTGCCCCGATCGCCATACATATACAC
  301 ---------+---------+---------+---------+---------+---------+ 360
      GGCAGCTATATATAGTGCAGGCTATATACGCGTGCACGGGGCTAGCGGTATGTATATGTG

      ACACAATTACTATATGCATAATACCAGTGACATACATACATATGTACTACTCGAATCAAT
  361 ---------+---------+---------+---------+---------+---------+ 420
      TGTGTTAATGATATACGTATTATGGTCACTGTATGTATGTATACATGATGAGCTTAGTTA

      CCGATCGGAGGAGGAAATGAGGAGGCAAAACAAATCTATATACAGTACATGGCTAACACT
  421 ---------+---------+---------+---------+---------+---------+ 480
      GGCTAGCCTCCTCCTTTACTCCTCCGTTTTGTTTAGATATATGTCATGTACCGATTGTGA

      TATATAAAGCCCCTATATTGAGGTATAAAACGTGAATTTTTAGCAGACTAAGTAATAAAT
  481 ---------+---------+---------+---------+---------+---------+ 540
      ATATATTTCGGGGATATAACTCCATATTTTGCACTTAAAAATCGTCTGATTCATTATTTA

      TATACACCACAGGCGGTCATATCGACTACACACACACAGGGCGGTT
  541 ---------+---------+---------+------ 586
      ATATGTGGTGTCCGCCAGTATAGCTGATGTGTGTGTGTCCCGCCAA
```

A5.6.6 Sequence obtained from the end of the 3.4kb *Hin* dIII to *Eco* RI genomic subclone nearest to the *Eco* RV site in the pC10 clone. The primer used was the T7 primer.

```
Hin dIII
      AGCTTTAATTTGAAGCTATATAGGTATATATATATACGCAAATATATATATATATATATA
    1 ---------+---------+---------+---------+---------+---------+ 60
      TCGAAATTAAACTTCGATATATCCATATATATATATGCGTTTATATATATATATATATAT

      CCCACACACTAATACACACCCGACGCAATACACACAAATAGACACATACAGACAAGGAAA
   61 ---------+---------+---------+---------+---------+---------+ 120
      GGGTGTGTGATTATGTGTGGGCTGCGTTATGTGTGTTTATCTGTGTATGTCTGTTCCTTT

      CGTCCTCAGTTTATAAATTTTCAAATGAGTGACACGACAAACTTGAAACGAAACGAACGA
  121 ---------+---------+---------+---------+---------+---------+ 180
      GCAGGAGTCAAATATTTAAAAGTTTACTCACTGTGCTGTTTGAACTTTGCTTTGCTTGCT

      AACGAAACACGTAGGCAGTACGGATTAGGTCTTAAGGCCCCATATAAGTAATATACAATT
  181 ---------+---------+---------+---------+---------+---------+ 240
      TTGCTTTGTGCATCCGTCATGCCTAATCCAGAATTCCGGGGTATATTCATTATATGTTAA

      TAAATTATTTAACAATCAATTTATTATTGATGGCAAACGGAAGAGAAAGCGGAACGATGG
  241 ---------+---------+---------+---------+---------+---------+ 300
      ATTTAATAAATTGTTAGTTAAATAATAACTACCGTTTGCCTTCTCTTTCGCCTTGCTACC

      CGATACGATAGTGATGACGATAACGACAGCGAGAGGGGAGAGCCAGGCAGTATAGATCGG
  301 ---------+---------+---------+---------+---------+---------+ 360
      GCTATGCTATCACTACTGCTATTGCTGTCGCTCTCCCCTCTCGGTCCGTCATATCTAGCC
```

```
                                        RV
    AGCCGCGAACATATACATGTGATGATATATCGATATCGTCGATATATATGCACGTCGATA
361 ---------+---------+---------+---------+---------+---------+ 420
    TCGGCGCTTGTATATGTACACTACTATATAGCTATAGCAGCTATATATACGTGCAGCTAT

    TAT
421 --- 423
    ATA
```

A5.6.7 Sequence obtained from the 5' end of the pC10 clone. The T3 primer was used. (x = *Xho* I).

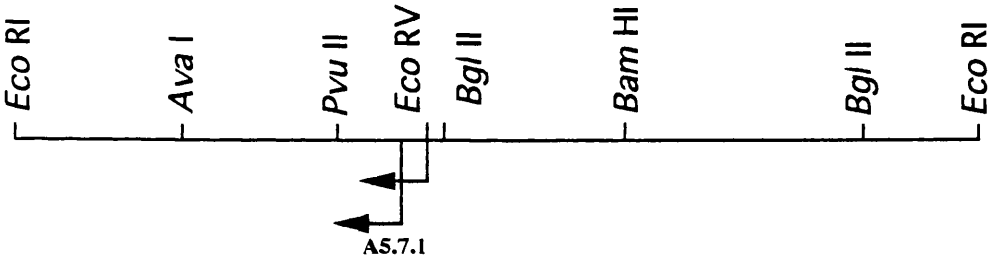
```
      X
    CTCGAGCGTCGATTACTTCCATTACCCATTCCTATCGAAAGACCACCACCCCTAGATACG
  1 ---------+---------+---------+---------+---------+---------+ 60
    GAGCTCGCAGCTAATGAAGGTAATGGGTAAGGATAGCTTTCTGGTGGTGGGGATCTATGC

    CACACCCCCAAGAATCGAATACGAAAACGAAATCGATGGCTGAAACGAGATTACAACCAC
 61 ---------+---------+---------+---------+---------+---------+ 120
    GTGTGGGGGTTCTTAGCTTATGCTTTTGCTTTAGCTACCGACTTTGCTCTAATGTTGGTG

    AATTACGATAACTATTACAATGGCCACCATAANCACAGATGCACAACCACTGGAGCGAAG
121 ---------+---------+---------+---------+---------+---------+ 180
    TTAATGCTATTGATAATGTTACCGGTGGTATTNGTGTCTACGTGTTGGTGACCTCGCTTC

    GACATGCGAATTAAGAATGCGGATGTAGAAGCTGAAAGGAGTAGATCGAATCTAGTGAAT
181 ---------+---------+---------+---------+---------+---------+ 240
    CTGTACGCTTAATTCTTACGCCTACATCTTCGACTTTCCTCATCTAGCTTAGATCACTTA

    ACTTAAGGCGTTAAGCGTACTTCTAGTTTGTATCTCTATATCTCGC
241 ---------+---------+---------+---------+------ 286
    TGAATTCCGCAATTCGCATGAAGATCAAACATAGAGATATAGAGCG
```

A5.6.8 Sequence obtained from the genomic region containing the 5' end of the pC10 clone. The T7 primer was used. The area corresponding to the *Xho* I site in the pC10 clone is highlighted with bold text. There is no *Xho* I site in the genome however as the G residue at position 5 in the recognition sequence, is replaced with an A.

```
    ATTGTAATAGTTATCGTAATTGTGGGTGTCATCTCGTTTCAGCCATCGATTTCGTTTTCG
  1 ---------+---------+---------+---------+---------+---------+ 60
    TAACATTATCAATAGCATTAACACCCACAGTAGAGCAAAGTCGGTAGCTAAAGCAAAAGC

    TATTCGATCTTGGGGGTGTGCGTATCTAGGGGTGGTGGTCTTTCGATAGGAATGGGAATG
 61 ---------+---------+---------+---------+---------+---------+ 120
    ATAAGCTAGAACCCCCACACGCATAGATCCCCACCACCAGAAAGCTATCCTTACCCTTAC

    GAAAGTAATCGACGTCTTTCATTGGATATGTGCGAGTGTCTCTAGTTAAAAAAAATATTA
121 ---------+---------+---------+---------+---------+---------+ 180
    CTTTCATTAGCTGCAGAAAGTAACCTATACACGCTCACAGAGATCAATTTTTTTTATAAT

                        P2
    ATNCAATTTTAAGCAGCTGAAGATGGGGATTTTTCGGAAACCCTCTCATTTCTTCTGTGC
181 ---------+---------+---------+---------+---------+---------+ 240
    TANGTTAAAATTCGTCGACTTCTACCCCTAAAAAGCCTTTGGGAGAGTAAAGAAGACACG

    TGGGGGCTGGGGGAAATCTATCGGTTATTGGTTATCGAAGTCGGTTTAGTCTAATATTGC
241 ---------+---------+---------+---------+---------+---------+ 300
    ACCCCCGACCCCCTTTAGATAGCCAATAACCAATAGCTTCAGCCAAATCAGATTATAACG

    GTGTATAAGTTGCGTAGTCCTTAACGGATCTTTAGTGTACAAATATGAATGCTTTATCTT
301 ---------+---------+---------+---------+---------+---------+ 360
    CACATATTCAACGCATCAGGAATTGCCTAGAAATCACATGTTTATACTTACGAAATAGAA

    GATTTTTTTATTTTTTTTTCGTTTGTTCTCTTTTGTTTGCCATTTGTTACTGCTTTAAAA
361 ---------+---------+---------+---------+---------+---------+ 420
    CTAAAAAAATAAAAAAAAAGCAAACAAGAGAAAACAAACGGTAAACAATGACGAAATTTT

    TCTTTGTTTTTACTTCAGTCTTTTTGAATTTCATTTTTGTGTAGATTTGTTCTTTTCTTG
421 ---------+---------+---------+---------+---------+---------+ 480
    AGAAACAAAAATGAAGTCAGAAAAACTTAAAGTAAAAACACATCTAAACAAGAAAAGAAC
```

```
    ATGCGCCTGCCTTCAACGCGTACTTCTAGTACAAGTACAAAAAAAAAGATTGGATCTACT
481 ---------+---------+---------+---------+---------+---------+ 540
    TACGCGGACGGAAGTTGCGCATGAAGATCATGTTCATGTTTTTTTTTCTAACCTAGATGA

    AAATCTCTATCGACTTCGATCGATTACTATCGATATTCAGCGACTTACACTCGACATCGA
541 ---------+---------+---------+---------+---------+---------+ 600
    TTTAGAGATAGCTGAAGCTAGCTAATGATAGCTATAAGTCGCTGAATGTGAGCTGTAGCT

    TCCACTGGCTAGTTGTATGCGTTATCTGAATTATTCATAGGTCTGTATATCTGTATAATT
601 ---------+---------+---------+---------+---------+---------+ 660
    AGGTGACCGATCAACATACGCAATAGACTTAATAAGTATCCAGACATATAGACATATTAA

    ATGCTATGTCCATATACCTAGATATATCTGCGTGTATATAGAGATACAAACTAGAAGTAC
661 ---------+---------+---------+---------+---------+---------+ 720
    TACGATACAGGTATATGGATCTATATAGACGCACATATATCTCTATGTTTGATCTTCATG

    GCTTAACGCCTTAAAGTATCACTAGATTCGATCTACTCCTTNCAGCTTCTACATCGCATT
721 ---------+---------+---------+---------+---------+---------+ 780
    CGAATTGCGGAATTTCATAGTGATCTAAGCTAGATGAGGAANGTCGAAGATGTAGCGTAA

    GTTAATTCGCATGTCTCGCTCCAGTCCTTGTGCATCTGTGGTTATGGTGGCCAT
781 ---------+---------+---------+---------+---------+---- 834
    CAATTAAGCGTACAGAGCGAGGTCAGGAACACGTAGACACCAATACCACCGGTA
```

A5.7 Sequencing rationale for the region containing the repetitive area in the 6.0kb *Eco* RI-*Eco* RI subclone, p94.R1 from the genomic clone λ9405.



A5.7.1 Sequence obtained from the region containing the repetitive area in the 6.0kb *Eco* RI-*Eco* RI subclone, p94.R1 from the genomic clone λ9405. The repetitive region is indicated in bold.

```
    ACGGTGTTCATACGGACAGACGGCTGGATTGACTCGGATATTGATCCTGATCAAGAATAT
  1 ---------+---------+---------+---------+---------+---------+ 60
    TGCCACAAGTATGCCTGTCTGCCGACCTAACTGAGCCTATAACTAGGACTAGTTCTTATA

    ATATACTTTATATAGTCGGAAACGCTTACTTCTGCCTGTTACATACTTTTCAACAAATCT
 61 ---------+---------+---------+---------+---------+---------+ 120
    TATATGAAATATATCAGCCTTTGCGAATGAAGACGGACAATGTATGAAAAGTTGTTTAGA

    AGTATACCCTTTCAATCTACGAGTAACGGGTATAATAATGTGTAAGGGATGGTGCCACTC
121 ---------+---------+---------+---------+---------+---------+ 180
    TCATATGGGAAAGTTAGATGCTCATTGCCCATATTATTACACATTCCCTACCACGGTGAG

    CCAAAATTATGATATGAGCTTGTGTAACCCAATACAAAAAACACAAAAAAAAAAAATATCT
181 ---------+---------+---------+---------+---------+---------+ 240
    GGTTTTAATACTATACTCGAACACATTGGGTTATGTTTTTTGTGTTTTTTTTTTTATAGA

    GTCTCCGCTTTGAAGTTATTTATTTTGGCCACAATAAGGTATCATTCAGCCAGTAGTGCA
241 ---------+---------+---------+---------+---------+---------+ 300
    CAGAGGCGAAACTTCAATAAATAAAACCGGTGTTATTCCATAGTAAGTCGGTCATCACGT

    TCCCGGCCGTTACNCATCTTCTTATTATAATATCTTTAGATTGTCGCTTTCTCTCTCTAT
301 ---------+---------+---------+---------+---------+---------+ 360
    AGGGCCGGCAATGNGTAGAAGAATAATATTATAGAAATCTAACAGCGAAAGAGAGAGATA

    CCCCCTCTCTTTTTCGCCCAGTATCCCCGTCTCTGCCTGACATCCCCTACGTCATAGACG
361 ---------+---------+---------+---------+---------+---------+ 420
    GGGGGAGAGAAAAAGCGGGTCATAGGGGCAGAGACGGACTGTAGGGGATGCAGTATCTGC
```

```
      AACTATCTTTGGGTA
  421 ---------+----- 435
      TTGATAGAAACCCAT
```

A5.8 Sequence and map obtained from the T3 end of the 3.2kb *Sal* I -*Sal* I restriction fragment from λ9915. Only this end was sequenced. See Sections 6.3.3 and 6.5.2.3 for more details.

```
      TCATTCGATGCTGAAGATTTCGGCTTAAGCGCGTTCAGCTTGTTAAACGCGCTGCTGGCA
    1 ---------+---------+---------+---------+---------+---------+ 60
      AGTAAGCTACGACTTAGGAAGCCGAATTCGCGCAAGTCGAACAATTTGCGCGACGACCGT
FRAME 1 S  F  D  A  E  D  F  G  L  S  A  F  S  L  L  N  A  L  L  A   -

      AATTGCATTCCCCATTTCCCTGCCATTTTCCATTTCCCACCCACACTTTCCGGCCACTTC
   61 ---------+---------+---------+---------+---------+---------+ 120
      TTAACGTAAGGGGTAAAGGGACGGTAAAAGGTAAAGGGTGGGTGTGAAAGGCCGGTGAAG
FRAME 1 N  C  I  P  H  F  P  A  I  F  H  F  P  P  T  L  S  G  H  F   -

      CCGACAGGACACAGATGCTGGCAACGAACCGTGCGACTGAGATGGGATAGTTATTCTCGC
  121 ---------+---------+---------+---------+---------+---------+ 180
      GGCTGTCCTGTGTCTACGACCGTTGCTTGGCACGCTGACTCTACCCTATCAATAAGAGCG
FRAME 1 P  T  G  H  R  C  W  Q  R  T  V  R  L  R  W  D  S  Y  S  R   -

      GTCGTAAATTACAATAATTTCGGTTTAAACTTGGGAAAATGGGGTTTCGGCCATGTTTGG
  181 ---------+---------+---------+---------+---------+---------+ 240
      CAGCATTTAATGTTATTAAAGCCAAATTTGAACCCTTTTACCCCAAAGCCGGTACAAACC
FRAME 1 V  V  N  Y  N  N  F  G  L  N  L  G  K  W  G  F  G  H  V  W   -

      GAGTGTAACTCGCAGGATTTTTTTCATCAAGAATCAACTGAGCATCC
  241 ---------+---------+---------+---------+------- 287
      CTCACATTGAGCGTCCTAAAAAAAGTAGTTCTTAGTTGACTCGTAGG
FRAME 1 E  C  N  S  Q  D  F  F  H  Q  E  S  T  E  H     -
```

A5.9 Both ends of the 1.4kb *Sal* I-*Sal* I restriction fragment from λS2 were sequenced using T3 and T7 primers

A5.91 Sequence and map obtained from the T3 end of the 1.4kb *Sal* I-*Sal* I restriction fragment from λS2. The area with putative homology to a transposable element (see Section 4.2.4 and 4.3.2) is indicated by bold text.

```
      GCGTCGATTACTTTCCATTCCCATTCCTATCGAAAGACCACCACCCCTAGTATACGCACA
    1 ---------+---------+---------+---------+---------+---------+ 60
      CGCAGCTAATGAAAGGTAAGGGTAAGGATAGCTTTCTGGTGGTGGGGATCATATGCGTGT

      CCCCCAAGAATCGAATACGAAAACGAAATCGATGGGTGAAACGAGAACACAACCACAAGT
   61 ---------+---------+---------+---------+---------+---------+ 120
      GGGGGTTCTTAGCTTATGCTTTTGCTTTAGCTACCCACTTTGCTCTTGTGTTGGTGTTCA

      TACGATAACTATTACAATGGCCACCATAACCACAGATGCACAACCACTGGAGCGAAGGAC
  121 ---------+---------+---------+---------+---------+---------+ 180
      ATGCTATTGATAATGTTACCGGTGGTATTGGTGTCTACGTGTTGGTGACCTCGCTTCCTG

      ATGCGAATTAAGAATGCGGATGTAGAAGCTGAAAGGAGTAGATCGAATCTAGGTGAATAC
  181 ---------+---------+---------+---------+---------+---------+ 240
      TACGCTTAATTCTTACGCCTACATCTTCGACTTTCCTCATCTAGCTTAGATCCACTTATG

      TTAAAGGGGTTAAGCGTACTTCTAGTTTGGGATCTCTTATATACACGCAGGTATTTTCTA
  241 ---------+---------+---------+---------+---------+---------+ 300
      AATTTCCCCAATTCGCATGAAGATCAAACCCTAGAGAATATATGTGCGTCCATAAAAGAT

      GGTATTTGGACATAGCATACTTCATACAGATATC
  301 ---------+---------+---------+---- 334
      CCATAAACCTGTATCGTATGAAGTATGTCTATAG
```

A5.9.2 Sequence and map obtained from the T7 end of the 1.4kb *Sal* I -*Sal* I restriction fragment from λS2.

```
    GGGGATAAAATATAATTGTAGTCGCTTTTCGGACAACAGCAATAAGGTAATGCCATAAAC
  1 ---------+---------+---------+---------+---------+---------+ 60
    CCCCTATTTTATATTAACATCAGCGAAAAGCCTGTTGTCGTTATTCCATTACGGTATTTG

    TCTCCTCTAATTTCATTTACTAATTGCCATTAAAATTAATCCAAATCCAGCTCAGTGGGT
 61 ---------+---------+---------+---------+---------+---------+ 120
    AGAGGAGATTAAAGTAAATGATTAACGGTAATTTTAATTAGGTTTAGGTCGAGTCACCCA

    CAAAAAATAGGGACAATGATGGCCAACACCTTGGNTCTGTTGGCCATGGTCCCCAAAATA
121 ---------+---------+---------+---------+---------+---------+ 180
    GTTTTTTATCCCTGTTACTACCGGTTGTGGAACCNAGACAACCGGTACCAGGGGTTTTAT

    GCTTTGGAGGTCAGCTATAAACAATTTGCGAGCAATTCCACAAAAATATTGTGGGTTAAT
181 ---------+---------+---------+---------+---------+---------+ 240
    CGAAACCTCCAGTCGATATTTGTTAAACGCTCGTTAAGGTGTTTTTATAACACCCAATTA

    TGTAGGAAAATGTGGGGAAATGTTTCTACAAGCCCCGAGGACCCGAAACTCAAAAGGACT
241 ---------+---------+---------+---------+---------+---------+ 300
    ACATCCTTTTACACCCCTTTACAAAGATGTTCGGGGCTCCTGGGCTTTGAGTTTTCCTGA

    CGGCAATTTCAACCCACAAATTAACATGGGGGT
301 ---------+---------+---------+--- 333
    GCCGTTAAAGTTGGGTGTTTAATTGTACCCCCA
```

Reference List

References.

Aquade, M., N. Miyashita and C. H. Langley (1992). Polymorphism and divergence in the *mst26a* male accessory-gland gene region in *Drosophila*. Genetics: **132,** 755-770.

Albright, T. D. (1991). Motion perception and the mind-body problem. Current Opinions in Biology: **1,** No 6, 391-393.

Allen, N. D., DG. Cran, SC. Barton, S. Hettle, T. Reik and MA. Surani (1988). Nature: **333**, 852-855.

Andersson, S. and A. Lambertsson (1993). Evolution of the *dec-1* eggshell locus in *Drosophila*. 2. Intraspecific DNA-sequence analysis reveals length mutations in a repetitive region in *Drosophila melanogaster*. Journal of Molecular Evolution: **36,** 536-544.

Andersson, S., S. Saeboelarssen, A. Lambertsson, J. Merriam and M. Jacobslorena (1994). A *Drosophila* 3rd chromosome *minute* locus encodes a ribosomal-protein. Genetics: **137**, 513-520.

Ananiev, E. V., V. E. Barsky, Y. V. Ilyin and M. V. Ryzic (1984). The arrangement of transposable elements in the polytene chromosomes of *Drosophiila melanogaster*. Chromosoma: **90**, 366-377.

Ashburner, M. (1989). *Drosophila*: A Laboratory Handbook (1 ed.). New York: Cold Spring Harbor Laboratory Press.

Ashburner, M. (1989). *Drosophila*: A Laboratory Manual (1 ed.). New York: Cold Spring Harbor Laboratory Press.

Avery, L. (1993). The genetics of feeding in *Caenorhabditis elegans*. Genetics: **133**, No. 4, 897-917.

Bacon, J. P. and N. J. Strausfeld (1986). The dipteran giant fiber pathway - neurons and signals. Journal of Comparative Physiology, A Sensory Neural and Behavioral Physiology: **A158**, 529-548.

Baird, D. H., A. P. Schalet and R. J. Wyman (1990). The *Passover* locus in *Drosophila melanogaster* - complex complementation and different effects on the giant fiber neural pathway. Genetics: **126**, 1045-1059.

Baird, D. H., M. Koto and R. J. Wyman (1993). Dendritic reduction in *Passover*, a *Drosophila* mutant with a defective giant fiber neuronal pathway. Journal of Neurobiology: **24**, No 7, 971-984.

Baird, D. H. (1988). Genetics of *Passover*, a *Drosophila* mutant affecting neuronal connectivity. Ph. D. Thesis. Yale University, CT.

Balakrishnan, R. and V. Rodrigues (1991). The *Shaker* and *shaking-B* genes specify elements in the processing of gustatory information in *Drosophila melanogaster*. Journal of Experimental Biology: **157**, 161-181.

Ball, D. W., Azzoli, C. G. Baylin, S. B., Chi, D., Dou, S., DonisKeller, H., Cumaraswamy, A., Borges, M. and Nelkin, B. D. (1993). Identification of a human *achaete-scute* homolog highly expressed in neuroendocrine tumors. Proceedings of the National Academy of Science (U. S. A.): **90**, No. 12, 5648-5652

Barnes, T. M. (1994). *Opus* - a growing family of gap junction proteins. Trends In Genetics: **10**, No 9, 303-305.

Barrio, R., A. Delarco, H. Cabrera and C. Arribas (1993). Cloning and analysis of the *S2* ribosomal-protein cDNA from *Drosophila*. Nucleic Acids Research: **21**, 351.

Bassemir, U. K. and N. J. Strausfeld (1983). Cytology of cobalt-filled neurons in flys - cobalt deposits at pre-synaptic and post-synaptic sites, mitochondria and the cytoskeleton. Journal of Neurocytology: **12**, 949-970.

Baumgartner, S., D. Martin and R. Chiquetehrismann (1993). *Drosophila* ribosomal-protein *S19* cDNA sequence. Nucleic Acids Research: **21**, 3897.

Belasco J. and G. Brawerman (1993). Control of messenger RNA stability. Academic Press Inc. (London).

Bender, W., P. Spierer and D. S. Hogness (1983a). Chromosomal walking and jumping to isolate DNA from the *ace* and *rosy* loci and the *Bithorax* complex in *Drosophila melanogaster*. Journal of Molecular Biology: **168**, 17-33.

Bender, W., M. Akam, F. Karch, P. A. Beachy, M. Peifer, P. Spierer, E. B. Lewis and D. S. Hogness (1983b). Molecular-genetics of the *bithorax* complex in *Drosophila melanogaster*. Science: **221**, 23-29.

Berg D. E. and M. M. Howe (Eds). (1989). Mobile DNA. American Society for microbiology, Washington D. C.

Berger S. L., and Kimmel A. R., (1987). Guide to molecular cloning techniques. Methods in Enzymology: **152.**

Bingham, PM, MG Kidwell, GM Rubin, (1982). The molecular basis of P-M hybrid dysgenesis: The role of the P element, a P-strain-specific transposon family. Cell: **29**, No- 3, 995-1004.

Birnboim, H. C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Research: **7**, 1513-1523.

Bolivar F and R. L. Rodriguez, P. J. Greene, M. C. Betlach, H. L. Heyneker and H. W. Boyer (1977). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene: **2**, No. 2, 95-113.

Bonini N. M., W. M. Leiserman and S. Benzer (1993). The *eyes absent* gene: Genetic control of cell survival and differentiation in the developing *Drosophila* eye. Cell: **72**, 379-395.

Bonthron, DT, Sultan, P, Collins, T (1991). Structure of the murine c-sis proto-oncogene (Sis, PDGFB) encoding the B chain of platelet-derived growth factor. Genomics: **10**, No. 1, 287-292

Bossy, B., L. Hall and P. Spierer (1984). Genetic-activity along 315kb of the *Drosophila* chromosome. Embo Journal: **3**, 2537-2541.

Bridges, C. B. (1938). A revised map of the salivary gland X chromosome of *Drosophila melanogaster*. Journal of Heridity: **29**: 81 - 86.

Britten R. J., Graham D. E. and Neufeld B. R. (1974). Analysis of repeating DNA sequences by reassociation. Methods in Enzymology: **29E**, 363-380.

Brizuela, B. J., L. Elfring, J. Ballard, J. W. Tamkun and J. A. Kennison (1994). Genetic-analysis of the *brahma* gene of *Drosophila-melanogaster* and polytene chromosome subdivisions 72AB. Genetics: **137**, 803-813.

Brookfield, J. (1993). The generation of sequence similarity in *SINES* and *LINES*. Trends In Genetics: **9**, No. 2, 38.

Brown, N. H. and Kafatos, F. C. (1988). Functional cDNA libraries from *Drosophila* embryos. Journal of Molecular Biology: **203**, 425-437.

Brutlag, D. L. (1980). Molecular arrangement and evolution of heterochromatic DNA. Annual Review of Genetics: **14**, 121-144.

Bullock W. O., Fernandez J. M., and Short J. M (1987). *XL1-Blue* A high efficiency plasmid for transforming *recA Escherichia coli* strains with beta-galactosidiase selection. Biotechniques: **5**, 376-379.

Burns, D. K., B. C. Stark, M. D. Macklin and W. Y. Chooi (1984). Isolation and characterization of cloned DNA-sequences containing ribosomal-protein genes of *Drosophila melanogaster*. Molecular and Cellular Biology: **4**, 2643-2652.

Busseau, I., R. J. Diederich, T. Xu and S. Artavanistsakonas (1994). A member of the *notch* group of interacting loci, *deltex* encodes a cytoplasmic basic-protein. Genetics: **136**, 585-596.

Caizzi, R., C. Caggese and S. Pimpinelli (1993). *Bari-1*, a new transposon-like family in *Drosophila melanogaster* with a unique heterochromatic organization. Genetics: **133**, 335-345.

Campbell, H. D., T. Schimansky, C. Claudianos, N. Ozsarac, A. B. Kasprzak, J. N. cotsell, I. G. Young, H. G. Decouet and G. Miklos (1993). The *Drosophila*

melanogaster flightless-I gene involved in gastrulation and muscle degeneration encodes gelsolin-like and leucine-rich repeat domains and is conserved in *Caenorhabditis-elegans* and humans. Proceedings of the National Academy of Science (U. S. A.): **90**, 11386-11390.

Campos-Ortega, J. A. and M. Haenlin (1992). Regulatory signals and signal molecules in early neurogenesis of *Drosophila melanogaster*. Rouxs Archives of Developmental Biology: **201**, 1-11.

Campuzano, S, Carramolino, L, Cabrera, CV, *et al.*, (1985). Molecular genetics of the *achaete-scute* gene complex of *D. melanogaster*. Cell, **40**, No.2, pp.327-338

Carlson, M., and D. Brutlag (1977). Cloning and characterization of a complex satellite DNA from *Drosophila melanogaster*. Cell: **11**, 371-381.

Cavarec, L. and T. Heidmann (1993). The *Drosophila copia* retrotransposon contains binding sites for transcriptional regulation by homeoproteins. Nucleic Acids Research: **21**, No. 22, 5041-5049.

Cavener, D., and Ray, S. C. (1991). Eukaryotic start and stop translation sites. Nucleic Acids Research: **19**, No 12, 3185-3192.

Chan, Y. L., J. Olvera, V. Paz and I. G. Wool (1992). The primary structure of rat ribosomal protein *L11*. Biochemical and Biophysical ResearchCommunications: **185**, No. 1,356-362.

Chan, T. A., C. A. Chu, K. A. Rauen, M. Kroiher, S. M. Tatarewicz and R. E. Steele (1994). Identification of a gene encoding a novel protein-tyrosine kinase containing *sh2* domains and ankyrin-like repeats. Oncogene: **9**,1253-1259.

Charlesworth, B. P. Sniegowski, W. Stephan (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. Nature, **371**, 215 - 220.

Chen, C. N., T. Malone, S. K. Beckendorf and R. L. Davis (1987). At least 2 genes reside within a large intron of the *dunce* gene of *Drosophila*. Nature: **329**, 721-724.

Chomczynski, P. and N. Sacchi (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate phenol chloroform extraction. Analytical Biochemistry: **162**, No 1, 156-159.

Chooi, W. Y., L. M. Sabatini, M. Macklin and D. Fraser (1980). Group fractionation and determination of the number of ribosomal subunit proteins from *Drosophila melanogaster*embryos. Biochemistry: **19**, 1425-1433.

Chooi, W. Y. (1981). A comparison of the ribosomal-proteins of *Drosophila* ovary, adult, and embryo. Molecular and General Genetics: **184**, 342-346.

Chooi, W. Y., M. D. Macklin, K. R. Leiby, T. H. Hong, S. R. Scofield, L. M. Sabatini and D. K. burns (1982). Purification of *Drosophila* acidic ribosomal-proteins. European Journal Of Biochemistry: **127**, 199-205.

Coggshall, J. C. (1978). Neurons associated with the dorsal longitudinal flight muscles of *Drosophila melanogaster*. Journal of Comparative Neurology: **177**, 707-720.

Cohen, S., Chang, A. and L. Hsu (1972). Chromosomal antibiotic resistance in bacterial genetic transformation of *E. coli* by R-factor. Proceedings of the National Academy of Science (U. S. A.): **69**: 2110-2114.

Coombe, P. E. (1986). The large monopolar cells L1 and L2 are responsible for ERG transients in *Drosophila*. Journal of Comparative Physiology. A Sensory Neural and Behavioral Physiology: **159**, 655-665.

Costa, R., A. A. Peixoto, J. R. Thackeray, R. Dalgleish and C. P. Kyriacou (1991). Length polymorphism in the threonine-glycine-encoding repeat region of the *period* gene in *Drosophila*. Journal of Molecular Evolution: **32**, No. 3, 238-246.

Covarrubias L and F. Bolivar (1982). Construction and characterization of new cloning vehicles. VI. Plasmid pBR329, a new derivative of pBR328 lacking the 482-base-pair inverted duplication. Gene: **17**, No. 1, 79-89.

Craigie, R., (1992). Hotspots and warmspots: intergration specificity of retroelements. Trends in Genetics: **8**, No 6, 187-190.

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

Crews, S. T., J. B. Thomas and C. S. Goodman (1988). The *Drosophila single minded* gene encodes a nuclear protein with sequence similarity to the *per* gene product. Cell: **52**, 143-151.

Crompton, D. E., A. Griffin, J. A. Davies and G. Miklos (1992). Analysis of a cDNA from the neurologically active locus *shaking-B* (*Passover*) of *Drosophila melanogaster*. Gene: **122**, 385-386.

Crompton, D. E., M. Todman, M. Wilkin, S. Ji and J. Davies (1995). Essential and neural transcripts from the *Drosophila shaking-B* locus are differentially expressed in the embryonic mesoderm and pupal nervous system. Developmental Biology: (1995) In press.

Crompton, D. E., (1995). Characterisation of the *shaking-B* locus of *Drosophila melanogaster*. Ph. D. Thesis. Dept. of Genetics, Church Street, University of Glasgow, Glasgow, Scotland.

Davidson E. H. and R. J. Britten (1973). Organization, transcription, and regulation in the animal genome. Quantatative Review of Biology: **48**, No. 4, 565-613.

Davis, T., J. Trenear and M. Ashburner (1990). The molecular analysis of the *el-noc* complex of *Drosophila melanogaster*. Genetics: **126**, 105-119.

Dawid, I. B., E. O. Long, P. P. Dinocera and M. L. Pardue (1981). Ribosomal insertion-like elements in *Drosophila melanogaster* are interspersed with mobile sequences. Cell: **25**, No. 2,, 399-408.

Dealy, C. N., E. C. Beyer and R. A. Kosher (1994). Expression patterns of messenger-RNAs for the gap junction proteins *connexin43* and *connexin42* suggest their involvement in chick limb morphogenesis and specification of the arterial vasculature. Developmental Dynamics: **199**, 156-167.

De Frutos, R., K. R. Peterson and M. G. Kidwell (1992). Distribution of *Drosophila melanogaster* transposable element sequences in species of the *obscura* group. Chromosoma: **161**, No. 5-6, 293-300.

Del Franco, F. F., M. Gendronmaguire, P. J. Swiatek and T. Gridley (1993). Cloning, sequencing and expression of the mouse mammalian *achaete-scute* homolog-1 (*mash1*). Biochimica Et Biophysica Acta: **1171**, No. 3, 323-327.

Devereux, J., P. Haeberli and O. Smithies (1984). A comprehensive set of sequence-analysis programs for the vax. Nucleic Acids Research: **12**, 387-395.

Devlin, R. H., D. G. Holm, K. R. Morin and B. M. Honda (1990a). Identifying a single-copy DNA-sequence associated with the expression of a heterochromatic gene, the *light* locus of *Drosophila melanogaster*. Genome: **33**, No. 3, 405-415.

Devlin, R. H., B. Bingham and B. T. Wakimoto (1990b). The organization and expression of the *light* gene, a heterochromatic gene of *Drosophila melanogaster*. Genetics: **125**, No. 1, 129-140.

DiNocera, P. P. and I. B. Dawid (1983). Interdigitated arrangement of 2 oligo(A)-terminated DNA-sequences in *Drosophila*. Nucleic Acids Research: **11**, No16, 5475-5482.

Dong, J-T, P. W. Lamb, C. W. Rinker-Schaffer, J. Vukanovic, T Ichikawa, J. T. Isaacs and J. C. Barrett (1995). *kal1*, a metastasis suppressor gene for prostate cancer on human chromosome 11p11. 2. Science: **268**, 884-886.

Donnelly, R. J. and B. I. Kiefer (1986). DNA sequences adjacent to and specific for the 1. 672g/cm^3 satellite DNA in the *Drosophila* genome. Proceedings of the National Academy of Science (U. S. A.): **83**: 7172-7176.

Donnelly, R. J. and B. I. Kiefer (1987). DNA sequence of a *Drosophila* satellite associated sequence. Nucleic Acids Research, **15**: 10061.

Doolittle, W. F. and C. Sapienza (1980). Selfish genes, the phenotype paradigm and genome evolution. Nature: **284**, No. 5757, 601-603.

Doshi, P., S. Kaushal, C. Benyajati and C. I. Wu (1991). Molecular analysis of the responder satellite DNA in *Drosophila melanogaster* - DNA bending, nucleosome structure, and *rsp*-binding proteins. Molecular Biology and Evolution: **8**, No. 5, 721-741.

Duboule, D., M. Haenlin, B. Galliot and E. Mohier (1987). DNA sequences homologous to the *Drosophila opa* repeat are present in murine messenger-RNAs that are differentially expressed in fetuses and adult tissues. Molecular and Cellular Biology: 7, 2003-2006.

Duffy, J. B. and J. P. Gergen (1991). The *Drosophila* segmentation gene *runt* acts as a position-specific numerator element necessary for the uniform expression of the sex-determining gene *sex-lethal*. Genes and Development: **15**, No. 12A, 2176-2187.

Duncan I.M. (1987). The *bithorax* complex. Annual Review of Genetics. **21**: 285-319.

Eanes, W. F., M. Kirchner and J. Yoon (1993). Evidence for adaptive evolution of the *g6pd* gene in the *Drosophila melanogaster* and *Drosophila*-simulans lineages. Proceedings of the National Academy of Science (U. S. A.): **90**, No. 16, 7475-7479.

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

Eeken, J., F. H. Sobels, V. Hyland and A. P. Schalet (1985). Distribution of mr-induced sex-linked recessive lethal mutations in *Drosophila melanogaster*. Mutation Research: **150**, 261-275.

Egger, M. D. B. Peng and R. J. Wyman (1989). Patterns of synaptic interactions in a peripheral nerve in wild-type fruitflies and in a neurological mutant (*Passover*). Anatomical Research: **223**: 36A.

Eggleston, W. B., D. M. Johnston-Schlitz and W. R. Engels (1988). P-M hybrid dysgenesis does not mobilise other transposable element families in *Drosophila melanogaster*. Nature, **331**: 368 - 370.

Eisenberg, D, Weiss, RM, Terwilliger, TC (1984). The hydrophobic moment detects periodicity in protein hydrophobicity. Proceedings of the National Academy of Science (U. S. A.): **81**: No. 1I, 140-144.

Endow S. A. and J. G. gall (1975). Differential replication of satellite DNA in polyploid tissues of *Drosophila virilis*. Chromosoma: **50**, 175-192.

Evgenev M. B., V. G. Corces and D-H, Lankenow (1992). *Ulysses* transposable element of *Drosophila* shows structural similarities to functional domains of retroviruses. Journal of Molecular Biology: **225**, 917-924.

Fawcett, D. H., C. K. Lister and D. J. Finnegan (1986). Transposable element controling IR hybrid dysgenesis in *Drosophila melanogaster* are similar to mammalin *LINEs*. Cell, **47**: 1007 - 1015.

Feinberg, A. P. and B. Vogelstein (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Analytical Biochemistry: **137**, 6-13.

Feinberg, A. P. and Vogelstein, B. (1984). Addendum: A technique for radiolabelling DNA restiction endonuclease fragments to high specific activity. Analytical Biochemistry: **137**, 266-267.

Ferreiro, B., P. Skoglund, A. Bailey, R. Dorsky and W. A. Harris (1992). *XASH1*, a *Xenopus* homolog of *achaete-scute*: a proneural gene in anterior regions of the vertebrate CNS. Mechanisms of Development: **40**, No 1-2, 25-36.

Finbow, M. E., S. F. Goodwin, L. Meagher, N. J. Lane, J. Keen, J. Findlay and K. Kaiser (1994). Evidence that the 16-kda proteolipid (subunit-c) of the vacuolar H+-ATPase and ductin from gap-junctions are the same polypeptide in *Drosophila* and *Manduca* - molecular-cloning of the *vha16k* gene from *Drosophila*. Journal of Cell Science: **107**, No. 7, 1817-1824.

Finnegan, D. J., G. M. Rubin, M. W. Young and D. S. Hogness (1978). Repeated gene families in *Drosophila melanogaster*. Cold Spring Harbour Symposium on Quantatative Biology: **42**, 1053-1063.

Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. Trends In Genetics: **5**, 103-107.

Finnegan, D. J. and D. H. Fawcett (1986). Transposable elements in *Drosophila melanogaster*. Offprint from Oxford surveys on Eukaryotic genes. Vol 3.

Franks N. P. and W. R. Lieb, (1991). Stereospecific effects of inhalational general anaesthetic optical isomers on nerve ion channels. Science: **254**, 427-430.

Frigerio, G., M. Burri, D. Bopp, S. Baumgartner and M. Noll (1986). Structure of the segmentation gene *paired* and the *Drosophila prd* gene set as part of a gene network. Cell: **47**, No. 5, 735-746.

Frischauf, A. M., H. Lehrach, A. Poustka and N. Murray (1983). Lambda-replacement vectors carrying polylinker sequences. Journal of Molecular Biology: **170**, 827-842.

Frohman, M. A. (1990). RACE: Rapid amplification of cDNA ends. In M. A. Innis, D. H. Gelfand, J. J. Sninsky, and T. J. White (Eds.), PCR Protocols: A Guide to Methods and Applications San Diego: Academic Press.

Furia, M., F. A. Digilio, D. Artiaco, P. P. Davino, Cavaliere and L. C. Polito (1991). Molecular organisation of the *D. melanogaster Pig-1* gene. Chromosoma: **101**, 49-54.

Furia, M., P. P. Davino, S. Crispi, D. Artiaco and L. C. Polito (1993). Dense cluster of genes is located at the ecdysone-regulated 3C puff of *Drosophila melanogaster*. Journal of Molecular Biology: **231**, 531-538.

Fyrberg, E. A., B. J. Bond, N. D. Hershey, K. S. Mixter and N. Davidson (1981). The actin genes of *Drosophila*: protein coding regions are highly conserved but intron positions are not. Cell: **24**, 107-116.

Fyrberg, E. A., J. W. Mahaffey, B. D. Bond and N. Davidson (1983). Transcripts of the six *Drosophila* actin genes accumulate in a stage and tissue specific manner. Cell: **33**, 115-123.

Ganetzky, B (1977). On the components of segregation distortion in *Drosophila melanogaster*. Genetics: **86**, 321-355.

Gelbart, W. M., V. F. Irish, F. M. Johnston, F. M. Hoffmann, R. K. Blackman, D. Segal, L. M. Posakony and R. Grimaila (1985). The *decapentaplegic* gene complex in *Drosophila melanogaster*. Cold Spring Harbour Symposium on Quantatative Biology: **50**, 119-125.

Gergen, J. P. and E. Wieschaus (1986). Dosage requirements for *runt* in the segmentation of *Drosophila* embryos. Cell: **45**, No. 2, 289-299.

Gergen, J. P. and B. A. Butler (1988). Isolation of the *Drosophila* segmentation gene *runt* and analysis of its expression during embryogenesis. Genes and Development: **2** No. 9, 1179-1193.

Ghysen, A., C. Dambly-Chaudiere, L. Y. Jan and Y. N. Jan (1993). Cell-interactions and gene interactions in peripheral neurogenesis. Genes and Development: **7**, 723-733.

Glisin V., R. Crkvenjakov and C. Byus (1974). Ribonucleic acid isolated by cesium chloride centrifugation. Biochemistry: **13**, No. 12, 2633 - 2637.

Goldberg R. B., W. R. Crain, J. V. Ruderman, G. P. Moore, T. R. Barnett, R. C. Higgins, R. A. Gelfand, G. A. Galau, R. J. Britten and E. H. Davidson (1975). DNA sequence organization in the genomes of five marine invertebrates. Chromosoma: **51**, No. 3, 225-51.

Goodenough, D. A. and L. S. Musil (1993). Gap-junctions and tissue business - problems and strategies for developing specific functional reagents. Journal of Cell Science: **106**, 133-138.

Goodman, C. S., (1994). The likeness of being: phylogenetically conserved molecular mechanisms of growth cone guidance. Cell: **78**, 353-356.

Goodwin, A. W. (1993). Touch - the code for roughness. Current Biology: **3**, 378-379.

Gossler, A., A. L. Joyner, J. Rossant and W. C. Skarnes (1989). Mouse embryonic stem-cells and reporter constructs to detect developmentally regulated genes. Science, **244**, 463-465.

Grabowski, D. T., J. P. Carney and M. R. Kelley (1991a). An adult male specific gene in *Drosophila* containing the repetitive element *opa*. Biochimica Et Biophysica Acta: **1090**, No. 1, 115-118.

Grabowski, D. T., J. P. Carney and M. R. Kelley (1991b). A *Drosophila* gene containing the *opa* repetitive element is exclusively expressed in adult male abdomens. Nucleic Acids Research: **19**, No. 7, 1709.

Gray, M. R., H. V. Colot, L. Guarente and M. Rosbash (1982). Open reading frame cloning - identification, cloning, and expression of open reading frame DNA. Proceedings of the National Academy of Science (U. S. A.): **79**, 6598-6602.

Green, M. M., M. T. Yamamoto and G. Miklos (1987). Genetic instability in *Drosophila melanogaster* - cytogenetic analysis of mr-induced X-chromosome deficiencies. Proceedings of the National Academy of Science (U. S. A.): **84**, 4533-4537.

Greenwald, I. and G. M. Rubin (1992). Making a difference - the role of cell-cell interactions in establishing separate identities for equivalent cells. Cell: **68**, 271-281.

Gribskov, M., Devereux, J., and Burgess, R. R. (1984). The codon preference plot: graphic analysis of protein coding sequences and prediction of gene expression. Nucleic Acids Research: **12**, 539-549.

Grunstein, M., and Hogness, D. S. (1975). Colony Hydridisation: A Method for the Isolation of Cloned DNAs that Contain a Specific Gene. Proceedings of the National Academy of Science (U. S. A.): **72**, 3961.

Gubler, U. and B. J. Hoffman (1983). A simple and very efficient method for generating cDNA libraries. Gene,**25**: No. 2-3, 263 - 283.

Hahn, EJ. (1983). Autoradiography, a review of basic principles. American Laboratory.

Hall, L., P. J. Mason and P. Spierer (1983). Transcripts, genes and bands in 315,000 base-pairs of *Drosophila* DNA. Journal Of Molecular Biology: **169,** 83-96.

Hall L., J. E. Laird J. C. Pascall and R. K. Craig (1984). Guinea pig casein A cDNA. European Journal of Biochemistry: **138**, 585 - 589.

Hanks, S. K., A. M. Quinn and T. Hunter (1988). The protein-kinase family - conserved features and deduced phylogeny of the catalytic domains. Science: **241**, 42-52.

Hanks, S. K., and A. M. Quinn (1991). Protein kinase catalytic domains sequence database: identifacation of conserved features of primary structure and classifacation of family members. Methods in Enzymology: **200**, 38-62.

Hardie, G., and S. K. Hanks (1995). The protein kinase facts book - protein-serine kinases. Academic Press Ltd. Factsbook Series.

Hartenstein, V. and J. A. Campos-Ortega (1986). The peripheral nervous-system of mutants of early neurogenesis in *Drosophila melanogaster*. Rouxs Archives of Developmental Biology: **196**, 210-221.

Hawley, R. S. (1993). Meiosis as an m-thing - 25 years of meiotic mutants in *Drosophila*. Genetics: **135**, No. 3, 613-618.

Haynes, S. R., M. L. Rebbert, B. A. Mozer, F. Forquig and I. B. Dawid (1987). Pen repeat sequences are GGN clusters and encode a glycine-rich domain in a *Drosophila* cDNA homologous to the rat helix destabilizing protein. Proceedings of the National Academy of Science (U. S. A.): **84**, No. 7, 1819-1823.

Hayward, D. C., S. J. Delaney, H. D. Campbell, A. Ghysen, S. Benzer, A. B. Kasprzak, J. N. cotsell, I. G. Young and G. Miklos (1993). The *sluggish-A* gene of *Drosophila melanogaster* is expressed in the nervous-system and encodes proline oxidase, a mitochondrial enzyme involved in glutamate biosynthesis. Proceedings of the National Academy of Science (U. S. A.): **90**, 2979-2983.

Healy, M. J., R. J. Russell and G. Miklos (1988). Molecular studies on interspersed repetitive and unique sequences in the region of the complementation group *uncoordinated* on the X-chromosome of *Drosophila melanogaster*. Molecular and General Genetics: **213**, 63-71.

Healy, MJ, Russell, RJ (1990). Transcriptional analyses of the uncoordinated region of *Drosophila melanogaster*. Genome, **33**, No.6, pp.829-836

Heery, D. M., Gan, F. and Powell, R. (1990). A simple method for subcloning DNA fragments from gel slices. Trends in Genetics: **6**, 173.

Hellman, L. (1993). Characterization of four novel epsilon chain mRNA and a comparative analysis of genes for immunoglobulin E in rodents and man. European Journal of Immunology: **23**, 159-167.

Henikoff S., (1984). Unidirectional digestion with *exonuclease* III creates targeted breakpoints. Gene, **28**, No 3, 351-359.

Henikoff, S., M. A. Keene, K. Fechtel and J. W. Fristrom (1986). Gene within a gene - nested *Drosophila* genes encode unrelated proteins on opposite DNA strands. Cell: **44**, 33-42.

Hey, J. (1987). Transposable elements in *Drosophila*. P. hD. Thesis: State University of New York at Stony Brook.

Heyeck, S. D. and L. J. Berg (1993). Developmental regulation of a murine T-cell specific tyrosine kinase gene, TSK. Proceedings of the National Academy of Science (U. S. A.): **90**, 669-673.

Hilliker, A. J. R. Appels and A. Schalet (1980a). The genetic analysis *of D. melanogaster* heterochromatin. Cell: **21**, 607-619.

Hilliker, A. J., S. H. Clark, A. Chovnick and W. M. Gelbart (1980b). Cytogenetic analysis of the chromosomal region immediately adjacent to the *rosy* locus in *Drosophila melanogaster*. Genetics: **95**, 95-110.

Hilliker, A. J. and R. Appels (1982). Pleiotropic effects associated with the deletion of heterochromatin surrounding rDNA on the X-chromosome of *Drosophila*. Chromosoma: **86**, No. 4, 469-490.

Hodgson, C. P. and R. Z. Fisk (1987). Hybridization probe size control - optimized oligolabelling. Nucleic Acids Research: **15**, No 15, 6295.

Homyk, T., Szidonya, J. and Suzuki, D. T. (1980). Behavioral mutants of *Drosophila melanogaster* III. Isolation and Mapping of Mutations by Direct Visual Observation of Behavioral Phenotypes. Molecular and General Genetics: **177**, 553-565.

Hongo, S. and M. Jacobs-Lorena (1991). Translational discrimination of ribosomal-protein messenger-RNAs in the early *Drosophila* embryo. Developmental Biology: **145**, 338-343.

Horejsi, V and C. Vlcek (1991). Novel structurally distinct family of leucocyte surface glycoproteins including CD9, CD37, CD53 and CD63. FEBS Letters: **288**, No. 1, 1-4.

Horii Z. and A. J. Clarke (1973). Genetic analysis of the *recF* pathway to genetic recombination in *Escherisia coli* K12: isolation and characterisation of mutants. Journal of Molecular Biology: **80**, No. 2, 327 - 344.

Hsieh, T. S. and D. Brutlag (1979). Sequence and sequence variation within the 1. 688g/cm^3 satellite DNA of *Drosophila melanogaster*. Journal of Molecular Biology: **135**, 465-481.

Huber, P. W. and I. G. Wool (1984). Nuclease protection analysis of ribonucleoprotein complexes - use of the cyto-toxic ribonuclease alpha-sarcin to determine the binding-sites for *Escherichia-coli* ribosomal protein-L5, protein-L18, and protein-L25 on 5S RNA. Proceedings of the National Academy of Science (U. S. A.): **81**: 322-326.

Huber, P. W. and I. G. Wool (1986). Identifacation of the binding site on 5S RNA for the transcription factor IIIA: Proposed structure of a common binding site on 5S rRNA and on the gene. Proceedings of the National Academy of Science (U. S. A.): **83**: No. 6, 1593 - 1597.

Hultmark, D., R. Klemenz and W. J. Gehring (1986). Translational and transcriptional control elements in the untranslated leader of the heat-shock gene hsp22. Cell: **44**, 429-438.

Ikeda, K., Koenig, J. H., and Tsuruhara, T. (1980). Organisation of identified axons innervating the dorsal longitudinal flight muscles of *D. melanogaster*. Journal of Neurocytology: **9**, 799-823.

Ilyin, Y. V., V. G. Chmeliauskaite, E. V. Ananiev and G. P. Georgiev (1980a). Isolation and characterisation of a new family of mobile dispersed genetic elements, *mdg3* in *Drosophila melanogaster*. Chromosoma: **81**, 27-53.

Ilyin, Y. V., V. G. Chmeliauskaite, E. V. Ananiev, N. V. Lyubomirskya, A. A. Bayev, and G. P. Georgiev (1980b). Mobile dispersed genetic element *MDG1* of *Drosophila melanogaster*: structural organisation. Nucleic Acids Research: **8**, 5333-5346.

Ilyin, Y. V., V. G. Chmeliauskaite, V. V. Kulguskin and G. P. Georgiev (1980d). Double stranded sequences in RNA of *Drosophila melanogaster*: relation to mobile dispersed genes. Nucleic Acids Research: **8**, 5347-5361.

Ikemura, T (1985). Codon usage and tRNA content in unicellular and multicellular organisms. Molecular Biology and Evolution,.**2**, No.1, pp.13-34.

Irick, H. (1994). A new function for heterochromatin. Chromosoma: **103**, No. 1, 1-3.

Itoh, N., K. Ohta, M. Ohta, T. Kawasaki and I. Yamashina (1989). The nucleotide-sequence of a gene for a putative ribosomal-protein *S31* of *Drosophila*. Nucleic Acids Research: **17**, 2121.

Itoh, N., K. Ohta, M. Ohta, T. Kawasaki and I. Yamashina (1989). The nucleotide-sequence of cDNA for a *Drosophila* ribosomal-protein with homology to rat ribosomal-protein *S26*. Nucleic Acids Research: **17**, 441.

Jacob, L., Opper, M., Metzroth, B., Phannavong, B., and Mechler, B. M. (1987). Structure of the *l(2)gl* gene of *Drosophila* and delimitation of its tumour suppressor domain. Cell, **50**, 215-225.

Ji., S. (1995). Chromosomal walking and isolation of a transcript from the *shaking-B* locus of *Drosophila melanogaster*. MSc. Thesis, University of Glasgow, Glasgow, Scotland.

Jimenez, F. and J. A. Camposortega (1990). Defective neuroblast commitment in mutants of the *achaete-scute* complex and adjacent genes of *Drosophila melanogaster*. Neuron: **5**, 81-89.

John, B. and G. L. G. Miklos (1979). Functional aspects of satellite DNA and heterochromatin. International Review of Cytology: **58**, 1-114.

John, B. and G. L. G. Miklos (1988). The Eukaryote genome in Development and Evolution. 1-416. Unwin-Hyman, London.

Jongen, W., M. E. Fitzgerald, D. J. Asamoto, C. Picolli, T. J. Slaga, D. Gross, M. Takeichi and H. Yamasaki (1991). Regulation of connexin 43 - mediated gap junctional intercellular communication by Ca^{2+} in mouse epidermal cells is controlled by E-cadherin. Journal of Cell Biology: **114**, 545-555

Kania, M. A., A. S. Bonner, J. B. Duffy and J. P. Gergen (1990). The *Drosophila* segmentation gene *runt* encodes a novel nuclear regulatory protein that is also expressed in the developing nervous-system. Genes and Development: **4**, 1701-1713.

Kankel, D. R. and K. White (1993). Developmental neurobiology of the fruit-fly *Drosophila melanogaster*. Journal of Neurobiology: **24**, 699-700.

Kanter, H. L., J. E. Saffitz and E. C. Beyer (1994). Molecular-cloning of 2 human cardiac gap junction proteins, *connexin40* and *connexin45*. Journal of Molecular and Cell. Cardiology: **26**, 861-868.

Kaplan N. L., R. R. Hudson and C. H. langley (1989). The "hitchhiking effect" revisited. Genetics: **123**, 887-899.

Karn J. S. Brenner, L. Barnett and G. Cesareni (1980). Novel bacteriophage lambda cloning vector. Proceedings of the National Academy of Science (U. S. A.): **77**, No. 9, 5172-6.

Karn J, S. Brenner and L Barnett (1983). New bacteriophage lambda vectors with positive selection for cloned inserts. Methods in Enzymology: **101**, 3-19.

Karpen, G. H. and A. C. Spradling (1990). Reduced DNA polytenization of a minichromosome region undergoing position-effect variegation in *Drosophila*. Cell: **63**, No. 1, 97-107.

Kaufman T.C., M.A. Seeger, G.A. Olsen (1990). Molecular and genetic organisation of the *Antennapedia* gene complex of *Drosophila melanogaster*. Advances in Genetics, **27**: 309-362.

Kay, M. A., J. Y. Zhang and M. Jacobslorena (1988). Identification and germline transformation of the ribosomal-protein *rp21*-gene of *Drosophila* - complementation analysis with the *minute* qIII locus reveals identity. Molecular and General Genetics: **213**, 354-358.

Kay MA, Jacobs-Lorena M (1985). Selective translational regulation of ribosomal protein gene expression during early development of *Drosophila melanogaster*. Molecular and Cellular Biology: **5** No. 12, 3583-3592.

Kennison, JA (1995). The polycomb and trithorax group proteins of Drosophila: Trans-regulators of homeotic gene function. Annual Review of Genetics, No.29, pp.289-303.

Kidd, S. J. and D. M. Glover (1980). A DNA segment from D. *melanogaster* which contains five tandemly repeating units homologous to the major rDNA insertion. Cell: **19**, No. 1, 103-119.

Kidd, S., T. J. Lockett and M. W. Young (1983). The *Notch* locus of *Drosophila melanogaster*. Cell: **34**, 421-433.

Kidd, S., M. R. Kelley and M. W. Young (1986). Sequence of the *notch* locus of *Drosophila melanogaster* - relationship of the encoded protein to mammalian clotting and growth-factors. Molecular and Cellular Biology: **6**, 3094-3108.

King, D. G. and K. L. Valentino (1983). On neuronal homology - a comparison of similar axons in *Musca*, *Sarcophaga*, and *Drosophila* (diptera, schizophora). Journal of Comparative Neurobiology: **219**, 1-9.

King, D. G., and Wyman, R. (1980). Anatomy of the giant fibre pathway in *Drosophila*. I. Three thoracic components of the pathway. Journal of Neurocytology: **9**, 753-770.

King, D. G. and M. A. Tanouye (1983). Anatomy of motor axons to direct flight muscles in *Drosophila*. Journal of Experimental Biology: **105**, 231.

Kliman, R. M. and J. Hey (1993a). Reduced natural-selection associated with low recombination in *Drosophila melanogaster*. Molecular Biology and Evolution: **10**, No. 6, 1239-1258.

Kliman, R. M. and J. Hey (1993b). DNA sequence variation at the *period* locus within and among species of the *Drosophila melanogaster* complex. Genetics: **133**, No. 2, 375-387.

Knighton, D. R., J. Zheng, L. F. Ten-Eyck, V. A. Ashford, N. Xuong, S. S. Taylor and J. M. Sowadski (1991). Crystal structure of the catalytic subunit of cyclic adenosine monophosphate-dependent protein kinase. Science: **253**, No. 5018, 407-414.

Knipple, D. C., T. R. Fuiest and R. J. MacIntyre (1991). Molecular cloning and analysis of the chromosomal region 26A of *Drosophila melanogaster*. Molecular and General Genetics: **226**, 241-249.

Knust, E, Schrons, H, Grawe, F, CamposOrtega, JA (1992). Seven genes of the *Enhancer of split complex* of *Drosophila melanogaster* encode helix-loop-helix proteins. Genetics: **132**, No. 2, 505-518.

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

Koto, M., M. A. Tanouye, A. Ferrus, J. B. Thomas and R. J. Wyman (1981). The morphology of the cervical giant fiber neuron of *Drosophila*. Brain Research: **221**, 213-217.

Kramers, P., A. P. Schalet, E. Paradi and L. Huiserhoogteyling (1983). High proportion of multi-locus deletions among hycanthone-induced X-linked recessive lethals in *Drosophila melanogaster*. Mutation Research: **107**, 187-201.

Kreitman, M. and R. R. Hudson (1991). Inferring the evolutionary histories of the *adh* and *adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. Genetics: **127**, No. 3,, 565-582.

Krishnan, S. N., E. Frei, G. P. Swain and R. J. Wyman (1993). *Passover* - a gene required for synaptic connectivity in the giant fiber system of *Drosophila*. Cell: **73**, 967-977.

Krishnan, S. N., E. Frei, A. P. Schalet and R. J. Wyman (1995). Molecular basis of intracistronic complementation in the *Passover* locus of *Drosophila*. Proceedings of the National Academy of Science (U. S. A.): **92**, 2021-2025.

Kyte J. and Doolittle R. F. (1982). A simple method for displaying the hydropathic character of a protein. Journal of Molecular Biology: **157**, No. 1, 105-132.

Lakhotia, S. C., and J. Jacob, (1974). EM autoradiographic studies on polytene nuclei of *Drosophila melanogaster*. II. Organisation and transcriptive activity of the chromocentre. Experimental Cell Research: **86**, 253-263.

Langley, C. H., A. E. Shrimpton, T. Yamazaki, N. Miyashita, Y. Matsuo and C. F. Aquadro (1988). Naturally-occurring variation in the restriction map of the *amy* region of *Drosophila melanogaster*. Genetics: **120**, No. 1, 619-629.

Langley, C. H., R. Hudson, N. Kaplan and B. Charlesworth (1988). On the role of unequal exchange in the containment of transposable element copy number. Genetical Research: **52**, 223-235

Langley, C. H., J. Macdonald, N. Miyashita and M. Aguade (1993). Lack of correlation between interspecific divergence and intraspecific polymorphism at the *suppressor of forked* region in *Drosophila melanogaster* and *Drosophila-simulans*. Proceedings of the National Academy of Science (U. S. A.): **90**, No. 5, 1800-1803.

Laurie, C. C., J. T. Bridgham and M. Choudhary (1991). Associations between DNA-sequence variation and variation in expression of the *adh* gene in natural-populations of *Drosophila melanogaster*. Genetics: **129**, No. 2, 489-499.

Leer, RJ, Van, RaamsdonkDuin, MMC, Mager, WH, Planta, RJ (1985). Conserved sequences upstream of yeast ribosomal protein genes. Current Opinions in Genetics: **9**, No. 4, 273-277.

Lefevre, G. (1981). The distribution of randomly recovered X-ray-induced sex-linked genetic effects in *Drosophila melanogaster*. Genetics: **99**, 461-480.

Lefevre, G. and Watkins, W. (1986). The question of the total gene number in *Drosophila melanogaster*. Genetics: **113**, 869-895.

Lehmann, R., F. Jimenez, U. Dietrich and J. A. Camposortega (1983). On the phenotype and development of mutants of early neurogenesis in *Drosophila melanogaster*. Wilhelm Rouxs Archives of Developmental Biology: **192**, 62-74.

Levy L. S. and Manning J. E. (1981). Messenger RNA sequence complexity and homology in developmental stages of *Drosophila*. Developmental Biology: **85**, 141-149.

Levy W. B. and McCarthy B. J. (1975)., Messenger RNA complexity in *Drosophila melanogaster*. Biochemistry, **14**, 2440-2446.

Levy W. B. C. B. Johnson and B. J. McCarthy (1975). Diversity of sequences in total and polyadenylated nuclear RNA from *Drosophila* cells. Nucleic Acids Research: **3**, 1777-1789.

Lieber, T., S. Kidd, E. Alcamo, V. Corbin and M. Young (1993). Antineurogenic phenotypes induced by truncated *Notch* proteins indicate a role in signal transduction and may point to a novel function for *Notch* in nuclei. Genes and Development: **7**, 1949-1965.

Lifschytz, E. and Falk, R. (1968). Fine structure analysis of a chromosome segment in *Drosophila melanogaster*. Analysis of X-ray induced lethals. Mutation Research: **6**, 235-244.

Lifschytz, E. and Falk, R. (1969). Fine Structure Analysis of a Chromosome segment in *Drosophila melanogaster*. Analysis of Ethyl methanesulphonate - induced lethals. Mutation Research: **8**, 147-155.

Lifschytz, E. and Yakobovitz, N. (1978). The role of X-linked lethal and viable male sterile mutations in male gametogenesis of *Drosophila melanogaster*. Molecular and General Genetics: **161**, 275-284.

Lindsley D. and G. Zimm (1986). The genome of *Drosophila melanogaster*. Part 2: Lethals, maps. *Drosophila* Information Service, **64**: 29-36.

Lindsley D. and G. Zimm (1987). The genome of *Drosophila melanogaster*. Part 3: rearrangements. *Drosophila* Information Service, **65**: 6-29.

Lindsley, D. L. and Zimm, G. G. (1992). The Genome of *Drosophila melanogaster*. San Diego, CA. Academic Press, Inc

Lipman, D. J., and Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. Science: **227**, 1435-1441.

Lipshitz, H. D. and D. R. Kankel (1985). Specificity of gene-action during central nervous-system development in *Drosophila melanogaster* - analysis of the *lethal (1) optic ganglion reduced* locus. Developmental Biology: **108**, 56-77.

Livak, K. J., R. Freund, M. Schweber, P. C. Wensink and M. Meselson (1978). Sequence organisation and transcription at two heat shock loci in *Drosophila*. Proceedings of the National Academy of Science (U. S. A.): **75**, 5613-5617.

Lohe, A. R. and D. L. Brutlag (1986). Multiplicity of satellite DNA-sequences in *Drosophila melanogaster*. Proceedings of the National Academy of Science (U. S. A.): **83**, No. 3, 696-700.

Lovett, M, Kere, J, Hinton, LM (1991). Direct selection: A method for the isolation of cDNAs encoded by large genomic regions. Proceedings of the National Academy of Sciences of the United States of America, **88**, No.21, pp.9628-9632

Lovett, M (1994). Fishing for complements: Finding genes by direct selection. Trends in Genetics, **10**, No.10, pp.352-357.

Lowenhaupt, K. Y., A. Rich and M. L. Pardue (1989). random distribution of long mono-nucleotide and dinucleotide repeats in *Drosophila* chromosomes - correlations with dosage compensation, heterochromatin, and recombination. Molecular and Cellular Biology: **9**, 1173-1182.

Lubahn, D. B., D. R. Joseph, P. M. Sullivan, H. F. Willard, F. S. French and E. M. Wilson (1988). Cloning of human androgen receptor complementary-DNA and localization to the X-chromosome. Science: **240**, 327-330.

Lutzelschwab, R., Klambt, C., Rossa, R., and Schmidt, O. (1987). A protein product of the *Drosophila* recessive tumour gene *l(2) giant glarvae*, potentially has cell adhesion properties. EMBO Journal: **6,** No 6, 1791-1797.

Magoulas, C. and D. A. Hickey (1992). Isolation of 3 novel *Drosophila melanogaster* genes containing repetitive sequences rich in GCN triplets. Genome: **35**, No. 1, 133-139.

Maki, C., D. D. Rhoads, M. J. Stewart, B. Vanslyke and D. J. Roufa (1989). The *Drosophila melanogaster rpS17* gene encoding ribosomal-protein *S17*. Gene: **79**, 289-298.

Manning, J. E., C. W. Schmid and N. Davidson (1975). Interspersion of repetitive and -repetitive DNA sequences in the *Drosophila melanogaster* genome. Cell: **4**, 141-155.

Masai, I, Hosoya, T, Kojima, SI, Hotta, Y (1992). Molecular cloning of a *Drosophila* diacylglycerol kinase gene that is expressed in the nervous system and muscle. Proceedings of the National Academy of Science (U. S. A.): **89**, No. 13, 6030-6034.

Mason, J. M. and H. Biessman (1995). The unusual telomeres of *Drosophila*. Trends in Genetics: **11**, No. 2, 58-62.

Matsumoto H., Ozaki M., Isono L., Randall L., Florean M., Larrivee D., Pye Q. and Pak W. L., (1986). *Nina*: a class of mutations affecting photoreceptor-specific gene functions. In Molecular Neurobiology of *Drosophila*., **16** (Abst.), New York, Cold Spring Harbour Laboratory.

Matthews, R. (1992). A low fat theory of anaesthesia. Science: **255**, 156-157.

McGinnis, W., M. S. Levine, E. Hafen (1984). A conserved DNA sequence in homoeotic genes of the *Drosophila antennapedia* and *bithorax* complexes. Nature: **308**, No. 5958, 428-433.

Mckee, B. D. and G. H. Karpen (1990). *Drosophila* ribosomal-RNA genes function as an X-Y pairing site during male meiosis. Cell: **6**, No. 1, 61-72.

Mckee, B. D., L. Habera and J. A. Vrana (1992). Evidence that intergenic spacer repeats of *Drosophila melanogaster* ribosomal-RNA genes function as x-y pairing sites in male meiosis, and a general-model for a chiasmatic pairing. Genetics: **132**, 529-544.

McNabb, S. L. and M. Ashburner (1993). Identification of a *Drosophila* protein similar to rat s13 and archaebacterial-s11 ribosomal-proteins. Nucleic Acids Research: **21**, 2523.

Mechler B. M., W. McGinnis and W. J. Gehring (1985). Molecular cloning of *lethal (2) giant larva*, a recessive oncogene in *Drosophila melanogaster*. EMBO Journal: **4**, 1551 - 1557.

Mechler B. M., (1987). In Methods in Enzymology, Vol. 152, pp 241-248. Eds. Berger, S. L. and Kimmel, A. R. . Academic Press Inc.

Messing, J. (1983). New m13 vectors for cloning. Methods In Enzymology: **101**, 20-78.

Meyerowitz, E. M. and D. S. Hogness (1982). Molecular-organization of a *Drosophila* puff site that responds to ecdysone. Cell: **28**, No. 1, 165-176.

Miklos, G., M. J. Healy, P. Pain, A. J. Howells and R. J. Russell (1984). Molecular and genetic-studies on the euchromatin-heterochromatin transition region of the X-chromosome of *Drosophila melanogaster*. 1. a cloned entry point near to the *uncoordinated* (*unc*) locus. Chromosoma: **89**, 218-227.

Miklos, G., L. E. Kelly, P. E. Coombe, C. Leeds and G. Lefevre (1987). Localization of the genes *shaking-B*, *small optic lobes*, *sluggish-A*, *stoned* and *stress-sensitive-C* to a well-defined region on the X-chromosome of *Drosophila melanogaster*. Journal of Neurogenetics: **4**, 1-19.

Miklos, G., M. T. Yamamoto, J. Davies and V. Pirrotta (1988). Microcloning reveals a high-frequency of repetitive sequences characteristic of chromosome-4 and the beta-heterochromatin of *Drosophila melanogaster*. Proceedings of the National Academy of Science (U. S. A.): **85**, 2051-2055.

Miklos, G. and J. N. Cotsell (1990). Chromosome structure at interfaces between major chromatin types - alpha-heterochromatin and beta-heterochromatin. Bioessays: **12**, 1-6.

Mitchelson, A., M. Simonelig, C. Williams and K. Ohare (1993). Homology with *Saccharomyces-cerevisiae RNA14* suggests that phenotypic suppression in *Drosophila melanogaster* by *suppressor of forked* occurs at the level of RNA stability. Genes and Development: 7, 241-249.

Miyashita, N. and C. H. Langley (1988). Molecular and phenotypic variation of the *white* locus region in *Drosophila melanogaster*. Genetics: **120**, No. 1, 199-212.

Miyashita, N. (1990). Molecular and phenotypic variation at the *Zw* locus region in *Drosophila melanogaster*. Genetics: **125**, 407-419.

Miyashita, N. T., M. Aguade and C. H. Langley (1993). Linkage disequilibrium in the *white* locus region of *Drosophila melanogaster*. Genetical Research: **62**, No. 2, 101-109.

Mizrokhi, L. J., S. G. Georgieva and Y. V. Llyin (1988). *Jockey*, a mobile *Drosophila* element similar to mammalian *LINES*, is transcribed from the internal promoter by RNA polymerase-II. Cell: 685-691.

Moffatt, F., M. R. Hill, F. Cornelis, C. Schou, J. A. Faux, R. P. Young, A. L. James, G. ryan, P. Lesouef, A. W. Musk, J. M. Hopkin and W. Cookson (1994). Genetic-linkage of T-cell receptor alpha/delta complex to specific IgE responses. Lancet: 1597-1600.

Morata, G (1993). Homeotic genes of *Drosophila*. Current Opinions in Genetics and Development: **3**, No. 4, 606-614.

Morgan, JG, Dolganov, GM, Robbins, SE, Hinton, LM, Lovett, M (1992). The selective isolation of novel cDNAs encoded by the regions surrounding the human interleukin 4 and 5 genes. Nucleic Acids Research, **20**, No.19, pp.5173-5179.

Morgan, P. G., M. Sedensky and P. M. Meneely (1990). Multiple sites of action of volatile anaesthetics in *Caenorhabditis elegans*. Proceedings of the National Academy of Science (U. S. A.): 7, 2965-2969.

Mount, S. M. and G. M. Rubin (1985). Complete nucleotide sequence of the *Drosophila* transposable element *copia* - homology between *copia* and retroviral proteins. Molecular and Cellular Biology: **5**, No. 7, 1630-1638.

Muralidhar, M. G. and J. B. Thomas (1993). The *Drosophila bendless* gene encodes a neural protein related to ubiquitin-conjugating enzymes. Neuron: **11**, 253-266.

Murray, N. E., W. J. Brammar and K. Murray (1977). Lambdoid phages that simplify the recovery of *in vitro* recombinants. Molecular and General Genetics: **150**, 1,: 53-61.

Muskavitch, M. and D. S. Hogness (1982). An expandable gene that encodes a *Drosophila* glue protein is not expressed in variants lacking remote upstream sequences. Cell: **29**, No. 3, 1041-1051.

Norrander, J., T. Kempe and J. Messing (1983). Construction of improved M13 vectors using oligodeoxynucleotide-directed mutagenesis. Gene: **26**, 1,101-6.

Nose, A, Mahajan, VB, Goodman, CS (1992). *Connectin*: A homophilic cell adhesion molecule expressed on a subset of muscles and the motoneurons that innervate them in *Drosophila*. Cell: **70**, No. 4, 553-567.

Nose, A, Takeichi, M, Goodman, CS (1994). Ectopic expression of *Connectin* reveals a repulsive function during growth cone guidance and synapse formation. Neuron: **13**, No. 3, 525-539.

Noselli, S. and A. Vincent (1992). The *Drosophila melanogaster* ribosomal-protein *L17A* encoding gene. Gene: **118**, 273-278.

Nusslein-Volhard, C. (1991). Determination of the embryonic axes of *Drosophila*. Development: **112**, Suppl. 1, 1-10.

O'Connell P. and M. Rosbach(1984). Sequence, structure and codon preference of the *Drosophila ribosomal protein 49* gene. Nucleic Acids Research: **12**, 5495-5513.

Oh, CE, McMahon, R, Benzer, S, Tanouye, MA (1994). *bendless*, a *Drosophila* gene affecting neuronal connectivity, encodes a ubiquitin-conjugating enzyme homolog. Journal of Neuroscience, **14**, No.5 II, pp.3166-3179.

Ohta, T. (1993). Amino acid substitution at the *Adh* locus of *Drosophila* is facilaitated by small popualtion size. Proceedings of the National Academy of Science (U. S. A.): **90**, 4548-4551.

Okayama, H. and P. Berg (1982). High-efficiency cloning of full-length cDNA. Molecular and Cellular Biology: **2**, No. 2, 161-170.

Olson, P. F., T. Salo, K. Garrison and J. H. Fessler (1993). *Drosophila* acidic ribosomal protein-*rpAa2* - sequence and characterization. Journal Of Cellular Biochemistry: **51**, 353-359.

Orgel, L. E. and F. Crick (1980). Selfish DNA: The ultimate parasite. Nature: **284**, No. 5757, 604-607.

O'Tousa J. E., Baehr N., Martin R. L., Hirsh J., Pak W. L. and M. L Applebury. (1985). The *Drosophila ninaE* gene encodes an opsin. Cell, **40**, 839-850.

Pardue, M. L., K. Lowenhaupt, A. Rich and A. Nordheim (1987). (dC-dA)n. (dG-dT)n sequences have evolutionarily conserved chromosomal locations in *Drosophila* with implications for roles in chromosome structure and function. EMBO Journal: **6**, 1781-1789.

Parimoo, S, Patanjali, SR, Shukla, H, Chaplin, DD, Weissman, SM (1991). cDNA selection: Efficient PCR approach for the selection of cDNAs encoded in large chromosomal DNA fragments. Proceedings of the National Academy of Sciences of the United States of America, **88**, No.21, pp.9623-9627.

Parkhurst, S. M. and V. G. Corces (1986). Interactions among the *gypsy* transposable element and the *yellow* and the *suppressor of hairy wing* loci in *Drosophila melanogaster*. Molecular and Cellular Biology: **6**, No. 1, 47-53.

Patel, R. and M. Jacobs-Lorena (1992). Generation of minute phenotypes by a transformed antisense ribosomal-protein gene. Developmental Genetics: **13**, 256-263.

Peacock, W. J., D. L. Brutlag, E. Goldring, R. Appels, C. W. Hinton and D. L. Lindsley (1973). The organisation of highly repeated DNA sequences *in Drosophila melanogaster* chromosomes. Cold Spring Harbour Symposium on Quantatative Biology: **38**, 405-416.

Peacock, W. J., A. R. Lohe, W. L. Gerlach, P. Dunsmuir, E. S. Dennis and R. Appels (1977). Fine structure and evolution of DNA in heterochromatin. Cold Spring Harbour Symposium on Quantatative Biology: **42**, 1121-1135.

Pearson, W. R. and D. J. Lipman (1988). Improved tools for biological sequence comparison. Proceedings of the National Academy of Science (U. S. A.): 2444-2448.

Perrimon, N. and A. P. Mahowald (1986). *L(1)hopscotch*, a larval pupal zygotic lethal with a specific maternal effect on segmentation in *Drosophila*. Developmental Biology: **118**, 28-41.

Perrimon, N., L. Engstrom and A. P. Mahowald (1989a). Zygotic lethals with specific maternal effect phenotypes in *Drosophila melanogaster*. 1. Loci on the X-chromosome. Genetics: **121**, 333-352.

Perrimon, N., D. Smouse and G. Miklos (1989b). Developmental genetics of loci at the base of the X-chromosome of *Drosophila melanogaster*. Genetics: **121**, 313-331.

Petrovich, T. Z., J. Merakovsky and L. E. Kelly (1993). A genetic-analysis of the *stoned* locus and its interaction with *dunce*, *shibire* and *suppressor of stoned* variants of *Drosophila melanogaster*. Genetics: **133**, 955-965.

Phelan, P, Nakagawa, M, Wilkin, MB, Moffat, KG, OKane, CJ, Davies, JA,

Bacon, JP (1996). Mutations in *shaking-B* prevent electrical synapse formation in the *Drosophila* giant fiber system. Journal of Neuroscience, **16**, No.3, pp.1101-1113

Phillis, R. W., A. T. Bramlage, C. Wotus, A. Whittaker, L. S. Gramates, D. Seppala, F. farahanchi, P. Caruccio and R. K. Murphey (1993). Isolation of mutations affecting neural circuitry required for grooming behavior in *Drosophila melanogaster*. Genetics: **133**,581-592.

Pirrotta, V., C. Hadfield and G. Pretorius (1983). Microdissection and cloning of the *white* locus and the 3B1-3C2 region of the *Drosophila* X-chromosome. EMBO Journal: **2**, 927-934.

Pirrotta, V. (1990). Transvection and long distance gene regulation. Bioessays: **12**, No9, 409-414

Potter, S. S., W. J. Brorein, P. Dunsmuir and G. M. Rubin (1979). Transposition of elements of the *412*, *copia* and *297* dispersed repeated gene families in *Drosophila*. Cell: **17**, 415-427.

Power, M. E. (1948). The thoracico-abdominal nervous system of an adult insect, *Drosophila melanogaster*. Journal of Comparative Neurobiology: **88**, 347-409.

Preiss, A, UB Rosenberg, A. Kienlin, *et, al.* (1985). Molecular genetics of *Kruppel*, a gene required for segmentation of the *Drosophila* embryo. Nature: **313,** No. 5997, 27-32

Price, B. D., Z. Chang, R. Smith, S. Bockheim and A. Laughon (1993). The *Drosophila* neuralized gene encodes a C3HC4 Zinc-finger. EMBO Journal: **12**, 2411-2418.

Procunier, J. D. and R. J. Dunn (1978). Genetic and molecular organisation of the 5S locus and mutants in *Drosophila melanogaster*. Cell: **15**, 1087-1093.

Proudfoot, N. J., and Whitelaw, E. (1988). Termination and 3' end processing of eukaryotic RNA. In B. D. Hanes and D. M. Glover (Eds.), Transcription and Splicing (pp. 97-129). Oxford: IRL Press.

Prudhomme, N, Gans, M, Masson, M, Terzian, C, Bucheton, A (1995). *Flamenco*, a gene controlling the gypsy retrovirus of *Drosophila melanogaster*.Genetics, **139**, No.2, pp.697-711.

Puissant C, Houdebine LM (1990). An improvement of the single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. BioTechniques: **8,** No. 2, 148-149.

Qian, S., J. Y. Zhang, M. A. Kay and M. Jacobslorena (1987). Structural-analysis of the *Drosophila rpa1* gene, a member of the eukaryotic a-type ribosomal-protein family. Nucleic Acids Research: **15**, 987-1003.

Qian, S, Hongo, S, JacobsLorena, M (1988). Antisense ribosomal protein gene expression specifically disrupts oogenesis in *Drosophila melanogaster*. Proceedings of the National Academy of Sciences of the United States of America, **85**, No.24, pp.9601-9605.

Qiu, Y. H. and R. L. Davis (1993). Genetic dissection of the learning memory gene dunce of *Drosophila melanogaster*. Genes and Development: 7, 1447-1458.

Rafti, F., G. Gargiulo, A. Manzi, C. Malva, G. Grossi, S. Andone and F. Graziani (1988). Isolation and structural-analysis of a ribosomal-protein gene in *Drosophilamelanogaster*. Nucleic Acids Research: **16**, 4915-4926.

Ratner, V. A., S. A. Zabanov, O. V. Kolesnikova and L. A. Vasilyeva (1992). Induction of the mobile genetic element *Dm-412* transpositions in the *Drosophila* genome by heat-shock treatment. Proceedings of the National Academy of Science (U. S. A.): **89**, No. 12, 5650-5654.

Rhaman, R. and D. L. Linsley (1980). Male sterilising interactions between duplications and deficiencies for proximal X-chromosome material in *Drosophila melanogaster*. Genetics: **90**, 40-64.

Rheuben, M. B. and A. E. Kammer (1981). Membrane structures and physiology of an immature synapse. Journal of Neurocytology: **10**, 557-575.

Rigby, W. J., M. Dieckman, C. Rhodes and P. Berg (1977). Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. Journal Molecular Biology: **113**, 237-251.

Ritossa, F. M. (1976). The *bobbed* locus. In: The genetics and biology of *Drosophila*. (eds. M. Ashburner and E. Novitski). Vol 1b, 801-846. Academic Press, London and New York.

Rodrigues, V., S. Sathe, L. Pinto, R. Balakrishnan and O. Siddiqi (1991). Closely linked lesions in a region of the X-chromosome affect central and peripheral steps in gustatory processing in *Drosophila*. Molecular and General Genetics: **226**, 265-276.

Rosen, K. M. and L. Villa-Komaroff (1991). An alternative method for the visualisation of RNA in formaldehyde agarose gels. Focus: **12**, No. 2, 23-24.

Rosendahl, G, Andreasen, PH, Kristiansen, K (1991). Structure and evolution of the *Tetrahymena thermophila* gene encoding ribosomal protein L21. Gene: **98**, No. 2, 161-167.

Rothberg, JM, ArtavanisTsakonas, S (1992). Characterization of a conserved carboxy-terminal sequence in secreted proteins and a motif implicated in extracellular protein interactions. Journal of Molecular Biology: **227,** No. 2, 367-370.

Rubin G.M. (1986). A *Drosophila* laboratory manual. Unpublished Lab Copy.

Ruohola-Baker, H., L. Y. Jan and Y. N. Jan (1994). The role of gene cassettes in axis formation during *Drosophila* oogenesis. Trends in Genetics: **10**, 89-94.

Russell, R. J., M. J. Healy and J. G. Oakeshott (1992). Molecular analysis of the *lethal(1)B214* region at the base of the X-chromosome of *Drosophila melanogaster*. Chromosoma: **101**, No. 7, 456-466.

Russell, S. and K. Kaiser (1994). A *Drosophila melanogaster* chromosome-2L repeat is expressed in the male germ-line. Chromosoma: **103**, 63-72.

Ryerse J. S. (1989). Electron microscope immunolocation of gap junctions in Drosophila. Tissue and Cell: **121**, No. 6, 835-839.

Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989). Molecular Cloning: A Laboratory Manual (Second ed.). New York: Cold Spring Harbor Laboratory Press.

Sanger, F., Nicklen, S., and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Science (U. S. A.): **74,** 5463-5467.

Schalet, A. (1972). New Mutants. *Drosophila* Information Service: **49**, 36-37.

Schalet, A. and Lefevre, G. (1973). The Localization of "Ordinary" Sex-Linked Genes in Section 20 of the Polytene X Chromosome of *Drosophila melanogaster*. Chromosoma: **44**, 183-202.

Schalet, A. and Lefevre, G. (1976). The proximal region of the X chromosome. In The Genetics and Biology of *Drosophila*, ed. Ashburner, M. and Novitski, E., Academic Press, New York. Vol. 1b, pp. 845-902

Scherer G., J. Telford, C. Baldari and V. Pirriotta (1981). Isolation of cloned genes differentially expressed in at early and late stages of *Drosophila* embryonic development. Developmental Biology: **86**, 438-447.

Scherer G., C. Tschudi, J. Pererera, H. Delius and V. Pirriotta (1982). *B104*, a new dispersed repeated gene family in *Drosophila melanogaster* and its anologies with retroviruses. Journal of Molecular Biology: **157**, 435-452.

Schmid, C. W. and P. L. Deininger (1975) Sequence organisation of the human genome. Cell: **6**, 345-358.

Schneuwly, S., A. Kuroiwa and W. J. Gehring (1987). Molecular analysis of the dominant homeotic *antennapedia* phenotype. EMBO Journal: **6**, 201-206.

Scholzen, T, Arndt, E (1991). Organization and nucleotide sequence of ten ribosomal protein genes from the region equivalent to the *spectinomycin* operon in the archaebacterium *Halobacterium marismortui*. Molecular and Gen. Genetics: **228**, No. 1-2, 70-80.

Scofield, S. R. and W. Y. Chooi (1982). Structure of ribosomes and ribosomal-subunits of *Drosophila*. Molecular and General Genetics: **187**, 37-41.

Scott, M. P., A. J. Weiner, T. I. Hazelrigg, B. A. Polisky, V. Pirrotta, F. Scalenghe and T. C. Kaufman (1983). The molecular organisation of the *antennapedia* locus of *Drosophila*. Cell, **35**: 763 - 776.

Scott, M. P. and S. B. Carroll (1987). The segmentation and homeotic gene network in early *Drosophila* development. Cell: 689-698.

Scott, M. P., J. W Tamkun, G. W. Hartzell III. (1989). The structure and function of the homeodomain. Biochimica Et Biophysica Acta: **989**, 25-48.

Sederoff, R., L. Lowenstein and H. C. Birnboim (1975). Polypyrmidine segments in *Drosophila melanogaster* DNA: II. Chromosome location and nucleotide sequence. Cell: **5**, 183-194.

Sharp, P. M., Cowe, E., Higgins, D. G., Shields, D. C., Wolfe, K. H. and Wright, F. (1988). Codon usage patterns in *Escherichia coli, Bacillus subtilis, Saccharomyces cerevisiae, Schizosaccharomyces pombe, Drosophila melanogaster* and *Homo sapiens*: a review of the considerable within-species diversity. Nucleic Acids Research: **16**, 8207-8211.

Shigesada, H. K., M. satake, Y. Ito, H. Miyoshi, M. Ohki, M. Pepling and P. Gergen (1993). The *runt* domain identifies a new family of heteromeric transcriptonal regulators. Trends in Genetics: **9**, No 10, 338-341.

Shirakawa, T., A. R. Li, M. Dubowitz, J. W. Dekker, A. E. Shaw, J. A. Faux, C. S. Ra, W. Cookson and J. M. Hopkin (1994). Association between atopy and variants of the beta-subunit of the high-affinity immunoglobulin-E receptor. Nature Genetics: **7**, No. 2, 125-130.

Short J. M, Fernandez J. M., Sorge J. A and Huse W. D. (1988). Lambda Zap: a bacteriophage lambda expression vector with *in vivo* excision properties. Nucleic Acids Research: **16**, 7583-7600.

Smith, TF, Waterman, MS (1981). Overlapping genes and information theory. Journal of Theoretical Biology, **91**, No.2, pp.379-380

Sniegowski, P. D. and B. Charlesworth (1994). Transposable element numbers in cosmopolitan inversions from a natural population of *Drosophila melanogaster*. Genetics: **137**, No. 3, 815-827.

Southern, E. M. (1975). Detection of Specific Sequences Among DNA Fragments Separated by Gel Electrophoresis. Journal of Molecular Biology: **98**, 503.

Spencer, T. A. and G. A. Mackie (1993). The nucleotide-sequence of a cloned cDNA-encoding ribosomal-protein s6 from *Drosophila melanogaster*. Biochimica Et Biophysica Acta: **1172**, 332-334.

Spierer, P., A. Spierer, W. Bender and D. S. Hogness (1983). Molecular mapping of genetic and chromomeric units in *Drosophila melanogaster*. Journal of Molecular Biology: **168**, 35-50.

Spradling, A. C., G. Karpen, R. Glaser and P. Zhang (1993). DNA elimination in *Drosophila*. In: Evolutionary conservation of developmental mechanisms: 39-53. Wiley-Liss, Inc.

Staden, R., and McLachlan, A. D. (1982). Codon preference and its use in identifying protein coding regions in long DNA sequences. Nucleic Acids Research: **10**, 141-156.

Starich, T. A., Herman, R. K., and Shaw, J. E. (1993). Molecular and genetic analysis of *unc-7*, a *Caenorhabditis elegans* gene required for coordinated locomotion. Genetics: **133**, 527-541.

Starich, T. A., Lee, R. C., Avery, L. and Shaw, J. E. (1995). *unc7* and *eat5* represent a multigene family reminiscent of vertebrate connexins. Abstract number *C. elegans* database: 1995 entry.

Steffenson, D. M., R. Appels and W. J. Peacock (1981). The distribution of two highly repeated DNA sequences within *Drosophila melanogaster* chromosomes. Chromosoma, **82**: 525-541.

Stephan, W. and S. J. Mitchell (1992). Reduced levels of DNA polymorphism and fixed between-population differences in the centromeric region of *Drosophila ananassae*. Genetics: **132**, No. 4, 1039-1045.

Steward, R (1987). *Dorsal*, an embryonic polarity gene in *Drosophila*, is homologous to the vertebrate proto-oncogene, *c-rel*. Science, **238**, No.4827, pp.692-694

Stewart, M. J. and R. Denell (1993). Mutations in the *Drosophila* gene encoding ribosomal protein-**sS6** cause tissue overgrowth. Molecular and Cellular Biology: **13**, 2524-2535.

Stewart, W. W., (1978). Functional connections between cells as revealed by dye-coupling with a highly flourescent napthalimide tracer. Cell: **14**, 741-759.

Stocker, R. F. and M. Schorderet (1981a). Cobalt filling of sensory projections from internal and external mouthparts in *Drosophila*. Cell and Tissue Research: **216**, 513-523.

Stocker, R. F. and P. A. Lawrence (1981b). Sensory projections from normal and homoeotically transformed antennae in *Drosophila*. Developmental Biology: **82**, 224-237.

Strausfeld, N. J. and U. K. Bassemir (1983). Cobalt-coupled neurons of a giant fiber system in diptera. Journal of Neurocytology: **12**, 971-991.

Strobel, E., P. Dunsmuir and G. M. Rubin (1979). Polymorphisms in the chromosomal locations of elements of the *412*, *copia* and *297* dispersed repeated gene families in *Drosophila*. Cell, **17**: 429-439.

Sudol, M., S. Grant and P. C. Maisonpierre (1993). Protooncogenes and signaling processes in neural tissues. Neurochemistry International: **22**, 369-384.

Sutcliffe JG (1979). Complete nucleotide sequence of the *Escherichia coli* plasmid pBR322. Cold Spring Harbour Symposium on Quantatative Biology: **43** Pt 1, 77-90.

Suzuki, K., M. Iwata, S. Ito, K. Matsui, A. Uchida and Y. Mizoi (1994). Molecular basis for subtypic differences of the 'a' subunit of coagulation factor XIII with description of the genesis of the subtypes. Human Genetics: **94**, No. 2, 129-135.

Sved J. (1986). Eyes Absent (*eya*). *Drosophila* Information Service: **63**, 169.

Swain, G. P., R. J. Wyman and M. D. Egger (1990). A deficiency chromosome in *Drosophila* alters neuritic projections in an identified motoneuron. Brain Research: **535**, 147-150.

Swillens, S, Cochaux, P, Lecocq, R (1989). A pitfall in the computer-aided quantitation of autoradiograms. Trends in Biochemical Sciences,.**14**, No.11, pp.440-441.

Tabor, S. and C. C. Richardson (1989). Effect of manganese ions on the incorporation of dideoxynucleotides by bacteriophage-T7 DNA-polymerase and *Escherichia-coli* DNA-polymerase-1. Proceedings of the National Academy of Science (U. S. A.): **86**, No. 11, 4076-4080.

Tanouye, M. A., and Wyman, R. J. (1980). Motor outputs of the giant nerve fiber in *Drosophila*. Journal of Neurophysiol.: **44**, 405-421.

Tanouye, M. A. and D. G. King (1983). Giant fiber activation of direct flight muscles in *Drosophila*. Journal of Experimental Biology: **105**, 241-248.

Tautz, D., R. Lehmann, H. Schnurch, R. Schuh, E. Seifert, A. Kienlin, K. Jones and H. jackle (1987). Finger protein of novel structure encoded by *hunchback*, a 2nd member of the gap class of *Drosophila* segmentation genes. Nature: **327**, 383-389.

Temin, R. G., B. Ganetzky, P. A. Powers, T. W. Lyttle, S. Pimpinelli, P. Dimitri, C. I. Wu and Y. hiraizumi (1991). Segregation distortion in *Drosophila melanogaster* - genetic and molecular analyses. American Naturalist: **137**, 287-331.

Tepass, U. and V. Hartenstein (1994). The development of cellular junctions in the *Drosophila* embryo. Developmental Biology: **161**, 563-596.

Thisse, B., C. Stoetzel, C. Gorostizathisse and F. Perrinschmitt (1988). Sequence of the *twist* gene and nuclear-localization of its protein in endomesodermal cells of early *Drosophila* embryos. Embo Journal: 7, 2175-2183.

Trimarchi_JR, Schneiderman_AM (1994). The motor-neurones innervating the direct flight muscles of *Drosophila melanogaster* are morphologically specialised. Journal of Comparative Neurology, **340**, No.3, pp.427- 443.

Trimarchi_JR, Schneiderman_AM (1995a). Initiation of flight in the unrestrained fly, *Drosophila melanogaster*. Journal of Zoology, **235**, No.Pt. 2, pp.211-222

Trimarchi_JR, Schneiderman_AM (1995b). Flight initiations in *Drosophila melanogaster* are mediated by several distinct motor patterns. Journal of Comparative Physiology A-Sensory Neural and Behavioural Physiology, **176**, No.3, pp.355-364.

Trimarchi_JR, Schneiderman_AM (1995c). Different neural pathways co-ordinate *Drosophila* flight initiations evoked by visual and olfactory stimuli. Journal of experimental biology, **198**, No.5, pp.1099-1104.

Thomas, J. B. and R. J. Wyman (1983). Normal and mutant connectivity between identified neurons in *Drosophila*. Trends In Neuroscience: **6**, 214-219.

Thomas, J. B., and Wyman, R. J. (1982). A mutation in *Drosophila* alters connectivity between two identified neurones. Nature: **298**, 650-651.

Thomas, J. B., M. J. Bastiani, M. Bate and C. S. Goodman (1984). From grasshopper to *Drosophila* - a common plan for neuronal development. Nature: **310**, 203-207.

Thomas, J. B. and R. J. Wyman (1984a). Duplicated neural structure in *bithorax* mutant *Drosophila*. Developmental Biology: **102**, 531-533.

Thomas, J. B. and R. J. Wyman (1984b). Mutations altering synaptic connectivity between identified neurons in *Drosophila*. Journal of Neuroscience: **4**, 530-538.

Thomas, P. S. (1980). Hybridisation of denatured RNA and small DNA fragments transferred to nitrocellulose. Proceedings of the National Academy of Science (U. S. A.): **77**, No 9, 5201-5205.

Turner, S. H. and C. D. Laird (1973). Diversity of RNA sequences in *Drosophila melanogaster*. Biochemistry and Genetics: **10**, 263-274.

Ullrich A., J. Shine, J Chirgwin, R. Pictet, E. Tischer, W.J. Rutter and H.M. Goodman (1977). Isolation of cytoplasmic R.N.A.. Science, **196**: 1313.

Vander, A. J., JH Sherman and DS Luciano (1994) Human physiology, 6th International Edition. McGraw-Hill Publishers.

Van, Der Bliek, AM, Meyerowitz, EM (1991). *Dynamin*-like protein encoded by the *Drosophila shibire* gene associated with vesicular traffic. Nature: **351**, No. 6325, 411-414.

Vaslet, C. A., P. OConnell, M. Izquierdo and M. Rosbash (1980). Isolation and mapping of a cloned ribosomal protein gene of *Drosophila melanogaster*. Nature: **285**, 674-676.

Von Henje, G. (1986). A new method for predicting signal sequence cleavage sites. Nucleic Acids Research: **14**, 4683-4691

Waksman, G., D. Kominos, S. C. Robertson, N. Pant, D. Baltimore, R. B. Birge, D. cowburn, H. Hanafusa, B. J. Mayer, M. Overduin, M. D. Resh, C. B. Rios, L. Silverman and J. kuriyan (1992). Crystal-structure of the phosphotyrosine recognition domain *sh2* of *v-src* complexed with tyrosine-phosphorylated peptides. Nature: **358**, 646-653.

Watanabe, T. and D. R. Kankel (1990). Molecular-cloning and analysis of *l(1)ogre*, a locus of *Drosophila melanogaster* with prominent effects on the postembryonic development of the central-nervous-system. Genetics: **126**, 1033-1044.

Waye, J. S. and H. F. Willard (1985). Chromosome-specific alpha-satellite DNA - nucleotide-sequence analysis of the 2. 0 kilobasepair repeat from the human X-chromosome. Nucleic Acids Research: 2731-2743.

Weber, F., J. Devilliers and W. Schaffner (1984). An SV40 enhancer trap incorporates exogenous enhancers or generates enhancers from its own sequences. Cell: **36**, 983-992.

Wensink, P. C., D. J. Finnegan, J. E. Donelson and D. S. Hogness (1974). A system for mapping DNA sequences in the chromosomes of *Drosophila melanogaster*. Cell: **3**, 315-325.

Wensink, P. C. (1978). Sequence homology within families of *Drosophila melanogaster* middle repetitive DNA. Cold Spring Harbour Symposium on Quantatative Biology: **42**, 1033-1039.

Wensink, P. C., S. Tabata and C. Pachl (1979). The clustered and scrambled arrangement of moderately repetitive elements in *Drosophila* DNA. Cell: **18**, 1231-1246.

Wharton, K. A., B. Yedvobnick, V. G. Finnerty and S. Artavanistsakonas (1985). *Opa* - a novel family of transcribed repeats shared by the *Notch* locus and other developmentally regulated loci in *Drosophila melanogaster*. Cell: **40**, 55-62.

Whittaker P. A., Cambell A. J. B., Southern E. M. and Murray N. E (1988). Enhanced recovery and restriction mapping of DNA fragments cloned in a new lambda vector. Nucleic Acids Research: **16**, 6725-6736

Will B. M., A. A. Bayev and D. F. Finnegan (1981). Nucleotide sequence of terminal repeats of *412* transposable elements of *Drosophila melanogaster*. Journal of Molecular Biology: **153**, 897-915.

Wigboldus, J. D. (1987). cCDNA and deduced amino-acid sequence of *Drosophila rp21C*, another a-type ribosomal-protein. Nucleic Acids Research: **15**, 10064.

Wilson, D. M., W. A. Deutsch and M. R. Kelley (1993). Cloning of the *Drosophila* ribosomal protein-*S3* - another multifunctional ribosomal-protein with an pendonuclease DNA-repair activity. Nucleic Acids Research: **21**, 2516.

Wyman, R. J. and J. B. Thomas (1983). What genes are necessary to make an identified synapse? Cold Spring Harbour Symposium on Quantatative Biology: **48**, 641-652. .

Wyman, R. J., Thomas, J. B., Salkoff, L., and Costello, W. (1985). The *Drosophila* thorax as a model system for neurogenetics. In A. I. Selverston (Eds.), Model Neural Networks and Behaviour New York: Plenum Press.

Yamamoto, M. -T. and Miklos, G. (1987). Cytological Analysis of Deficiency 16-3-35 at the base of the X chromosome of *Drosophila melanogaster*. *Drosophila* Information Service: **66**, 154.

Yamamoto, M. T., A. Mitchelson, M. Tudor, K. Ohare, J. A. Davies and G. Miklos (1990). Molecular and cytogenetic analysis of the heterochromatin-euchromatin junction region of the *Drosophila melanogaster* X-chromosome using cloned DNA-sequences. Genetics: **125**, 821-832.

Yanisch-Perron C., Vieira J. and Messing J. (1985). Improved M13 cloning vectors and host strains: nucleotide sequence of M13mp18 and pUC19 vectors. Gene: **33**, 103-119.

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

Young, M. W. (1979). Middle repetitive DNA: a fluid component of the *Drosophila* genome. Proceedings of the National Academy of Science (U. S. A.): **76**, 6274-6278.

Young, M. W. and H. E. Schwartz (1981). Nomadic gene families in *Drosophila*. Cold Spring Harb. Symp. Quant. Biology: **45**, 629-640.

Young, B. S., A. Pession, K. L. Traverse, C. French and M. L. Pardue (1983). Telomeric regions in *Drosophila* share complex DNA sequences with pericentric heterochromatin. Cell, **34**: 85-94.

Zachar, Z. and P. M. Bingham (1989). Suppressible insertion-induced mutations in *Drosophila*. . Progress In Nucleic Acids Research and Molecular Biology: 87-98.

Zelentsova, E. S., Vashakidze, R. P., Krayev, A. S., and Evgenev, M. B. (1986). Dispersed repeats in *Drosophila virilis*: Elements mobilised by interspecific hybridisation. Chromosoma, **93**, 469-476.

Zimmerman, J. L., Fouts, D. L., Levy, L. S., and Manning, J. E. (1982). Non-adenylated mRNA is present as polyadenylated RNA in nuclei of *Drosophila*. Proceedings of the National Academy of Science (U. S. A.): **79**, 3148-3152.

Zusman S., D. Coulter, J.P. Geregen (1985). Lethal mutations induced in the proximal X chromosome of *Drosophila melanogaster*. *Drosophila* Information Service, **61**: 217-218.

Zusman, S. B. and E. F. Wieschaus (1985). Requirements for zygotic gene activity during gastrulation in *Drosophila melanogaster*. Developmental Biology: **111**, 359-371.