INSERTION-DELETION VARIATION IN THE DNA OF THREE NATURAL POPULATIONS OF *DROSOPHILA MELANOGASTER*

ROBIN NICHOLAS BEECH
PH.D.
UNIVERSITY OF EDINBURGH
1987

I would like to thank the following:

Andy, for most of the proof reading and being relatively patient.
Tony, for being the devil's advocate.
James, for proof reading and putting in the semicolons.
Jane, for moral support and a great deal of practical advice.
Debbie, for all the washing up and the fly food.
Jools, for the fly work and stock keeping.
Curt Strobeck, for the incentive to finish.
Trudy, for encouragement with the statistics.
Frank, without whom I might not have gone near the computer.
Juan Modellel, for supplying the clones for this project.

ABSTRACT

Variation due to insertion-deletion events in the DNA of three different population samples of *Drosophila melanogaster* has been examined using restriction fragment analysis. One of the samples was surveyed at the third chromosomal 87A7 heat shock locus and the other two at the X chromosomal *achaete-scute* locus. The frequency and distribution of insertions found at the heat shock locus was very similar to that observed in a previous study of this region (Leigh Brown 1983) and also to a survey of a second chromosomal region - the *Adh* locus (Aquadro *et al.* 1986). The X chromosomal locus differed significantly from the autosomal loci in the number of insertions found, having fewer insertions than expected. Such a difference between the X chromosome and the autosomes is expected if insertions induce generally recessive deleterious mutations. This does not seem to be the cause here but rather the lower number was due to an apparent non-uniformity in the distribution of insertions. Natural selection is not likely to be responsible for the observed distribution of insertions since they have been found to lie close to known transcripts where their deleterious effects on the phenotype are expected to be greatest. An explanation for the observed distribution could be that DNA is inserted preferentially close to transcription regions where the chromatin structure is known to be more accessible to external agents such as enzymes which interact with DNA, as presumably the proteins involved in the insertion process do.

TABLE OF CONTENTS

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

The phenotypic differences between the organisms of different species derive in part from the genetic differences encoded in their genomes. Understanding the process by which a species may diverge and eventually give rise to genetically isolated populations, the evolutionary process, requires a knowledge of the nature and behaviour of the underlying genetic constitution of a species. In order for this process to occur at all genetic variation must exist between individuals. The measurement and categorization of this variation has given rise to the discipline of population genetics.

The 'struggle to measure variation' is an old one and has undergone a revolution of scale within the last thirty years (reviewed by Lewontin 1974). Genetic observation of many organisms has revealed that there is a vast reservoir of variability within species bringing the concept of a 'wild-type' into doubt. The persistence of this variability in natural populations has been the source of much heated argument. The view propounded by Muller (1950) was that the great majority of the genome was homogeneously 'purified' by selection and was therefore expected to be almost entirely homozygous for 'wild-type' alleles. Occasionally variant forms would arise by mutation and, if deleterious in effect as a large fraction of mutations are, would persist for only a short time before being lost from the population. Mutations which could give rise to an advantage for an individual over contemporary members of the population would increase in frequency, becoming transiently polymorphic, and ultimately all members of the population would possess the new form.

This 'classical' view was opposed by the 'balance' hypothesis which supposed that the proportion of loci which segregate for alternative alleles could be very large, possibly almost all loci. The maintenance of such a degree of variation was supposed to occur by a process of 'balanced' selection for two or more alleles at certain loci which would maintain both genetic forms in the population. The significance of purifying selection under such a hypothesis must be smaller than supposed under the 'classical' view. The incorporation of advantageous mutations in a population was still assumed to be the mechanism by which populations evolved differences. The discovery of a vast wealth of polymorphism at the protein and DNA level (eg Lewontin and Hubby

1966; Harris 1966; Avise *et al* 1979; Jeffreys 1979; Langley *et al.* 1982) has shown that the strict 'classical' view is incorrect. It may indeed be true that the majority of loci in natural population do segregate for alternative alleles.

The classical-balance controversy has now been replaced, however, by a new debate about the nature of the differences which evolve between organisms. One view retains the belief that the substitution of one allele for another occurs by a systematic advantage of the new form over the older. Opponents to this 'selectionist' argument claim that the substitutional load involved during replacement at the rate observed for protein evolution is far larger than could be tolerated in most populations (Kimura 1968a). In two papers (Kimura 1968b; King and Jukes 1969) the proposal was made that the great majority of allelic substitutions which occur in nature do not affect the differential survival of individuals in a population. This neutralist view has received much attention in recent years (Kimura 1983).

1.1. Population Genetics: The Measurement of Variation

Phenotypic variation between individuals of a population is due to the interaction between the genotype of individuals and the environment in which they find themselves. In order to determine what part the genotype plays in the determination of phenotypic variation, genetic analysis of variants discovered is necessary. For this reason *Drosophila* has been the organism of choice since the understanding of the genetics of *Drosophila* far exceeds that for any other higher eukaryote. Variation was first systematically studied in *Drosophila* by surveying natural populations for variants analogous to the classical mutants observed in this organism. Approximately 2% of wild flies displayed phenotypes similar to these (eg Dubinin 1937, cited in Lewontin 1974). Genetic analysis of these revealed that only the order of 1% of wild flies were variant for identifiable genetic alleles. A more systematic approach was afforded by the study of lethal genes. Chromosomes extracted from natural populations of *Drosophila* contain on average one lethal mutation per genome (Dobzhansky and Spassky 1954). This represents about twice the variation found with visible variants.

Alleles which are not lethal may, however, severely reduce the viability, fertility or other components of fitness of an organism. The average viability of chromosome homozygotes in *Drosophila pseudoobscura* (Dobzhansky 1963)

was 76.06% of the average of heterozygotes. It is not clear from this experiment how many loci may be responsible for this reduction, and so be segregating for such alleles (Lewontin 1974). The general pattern of the distribution of homozygous fertility and viability remains the same over all *Drosophila* experiments conducted so far (Lewontin 1974; Tracey and Ayala 1974; Mackay 1986b). Chromosomes fall either at, or close to, lethality or have effects which produce a fitness close to that of heterozygotes.

The main criticism of the above methods is that they only provide information about a limited fraction of the genome. Variation of this type may, however, represent the significant fraction which is important in evolutionary terms. Analysis of the genetic basis for the isolation of *Drosophila* species is possible because some fertile hybrids between different species are sometimes produced. It has been found that alleles at relatively few loci are responsible for the observed genetic barriers between species (eg Coyne 1983, 1984, 1985; Coyne and Kreitman 1986).

The systematic use of starch gel electrophoresis in the detection of protein electrophoretic mobility polymorphisms (Lewontin and Hubby 1966; Harris 1966) opened up a new field of molecular population biology. The replacement of one amino acid with another in a protein, as a result of a change in the DNA, may produce an alteration in the mobility of a protein which is detectable. These first two surveys in *Drosophila* and man and many subsequent surveys (Lewontin 1974) have all revealed high frequencies of polymorphic loci. About one third of the enzyme loci surveyed were found to be segregating for electrophoretic variants. Allowing for synonymous changes in the genes themselves, which produce no observable change almost all loci were estimated to be segregating for alternative alleles.

Even though far more loci are potentially open to analysis by this method than were previously, only a proportion of the less than 5% of the genome, present as structural genes, are represented (Davidson *et al.* 1973; Spradling and Rubin 1981). The analysis of any portion of the entire genome has recently become possible with the advent of recombinant DNA technology. Cleavage of DNA with enzymes which recognise specific sequences produce DNA fragments which may be used to determine alterations in the DNA sequence directly. The ease of purification of mitochondrial DNA provided a

simple means for the first use of such techniques to analyse genetic variability and population structure (Avise *et al.* 1979). Following from this the use of cloned segments of genomic DNA in the detection of specific genomic DNA fragments (Southern 1975) has allowed the investigation of genetic variability in the nuclear genome itself (eg Jeffreys 1979; Langley *et al.* 1982; Kreitman 1983; Leigh Brown 1983; Aquadro *et al.* 1986). The variation discovered in this way is such that two individuals from, for example, a *Drosophila* population will differ from each other for about 0.5% of the DNA sequence on average.

1.2. Molecular Biology: The Structure of Genomes

The genomic DNA of higher eukaryotes contain sequences which vary in their degree of repetition in the genome. The rate of reassociation of denatured fragments of genomic DNA in solution provides a measure of the complexity and the degree of sequence repetition in the genome (Wetmuir and Davidson 1968; Waring and Britten 1966). Analysis of the reassociation kinetics of total genomic DNA has shown that there are three major components of the higher eukaryotic genome, classified by their degree of repetition (Wetmuir and Davidson 1968). This contrasts with the genomes of bacterial viruses such as T2 (Rubenstein *et al.* 1961) and of *E. coli* (Cairns 1963) which appear to be relatively simple by comparison (Britten and Kohne 1968). The different classes of DNA categorized in this way appear to play very different roles in the genome.

Satellite DNA consists of thousands of tandem duplications of a given sequence ranging from a few, up to a few hundred nucleotides (Gall and Atherton 1974; Southern 1970). These sequences are located in blocks within heterochromatin (Pardue and Gall 1970; Yunis and Yasmineh 1970) and are usually localized near the centromere or telomeres (Hennig *et al.* 1970; Peacock *et al.* 1973). Satellite DNA can vary between quite closely related species. The differences are of both total amount of satellite DNA and the sequences from which the satellite is composed (Peacock *et al.* 1977; Fry and Salser 1977; Gosden *et al.* 1977; Barnes *et al.* 1978; Brutlag 1980).

Although heterochromatin appears genetically inert (Gerhenson 1933) specific functions such as chromosome pairing and segregation (Cooper 1964), recombination suppression (Miklos and Nankivell 1976) and speciation (White 1978) have been attributed to it. The role of satellite sequences in these

functions is unclear since they may not be responsible for the observed effects, only associated with them. Charlesworth *et al.* (1986) and Stephan (1986) have claimed rather that satellite sequences may have accumulated in centric and telomeric heterochromatin simply because of the genetic properties of these regions of the chromosomes. In these regions recombination is suppressed and it is under such conditions that satellite sequences are likely to persist the longest.

About one quarter of moderately repetitive DNA is made up of tandem repeats of ribosomal RNA genes (Ritossa and Spiegelman 1965; Artavanis-Tsakonas *et al.* 1977) and histone genes (Lifton *et al.* 1978). The remaining three quarters of this DNA class are dispersed differently to satellite DNA. Rather than existing as tandem duplications these sequences are usually dispersed throughout the genome with only a few, and probably single, copies of a sequence at each location (Davidson *et al.* 1973; Manning *et al.* 1975). Organisms which have the *Xenopus* like genome structure have thousands of copies of each sequence, about 300 nucleotides long, separated by about 800 nucleotides of non-repeated DNA (Davidson *et al.* 1973). The number of such repeats compared to the number of genes in organisms, and their pattern of dispersal, has led some to the view that they may serve to regulate the expression of genes (Britten and Davidson 1969). Others believe that this portion of the genome may be insignificant or even detrimental and is present purely because the sequences are efficient at promoting their own replication within genomes (Doolittle and Sapienza 1980; Orgel and Crick 1980).

The short repeated families of higher eukaryotes or SINE's (Singer 1982), such as the *Alu* family appear to be mobile (Schmid and Jelink 1982). These elements may move by reverse transcription of RNA molecules generated by RNA polymerase III (Haynes *et al.* 1981; Jagadeeswaran *et al.* 1981). The RNA polymerase III promoter is contained within such RNA molecules and so new copies could presumably retain their mobility. For the *Alu* family itself it seems more likely that the sequence is derived from part of the 7SLRNA molecule (Walter and Blobel 1982; Ullu and Tschudi 1984) and has simply accumulated over the entire genome as a processed copy of this (Leigh Brown 1984). The sites of integration of *Alu* are remarkably stable, persisting for long evolutionary periods (eg Maeda *et al.* 1983; Hobbs *et al.* 1985). During the time period over which individual sites have been found to be stable the rapid

accumulation, or loss, of many hundreds of copies of the *Alu* sequence has also occurred (Ru Hwu *et al.* 1986).

A class of rather larger elements is also present in the *Xenopus*-like pattern. Sequences similar to the *Kpn* I family or LINE's (Singer 1982; Schafit-Zagardo *et al.* 1982) are also repeated thousands of times within the genome but are in general larger than about 4 kb. The mobility of these sequences may be greater than that of SINE's (Economou-Pachnis *et al.* 1985; Musich and Dykes 1986).

In *Drosophila* the SINE component of the genome appears to be missing or to be only a minor component. The majority of moderately repetitive DNA is present as relatively large elements, about 5 kb long (Manning *et al.* 1975). These elements are extremely mobile compared to the SINE fraction of other genomes. The position of these elements varies between strains of *Drosophila* and even between parent flies and derived tissue culture cells (Strobel *et al.* 1979; Potter *et al.* 1979). Individual flies from the same population may share no common sites of integration for specific members of a moderately repetitive family (Montgomery and Langley 1983; Leigh Brown and Moss 1987).

The third component of the genome probably contains a majority of the genetic complexity of organisms (Davidson and Hough 1971). The sequences in this fraction occur only once or a few times each in the genome and include the genes encoding the structural proteins. These genes are associated with the euchromatin and it may be that a single band of euchromatin is associated with a single gene (Zhimulev *et al.* 1981; Spierer and Spierer 1983; Bossy *et al.* 1984). It is this fraction which, until the advent of DNA technology, was open to observation of genetic variability by examination of the protein products which it encoded.

1.3. Molecular Population Genetics

The biological and evolutionary interplay between the various components of the genome is certainly complex. Understanding the effects of each in the laboratory, and the dynamic aspects revealed by the examination of natural populations, is essential to the understanding of evolution. Only recently has the development of molecular biological techniques been such that it has become possible to extrapolate from the experimental findings of a

survey to a prediction of population dynamics at the DNA level. Analysis at the molecular level is often tedious because of the work involved in gaining information of a small portion of the genome. The rapid introduction of new techniques in DNA technology from other disciplines is making the task easier.

Detailed analysis of observations made using techniques of low resolving power provides necessary insight into the exact causes of the more gross manifestations of genetic changes. The mechanisms by which genetic variation produces effects on the phenotype and hence how certain sequences may be important in evolutionary terms is becoming possible.

1.4. Background for the Survey

Drosophila has proved to be an ideal organism in which to study population genetics. Flies are easily maintained on simple media and have a generation period of two weeks or so in the laboratory. Ever since the discovery of the first genetic mutation in *Drosophila melanogaster* - *white* (Morgan 1910) the use of mutations has led to a great understanding of the genetics of this species. Various chromosomal rearrangements have been produced which allow genetic analysis of a degree which is impossible with other species of higher organism (Lindsley and Grell 1967).

The genome of *Drosophila melanogaster* is small compared to organisms such as man (Rasch *et al.* 1971) and is divided among four chromosomes, although only a few known functions have been attributed to the fourth (Lindsley and Grell 1967). This has allowed the rapid localisation of the chromosomal regions responsible for various minor phenotypic effects (Shrimpton and Robertson 1987a, 1987b) and the genetic analysis of the natural variation in fitness found in wild populations (Reviewed by Lewontin 1974).

Drosophila melanogaster is a cosmopolitan species with an extremely large population size (Mukai and Yamaguchi 1974; Kreitman 1983). The ease of analysis afforded by this species in the field of population genetics has allowed the discoveries made to be applied to that in most other organisms. This species has taken on the role of a model system in which experimental evidence can be used to make wide ranging predictions about the general nature of evolution. The thorough understanding of the population genetics of

Drosophila melanogaster is essential if predictions are to be made for other systems. One must also be careful if discrepancies exist between the model and the organism of study. The recent knowledge obtained on the moderately repetitive component of the *Drosophila melanogaster* genome suggests that this component may behave in a quantitatively, but not a qualitatively, different way to most other species. It is important therefore to understand the possible effects which such a difference may make in population genetic terms so that the model can be applied more accurately to other situations.

1.5. Transposable Elements: a New Process in Population Biology

In situ hybridization of cloned members of repetitive DNA families in *Drosophila* has shown that these sequences are extremely mobile (Wensink *et al.* 1974; Potter *et al.* 1979, 1980; Strobel *et al.* 1979; Finnegan *et al* 1978; Ilyin *et al.* 1978; Young 1979; Montgomery and Langley 1983; Leigh Brown and Moss 1986). These elements have been shown to be responsible for many deleterious mutations (Bender *et al.* 1983; Carramolino *et al* 1983; Campuzano *et al.* 1986; Parkhurst and Corces 1986; Levis *et al.* 1984; Pirrotta and Brockl 1984; Mackay 1985, 1986a; Fitzpatrick and Sved 1986; Yukohiro *et al.* 1985). The presence of these potentially harmful sequences has prompted the proposal of a new process of evolution - the evolution of sequence elements within genomes where individual sequences rather than the organism as a whole are the unit on which the forces of evolution act (Doolittle and Sapienza 1980; Orgel and Crick 1980). These sequences may promote their own replication within genomes, despite harmful effects on the host, and so become numerous both within the genome and throughout the gene pool of sexually reproducing populations. This process does not appear to continue indefinitely, since the numbers of mobile elements within genomes appears to be limited. Population models have been proposed in which the limitation of numbers by a reduction of host fitness (Charlesworth and Charlesworth 1983; Brookfield 1982) or by a process of self regulation (Charlesworth and Charlesworth 1983; Langley *et al.* 1983; Charlesworth and Langley 1986) is presumed to occur.

Although it was known that middle repetitive DNA in *Drosophila* was dispersed throughout the genome the first observations of these sequences in the neighbourhood of genes caused some surprise (Langley *et al.* 1982; Leigh Brown 1983). These insertions were responsible for a large part of the variability present at the loci studied. More recent analysis (Aquadro *et al.* 1986) has suggested this to be a general phenomenon in *Drosophila*. This type of variation is not restricted to *Drosophila* and has been reported in organisms such as the rat (Economou-Pachnis *et al.* 1985) and humans (Hobbs *et al.* 1985). It would seem prudent to try to understand how such variation is generated, how it persists and the long term effects of insertions in evolutionary terms.

1.6. Intentions of the Survey

The survey presented here was intended to provide information about the variability found at certain regions of the genome with an emphasis on insertion-deletion variation. DNA samples representing single chromosomes from natural populations of *Drosophila melanogaster* were examined by restriction fragment analysis. Fragments from a specific region of the genome were visualized by hybridization to unique sequence DNA which had been cloned previously from *Drosophila melanogaster*. The accumulation of information on the phenotypic effects of insertions (Mackay 1985, 1986a; Yukohiro *et al.* 1985; Fitzpatrick and Sved 1986) and their chromosomal distribution (Montgomery and Langley 1983; Leigh Brown and Moss 1987; Biemont 1986; Ronserray and Anxolabehere 1986; Ajioka and Eanes 1987) requires the description of insertion patterns at the DNA level before a full interpretation can be made.

Preliminary observations of DNA insertions at the molecular level in natural populations (Langley *et al.* 1982; Leigh Brown 1983) indicate that insertion-deletion variation may contribute a substantial fraction of the variation in natural populations. The generality of this situation over different populations, different sections of the genome and different combinations of population parameters needs investigation.

Previous studies have used samples of about 30 chromosomes from wild populations (eg Leigh Brown 1983; Aquadro *et al.* 1986). In order to compare results between populations, samples of at least this order of size need to be

examined. The choice of which regions to examine must be made with the potential implications of the results in mind. Two models of the population dynamics of mobile elements differ in the process of limitation of copy number which is assumed to predominate. To distinguish between the two it would be helpful to examine a region where the effects of selection are expected to be relatively high.

The X chromosome is an obvious choice for several reasons. Firstly the X chromosome in *Drosophila* is hemizygous in males, and therefore is expected to be under more intense selection than the autosomes. Autosomal alleles may be heterozygous if the population is polymorphic. The proportion of these on which selection has any effect is decreased by the degree of recessivity of the alleles in that heterozygous combination. In the effectively haploid state of the X chromosome in males, dominance is complete. The proportion of alleles affected by selection is therefore proportionally greater. The greatest difference between autosomal and X-linked alleles is expected to occur when they have a low degree of dominance.

Secondly the X chromosome is very easily manipulated. In order to examine the DNA of chromosomes in an unambiguous fashion it must be possible to reveal sequences which are from specific chromosomes. For the third chromosomes for example, as explained in the materials and methods section, several crosses must be performed before flies of the desired genotype are produced. These flies contain a wild chromosome heterozygous with a chromosome deleted for the region of interest. This ensures that the sequences examined all derive from the wild type chromosome. In the final cross the desired phenotype appears in about a quarter of the flies. The phenotypic markers of this class needed careful examination to be identified correctly.

The process of extraction of the X chromosome is relatively simple by comparison. Single males are mated to females carrying an attached X chromosome. All the male offspring receive their X chromosome from their fathers (Lindsley and Grell 1967). Since the attached X chromosome carries a mutation which produces a yellow bodied phenotype, the males can be identified easily. This only requires a single cross to be carried out. DNA produced from flies such as these contain X chromosomal DNA only from the

wild-derived chromosome. This DNA can therefore be used to study other X-linked loci as well.

1.7. The Regions Studied

As a preliminary study the heat shock region at the 87A7 locus on the third chromosome was chosen. This region had been examined previously (Leigh Brown 1983) and a similar study in a second population would be useful to establish if the variation found there was general between populations. The extent of DNA it was possible to examine with the probes available was about 25 kb. Within this region there are two known transcripts in a divergent orientation separated by about 1 kb. It is not known whether there are any other regions of transcription within this distance so the potential effects of any insertions found would be unknown.

A much larger region was chosen for study on the X chromosome. λ phage clones covering 110 kb at the *achaete-scute* complex (AS-C) were kindly provided by Dr J. Modellel. The length of this series of clones is important since it is possible to examine variation over a much larger contiguous section of the genome than was possible before. The pattern of transcription over this region has received much attention recently (Carramolino *et al.* 1983; Campuzano *et al.* 1985; Parkhurst and Corces 1986; Biessmann 1985). The analysis of transcriptional activity has clarified the genetic analysis at the AS-C locus, which is extensive.

Five phenotypic characters have been genetically mapped within this region extending from *yellow* through *achaete*, *scute*-α, *lethal of scute* to *scute*-β and mutations of the *hairy-wing* phenotype (Campuzano *et al.* 1985; Lindsley and Grell 1967). Mutations of the *yellow* locus produce flies with yellow cuticular structures (Lindsley and Grell 1967; Parkhurst and Corces 1986). *achaete* mutations affect the microchaetae number and distribution on the fly (Lindsley and Grell 1967; Garcia-Bellido 1979). The *scute* mutations may be subdivided into three categories. Both *scute*-α and *scute*-β mutations remove macrochaetae from the head and thorax but the *scute*-α mutations tend to be more extreme, removing more bristles (Lindsley and Grell 1967; Dubinin 1933; Garcia-Bellido 1979). The *lethal of scute* mutations are all chromosomal rearrangements with one breakpoint within the AS-C region which also affect *scute* functions (Garcia-Bellido 1979). Although no available

lethal point mutations have been shown to be allelic to lethal of *scute* it does appear to be a real entity (Garcia-Bellido 1979; Campuzano *et al.* 1985).

The dominant mutations of *hairy-wing* phenotype have been genetically mapped to the AS-C region (Demerec and Hoover 1939). Although these mutations have been classified as distinct from mutations at *achaete* or *scute* it has been suggested (Demerec and Hoover 1939; Garcia-Bellido 1979) that there is no distinct *hairy-wing* locus. This in fact appears to be the case, with lesions determining the mutant phenotype residing in the transcripts presumed to encode *achaete* or *scute*-α functions (Campuzano *et al.* 1985, 1986).

The genetic regulation of this region appears complex and may occur by some process of negative control (Falk 1963) by the products of loci such as *hairy* and *extramacrochaetae* (Moscoso del Prado and Garcia-Bellido 1984a, 1984b). It has also been suggested that mutant lesions within the complex have effects dependent on their position within the complex. A gradation of effects has been claimed for mutant lesions centromeric to the putative *scute*-α transcript. Lesions closer to the gene appear to produce a greater loss of macrochaetae.

Molecular analysis of the *yellow* locus has demonstrated that many transcripts are produced from this region (Campuzano *et al.* 1985; Parkhurst and Corces 1986; Biessmann 1985). The regulation of this cluster of transcripts is little understood, although interference of transcription of the *yellow* locus in the y^2 allele by the transposable element *gypsy* inserted close to the regulatory sequences of the *yellow* gene seems to occur by some form of steric hindrance when *gypsy* is being actively transcribed. This may be due to some form of competition for transcription between the two. The functions of many of the transcripts produced from this region is unknown (Parkhurst and Corces 1986).

To examine the various wild-derived chromosomes DNA was prepared from a few hundred flies from each line (see Chapter 2). The DNA from each was then digested with a restriction enzyme which would cut the DNA at specific points which had been mapped previously for the HS and AS-C regions. These samples were then loaded into adjacent sample wells of agarose gels and electrophoresis carried out (see Chapter 2). Under such a direct current fragments of different sizes migrate at different rates through

the gel matrix. The mobility has been found to be approximately proportional to the $\log_{10}$ molecular weight (Bishop *et al.* 1967; Peacock and Dingman 1968). Thus smaller fragments migrate faster through the gel than larger ones. Following electrophoresis the DNA was transferred and fixed to a membrane filter. The position of any one specific sequence from the whole genome was then found by hybridizing this total DNA to a radioactive DNA sequence homologous to the one of interest (Southern 1975). The DNA was prepared from flies such that the probes used would only be homologous to DNA from the wild-derived chromosome.

In this way the fragments produced by a given restriction enzyme homologous to the same probe were identified in the different DNA samples. Variability between individual wild-derived chromosomes would be manifest as an alteration in the size, and therefore the mobility of the fragment, or fragments, homologous to the probe. There are two ways in which such a size change could occur: either a single base of the DNA at a restriction site could have been replaced by another, or a gain or loss of DNA from the region encompassed by the fragment examined may have occurred. By examination of the pattern produced, the size of fragments homologous to the probe and comparisons between different DNA samples and a molecular map of the region, the nature of the change could be established.

Most restriction endonucleases which have been analysed in detail recognise a specific short palindromic DNA sequence and produce a double stranded break at, or close to, that sequence (Kelly and Smith 1970; Hedgepeth *et al.* 1972). These enzymes have been isolated from many different bacteria where they are involved in protection from the invading DNA of bacteriophage (Arber and Linn 1969; Boyer 1971). The enzymes used in the work presented here all recognise a hexanucleotide sequence. Nucleotide substitution within the recognition sequence of an enzyme will prevent cleavage and result in a larger fragment produced by that enzyme. The creation of a site, from a sequence which differs by a single nucleotide from the recognition sequence, will give rise to a smaller fragment, or two fragments if the site is generated within the region of homology with the probe.

Either of these changes, if due to a nucleotide substitution, will only affect fragments produced by a single restriction enzyme. Two substitutions

would be required to alter the pattern produced by two enzymes recognizing different sequences. In such a case the length of the restriction map of the other lines should equal that of the altered line. If this is not the case then DNA may have been either gained or lost from the region in question. If this has occurred then the pattern of fragments produced with any enzyme should be altered since the change is not due to a specific sequence recognized by one of the enzymes.

By comparing the fragments produced in different lanes, differences between the wild-derived chromosomes may be determined. Variation due to insertion-deletion events has only recently been observed at the DNA sequence level (Langley *et al.* 1982; Leigh Brown 1983; Aquadro *et al.* 1986). To understand the population dynamics of this type of event more clearly it is desirable that the nature of the inserted DNA be determined. This is most easily done by cloning the inserted DNA. Once the DNA has been cloned, further analysis can determine if the inserted sequence was, for example, repetitive.

By comparing a random sample of chromosomes at the heat shock or AS-C loci some inferences can be made about the processes occurring in wild populations. Comparison of the quantity and quality of the variation observed on the X chromosome in contrast to that found on the autosomes should indicate whether the action of selection influences the distribution of the variants found. The examination of a long contiguous region, 120 kb at the AS-C locus, is valuable because previously only much shorter regions have been surveyed. This region has also been characterized more fully for the presence of transcripts and genetically identifiable loci than other regions screened previously. The distribution of variation both between individuals and along the genome may also be investigated.

CHAPTER 2

MATERIALS AND METHODS

The compositions of solutions and media are detailed in Appendix II

2.1. Bacterial stocks

Standard sterile technique was employed with all bacterial and bacteriophage work.

Single colonies of bacteria were grown overnight at 37 °C on suitable agar by streaking bacteria from a stock onto the surface of the plate. Liquid cultures were prepared by inoculating 10 ml of suitable broth with a single colony from a plate culture and incubating with shaking at 37 °C. Larger cultures were usually obtained by inoculating 100 ml - 500 ml of broth with 5 ml of a small scale liquid culture and incubating as before. Stocks of each strain were prepared by centrifugation of a 10 ml liquid culture at 5 000 g for 10 min. and resuspending the bacterial pellet in 1 ml of L-broth. 1 ml of sterile glycerol was added to this, mixed thoroughly and stored frozen at -70 °C.

Escherichia coli strain JM103 (Messing *et al.* 1981) was used as a recipient during transformation with plasmid DNA. This strain is deleted for the genomic *lac* gene and carries *lac* I^q,Zϕ M15 on an F' plasmid. The cells cannot therefore metabolise lactose or convert the chromogenic lactose analogue 5-bromo-4-chloro-3-indolyl-2-D-galactoside (X-gal) to its blue product. Plasmids containing the alpha segment of the *lac* gene can change the cell phenotype to lac^+ by alpha complementation (Ullmann *et al.* 1967). If there has been an insertion of DNA into the cloning site within the *lac* gene then the plasmid can no longer produce the necessary functional alpha peptide (Gronenborn and Messing 1978). This difference forms the basis of the blue-white colony assay for plasmids containing inserted DNA sequences.

E. coli strains Q358, Q359, WL87 and WL95 were used as hosts for λ bacteriophage. Q359 is essentially isogenic with Q358 except that it contains an integrated P2 prophage (Karn *et al.* 1980). Similarly WL95 is a P2 lysogen of WL87. Only λ phage lacking the *red, gam* and *delta* functions, which will be lost by replacement with non-lambda DNA during cloning, will be capable of apparently normal growth on this strain (Zissler *et al.* 1971).

Bacteriophage plating cells were prepared from a suitable host strain, usually WL87, at mid log-phase in 10 ml of T-broth supplemented with 0.2% maltose. The cells were spun down at 5 000 g for 10 min. and resuspended in 5 ml of 10 mM $MgSO_4$. Maltose induces the *lamB* gene of *E. coli*, which encodes the lambda receptor, as it is part of the maltose operon (Thirion and Hofnung 1972).

2.2. Bacteriophage stocks

Single plaques were prepared by allowing an appropriate number of phage to adsorb to 0.1 ml of WL87 plating cells. After incubating for 15 min. at 37 °C, 4 ml of molten L-top agar at 45 °C was added to the cells and after mixing, poured onto the surface of an L-agar plate. The top agar was allowed to set and the plate incubated overnight at 37 °C.

For small scale liquid cultures, 10 ml of L-broth supplemented with 0.1% glucose and 10 mM $MgCl_2$ were inoculated with 0.1 ml of plating cells. A single plaque, removed from a plate stock with a sterile pasteur pipette was transferred to the flask and shaken vigorously at 37 °C overnight. By this time lysis had usually occurred producing 10^8 to 10^{10} plaque forming units (pfu) per ml.

To obtain larger quantities a suitable host strain, usually WL87 was allowed to grow from an $O.D._{600}$ of 0.05 to one of 0.3 in 400 ml of L-broth made 10 mM with $MgCl_2$. At this point the cells were infected with about 10^9 pfu from a 10 ml liquid lysate. After about 4 hours of vigorous shaking at 37 °C, when cell debris was clearly visible, 0.1 ml of chloroform, 400 µg of RNAse and 400 µg of DNAse were added and the lysate left shaking slowly for 15 min. Over 10^{12} pfu ml^{-1} could be obtained from such cultures.

2.3. Transformation

JM103 cells were made competent to take up plasmid DNA by a method similar to that given by Hanahan (1983). 5 ml of an overnight liquid culture was transferred to a flask containing 100 ml of ψ-broth and incubated with shaking at 37 °C until the $O.D._{600}$ of the culture reached 0.3. The cells were chilled on ice before being centrifuged at 5 000 g for 10 min. The pellets were drained thoroughly, resuspended in 25 ml of Tfb I and kept on ice for 10 min.

before the process was repeated. This time the cells were resuspended in 4 ml of Tfb II, kept on ice for 10 min. then dispensed into 100 µl aliquots and frozen rapidly with liquid nitrogen before being stored at -70 °C.

To obtain transformants, an aliquot of competent cells was thawed slowly on ice and left for 10 min. An appropriate quantity of DNA in TE was added to the cells and left for a further 10 min. on ice to allow adsorption of the DNA to the cell surfaces. After a 90 s heat shock at 42 °C the cells were cooled on ice, 0.5 ml of L-broth was added and the cells incubated at 37 °C for 30 min. to allow expression of the plasmid-encoded ampicillin resistance.

To reduce the volume to be plated out, the cells were spun for 15 s in a microfuge, all but 100 µl of the supernatant removed and the cells gently resuspended. This suspension was plated out with a minimum of spreading onto L-agar containing 100 µg ml^{-1} ampicillin. The plate was incubated at 37 °C overnight. If the presence of β-galactosidase activity was to be tested, 50 µl of a 2% w/v solution of X-gal in di-methyl formamide was added to the cells immediately before spreading. A transformation efficiency of 10^7 to 10^8 colonies per µg of DNA was routinely obtained.

2.4. Preparation of plasmid DNA

Small amounts of DNA were obtained from 1.5 ml of an overnight culture of bacteria grown in the presence of 100 µg ml^{-1} ampicillin by the boiling method (Holmes and Quigley 1981).

Plasmid DNA from a 400 ml stationary phase culture of bacteria in L-broth, supplemented with 0.1% glucose and 100 µg ml^{-1} ampicillin was prepared by the alkaline lysis method exactly as described by Maniatis *et al.* (1982). The DNA was purified by centrifugation to equilibrium at 60 000 g of a solution of the plasmid with 1 g ml^{-1} of CsCl added and made 0.1% w/w with ethidium bromide. The lower band of plasmid DNA, visible when viewed with long wave ultra violet light was collected from the top of the tube with a siliconised pasteur pipette.

The ethidium bromide was removed from the DNA by passing the sample through a 4 cm Dowex (AG 50W-X8, Bio-Rad laboratories) ion exchange resin column in a siliconised pasteur pipette. The resin was equilibrated with

1 M NaCl, 0.01 M EDTA and 0.1 M tris at pH 7.5 prior to use. Ethidium bromide is retained by the column while the DNA is allowed to pass through.

DNA was recovered after the removal of ethidium bromide by ethanol precipitation after diluting the sample three-fold with TE. The resulting pellet was washed with a 70% v/v ethanol-water mixture, resuspended in 2 ml of TE, made 0.7 M with ammonium acetate and the DNA reprecipitated with ethanol. The DNA was again washed with 70% ethanol before being dried thoroughly under vacuum and resuspended in approximately 0.5 ml of TE.

2.5. Preparation of bacteriophage DNA

For small quantities of bacteriophage DNA 0.8 ml of a small scale liquid lysate was made 1 μg ml^{-1} with DNAse and incubated at room temperature for 15 min. Bacterial debris was removed by centrifugation for 2 min. in a microfuge. 0.2 ml of TES was added, shaken gently and incubated at 70 °C for a further 15 min. After cooling on ice 0.15 ml of 8 M potassium acetate was added and the protein and SDS allowed to coagulate for 15 min. on ice before being removed by centrifugation as before. The supernatant was extracted once with phenol-chloroform and the DNA precipitated with ammonium acetate and ethanol. 1 – 2 μg of DNA was usually recovered.

For larger quantities, a large scale liquid culture was used. The lysate was made 1 M with NaCl and the cell debris removed by centrifugation at 16 000 g for 10 min. The cleared lysate was made 7% w/v with polyethylene glycol (8 000 d) and left on ice for at least an hour. The phage particles were collected by centrifugation at 16 000 g for 20 min. and the pellet allowed to drain well before being resuspended gently in 4 ml of TMN. This suspension was extracted twice with chloroform before being saturated with CsCl and introduced onto the bottom of a discontinuous gradient of 3 steps, 56% w/w, 45% w/w and 32% w/w of CsCl in TMN. After centrifugation at 40 000 g for 90 min. the phage particles were collected with a siliconised pasteur from the interface between the 45% and 32% steps. The resulting suspension was made 42% w/w CsCl in TMN before centrifugation to equilibrium at 40 000 g.

The band of phage particles was removed as before and dialysed against 10 mM Tris (pH 8.0), 5 mM NaCl, 1 mM EDTA. In the absence of magnesium ions the phage particles begin to dissociate but the DNA was finally liberated and purified by extraction three times with phenol and recovered by ethanol precipitation.

2.6. Marked Drosophila stocks

The *TM6B,Tb,e* / π_2 and *C(i)DX,y,f* / sn^w *(ii)* stocks were kindly provided by Dr W.R. Engels (Engels 1979; Engels and Preston 1981). These two stocks have essentially a π_2 background and so are strong P strains. The *Df(3R)E-229,red,e* / *TM3,Sb,Ser,e* stock was obtained from D. Ish-Horowicz. The deficiency has breakpoints on either side of the 87A7 heat-shock locus extending from 86F6-8 to 87A9-B2 (Ish-Horowicz *et al.* 1979).

The *TM6B,Tb,e* / *mwh,e* stock was constructed by Dr A. J. Leigh Brown by crossing virgin females from the *TM6B,Tb,e* ; π_2 stock to males from a stock homozygous for a third chromosome carrying mwh,e. The ebony progeny from this cross, which must be of the required genotype were then crossed to each other to set up the final stock. Although this cross is expected to result ultimately in hybrid dysgenesis the stock was maintained for a period of 5 - 10 generations before being used in the third-chromosome extraction procedure. By this time the stock should have become essentially a P strain (Kidwell *et al.* 1981). Subsequent tests have shown that this stock has switched to P cytotype (see Chapter 4).

2.7. Chromosome samples

All the fly crosses involved in the chromosome extraction procedures were performed by J. E. Moss, except for the cross involving the *Df(3R)E-229* deficiency.

For the molecular study of the 87A7 heat shock (HS) locus a series of lines, each homogeneous for a single wild-derived third chromosome were used. These lines were produced from a set of isofemale lines which had been established from a wild-caught sample of *Drosophila melanogaster* from a vineyard (Chateau Gensac), near Condom in France by Dr J. S. Jones. The isofemale lines were maintained for a period of about a year before they were

used in the chromosome extraction procedure.

Figure 2.1 shows the derivation of the extracted chromosome lines and flies from which DNA was extracted. A single male from each isofemale line was crossed to several virgin females from the *TM6B,Tb,e* ; π_2 stock. A single F_1 tubby male, which must carry a single wild-derived third chromosome and the *TM6B,Tb,e* balancer chromosome was then crossed to several virgin females from the *TM6B,Tb,e / mwh,e* stock. Of the four possible classes of progeny the *TM6B,Tb,e* homozygotes were lethal, flies heterozygous for the chromosome carrying mwh,e and a wild-derived chromosome would not display the tubby phenotype and flies carrying the *TM6B,Tb,e* and *mwh,e* chromosomes would have an ebony phenotype. A single tubby, non-ebony male was crossed to a similar female to establish the extracted third chromosome line. Both these flies must be heterozygous for identical copies of the paternal, wild-derived third chromosome and the *TM6B,Tb,e* balancer.

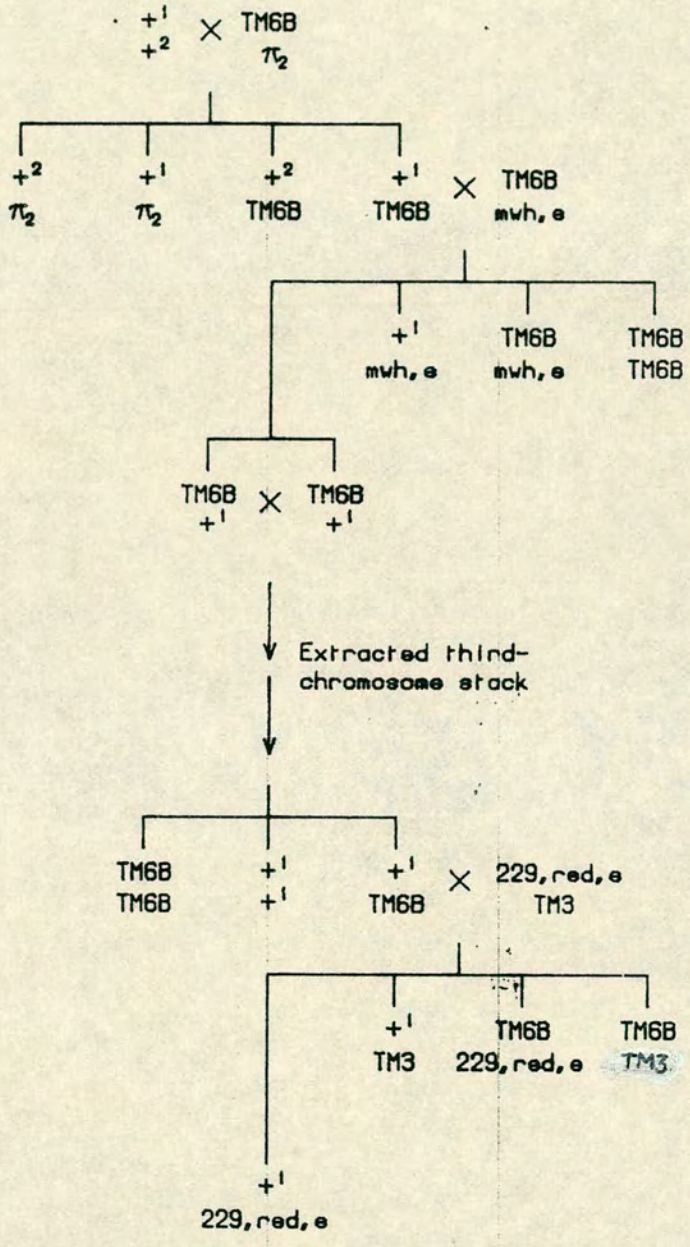
In these extracted third chromosome lines two third-chromosomes segregate. *TM6B,Tb,e* homozygotes are lethal but the wild-derived chromosome may or may not carry a recessive lethal. To simplify the subsequent analysis of DNA fragments, a single male from the extracted third-chromosome line was crossed to several virgin females from an *Df(3R)E-229,red,e / TM3,Sb,Ser,e* stock. Four phenotypic classes of progeny were observed: a) wild-type; b) stubble, serrate; c) tubby, ebony and d) stubble, serrate, tubby, ebony. Of these, the first class were collected and kept on ice before being frozen in liquid nitrogen.

If DNA from these flies was to be prepared within a few days they were stored at -20 °C otherwise at -70 °C. These flies were heterozygous for the wild-derived third chromosome and the *Df(3R)E-229* deficiency leaving the 87A7 heat shock locus effectively hemizygous. DNA prepared from these flies contained sequences homologous to the 87A7 heat shock locus only from the wild-derived chromosome.

Samples of *Drosophila melanogaster* collected from two populations in 1984 were used in the X chromosome study. The first was a sample of flies, kindly supplied by Dr A. E. Shrimpton which had been caught at a fruit market in Raleigh, North Carolina, U.S.A. The second sample was taken from flies found at a refuse tip in Zaharra, Spain, again supplied by Dr J. S. Jones. Both

Figure 2.1

Genealogy of the $+^1$ / 229,red,e flies from which DNA for the 87A7 heat shock locus survey was prepared. $+^1$ and $+^2$ represent wild-derived third chromosomes. A single male progeny from an inseminated wild-caught female was mated to several virgin females from the TM6B,Tb,e / π_2 stock. A single *tubby* male from the offspring was mated to several TM6B,Tb,e / mwh,e virgin females. Single *tubby* males and females were mated to each other to establish the extracted third chromosome stock. Male *tubby* flies from this stock were crossed to virgin 229,red,e / TM3,Sb,Ser,e females and phenotypically wild-type offspring were collected. DNA for the survey was prepared from these flies which were effectively hemizygous for the wild-derived 87A7 heat shock locus. This procedure was repeated with male offspring from many different wild-derived females and DNA from 32 of the lines was used in the survey. G


+1 +2 × TM6B π2
+2 π2
+1 π2
+2 TM6B
+1 TM6B
× TM6B mwh, e
+1 mwh, e
TM6B mwh, e
TM6B TM6B
TM6B +1 × TM6B +1
Extracted third-chromosome stack
TM6B TM6B
+1 +1
+1 TM6B
× 229, red, e TM3
+1 TM3
TM6B 229, red, e
TM6B TM3
+1 229, red, e

samples arrived as wild-caught flies from which inseminated females were separated into vials and their resultant progeny used to establish a series of isofemale lines. A single first generation male from each isofemale line was crossed simultaneously with several virgin female *TM6B,Tb,e / mwh,e* flies and *C(i)DX,y,f* virgin females. The two types of female were then separated and the progeny of the *TM6B,Tb,e* females used elsewhere (Leigh Brown and Moss 1987).

The progeny from the *C(i)DX,y,f* females were crossed to each other to establish the extracted X chromosome lines. In these lines the females carry two X chromosomes and a Y chromosome. The X chromosomes are attached and so segregate as a single chromosome (Lindsley and Grell 1967). Each female inherits the attached X chromosome from its mother, each male a single wild-derived X chromosome from its father. The wild-derived X chromosome, which cannot recombine with the maternally inherited Y chromosome and is never present in females unless the attached X chromosome breaks, will thus remain intact. Males from each of the lines were collected and stored in the same way as the flies from the third chromosome line, *Df(3R)E-229,red,e* cross progeny.

2.8. Cytotype tests

To determine whether a particular *Drosophila* stock had either the P or M cytotype the following procedure was followed. Males from the *C(i)DX,y,f / sn^w (ii)* stock, carrying the sn^w marker on a P chromosome were mated to several virgin females from the stock to be tested. The progeny of this cross were allowed to mass mate and males from the F_2 generation scored for the *singed* phenotype. The flies were reared at a constant temperature of 25 °C

If the stock tested was of M cytotype the original sn^w allele would have a greatly enhanced mutation rate to either sn^e or sn^+ (Engels 1979). Wild type males would be indistinguishable from males carrying the grandmaternal X chromosome so the mutation rate was calculated as the ratio of sn^e flies to the total number carrying either sn^w or sn^e. A high mutation rate would suggest that the tested stock was of M cytotype, not otherwise.

2.9. Preparation of genomic DNA

Between 250 and 500 flies were placed in a manual glass homogeniser (Dounce, Kontes scientific glassware) together with 5 to 10 µl per fly of buffer H. The flies were homogenised with two or three passes of the pestle, leaving a suspension of, mainly, intact nuclei. An equal volume of buffer I was added and mixed gently before incubation at 37 °C for an hour. By this time the nuclei would have lysed and the proteins been digested by the protease, releasing the DNA into solution. Contaminating peptides were removed by extraction twice with phenol, once with phenol–chloroform and once with chloroform alone. The aqueous phase was then made 0.7 M with ammonium acetate and the DNA precipitated by adding two volumes of ethanol.

Centrifugation for 10 min. at 5 000 g produced a pellet of DNA from which the supernatant was removed. This pellet was then redissolved in 2 ml of TE and left at 4 °C overnight before reprecipitating the DNA as before. The DNA pellet this time was rinsed with a 70% v/v ethanol to remove any traces of salt before being completely dried under vacuum. Finally the DNA was redissolved in TE at a concentration of about 0.5 µg ml^{-1} for the heat shock locus survey and 1 µg ml^{-1} for the X chromosome survey and stored at 4 °C. DNA samples were prepared from 32 CM lines, 27 NC lines, 22 FV lines and females from the *C(i)DX,y,f* stock.

2.10. Restriction fragment analysis of DNA

Restriction endonucleases (restriction enzymes) were perchased from Boerhinger–Mannheim, Amersham International and Anglian Biochemicals. Each DNA sample to be digested was made 10 mM $MgCl_2$, 10 mM tris (pH 7.5), 10 mM β–mercaptoethanol and 0, 50 or 100 mM with NaCl according to the manufacturers instructions, in an appropriate volume (usually 20 µl). The reaction was incubated at the recommended temperature until digestion should have been complete (except where stated otherwise). The restriction fragments produced by such digestion were analysed using separation by electrophoresis in TBE buffer through submerged horizontal agarose gels (McDonnel *et al.* 1977) and visualisation by illumination with transmitted short wave ultra violet light once the gel had been equilibrated with 1 µg ml^{-1} ethidium bromide.

Further analysis of the restriction fragments was performed by transferring the fragments to nitrocellulose, 'Gene Screen Plus' (New England Nuclear) or 'Hybond' membranes (Amersham International) and subsequently hybridising radioactively labelled probes to the filters using the procedure of Southern (1975) as modified by Wahl *et al.* (1979). DNA sequences homologous to the probes were visualised by autoradiography.

2.11. Analysis of genomic DNA fragments

For the survey of the 87A7 heat shock locus, restriction enzymes were chosen which would aid the rapid screening of the locus for the presence of any large insertion-deletion events and also allow a comparison to be made with a previous study at this locus on a different population (Leigh Brown 1983) with respect to restriction site variants observed in that survey. The enzymes chosen were *Bam* HI, *Eco* RI, *Pst* I, *Xba* I and *Xho* I.

In the X chromosome survey one of four restriction endonucleases was used to cleave the DNA into fragments. These enzymes, *Bam* HI, *Bgl* II, *Xba* I and *Xho* I, were chosen solely to aid the screening procedure for large insertion-deletion events. The recognition sequences of enzymes *Xba* I and *Xho* I occur approximately 10 times each within the 110 kb region surveyed and consequently give relatively large DNA fragments. The sequences which the enzymes *Bam* HI and *Bgl* II recognise occur approximately 20 and 30 times respectively in the same region. Fragments produced by the first two enzymes allow more of the genome to be surveyed at once in an interpretable fashion although not as accurately or as sensitively as fragments produced by the latter two enzymes. The smaller fragments however are more useful for determining the position of any insertion-deletion events accurately.

5 - 10 µl of the genomic DNA solution were incubated with about 10 units of the restriction enzyme for 5 hours. The fragments produced were separated by electrophoresis through an agarose gel as described above and then transferred to a membrane filter. Nitrocellulose filters were used throughout the 87A7 heat shock locus survey.

2.12. Hybridisation of membrane filters

The dry membrane filters were wetted with 2 x SSC and washed for half an hour each at 65 °C with 0.3% w/v low fat dried milk (LFDM, J. Sainsbury plc), 1 x SSC then 0.3% LFDM, 1 x SSC, 0.5% SDS, 50 μg ml^{-1} herring sperm DNA (HSDNA) and finally 0.3% LFDM, 1 x SSC, 0.5% SDS, 50 μg ml^{-1} HSDNA and 10% w/v dextran sulphate. The last two washes were performed in a polythene bag and the denatured, radioactive probe DNA introduced into the final wash after the half hour incubation period. Hybridisation was at 65 °C overnight.

Low fat dried milk provides an economical replacement for the 0.2% Denhardt's solution (0.2% BSA, 0.2% polyvinylpyrrolidine and 0.2% Ficoll (40 000 d)) usually used (Johnson *et al.* 1984). In my hands this technique produced a reduction in the non-specific binding of probe DNA which compared favourably to that obtained with Denhardt's solution.

After hybridisation, the filters were washed for half an hour each at 65 °C with 2 x SSC, 0.5% SDS then 1 x SSC, 0.5% SDS and finally twice with 0.1 x SSC, 0.5% SDS. Excess liquid was blotted from the filters with tissue paper and they were sealed into thin polythene bags to prevent contamination of the autoradiography cassettes. The filters were laid down against X-ray film (RX, Fuji photo film Co., ltd) with an intensifying screen at -70 °C overnight before the film was developed. If the signal observed on the film was sufficient a second film was laid down as before for one to two weeks.

Filters could be rehybridised to a second probe provided they were 'stripped' of the previous probe. Nitrocellulose filters were incubated at room temperature in 0.5 M NaOH, 1.5 M NaCl for 10 min. and neutralised with 1.0 M tris, 1.5 M NaCl (pH 8.0). Nylon based filters were washed in 0.4 M NaOH at 42 °C for 30 min. and then neutralised as the nitrocellulose was. These filters could be allowed to air dry before use.

2.13. Synthesis of radioactively labelled probes

Two methods of incorporating α-^{32}P labelled deoxycytosine triphosphate (dCTP, Amersham International, about 110 TBq $mmole^{-1}$) into probe DNA were used. Initially this was done by nick translation (Rigby *et al.* 1977) and later by a more recent and efficient method of primer extension (Feinberg and

Vogelstein 1984).

Nick translated probes were prepared by incubating approximately 50 ng of probe DNA with 2 x 10^{-4} units of DNAse (DN-EP, Sigma Chemical Company) and 5 units of Polymerase I (Anglian Biochemicals) in 50 µl of a solution containing 50 mM tris (pH 8.0), 5 mM $MgCl_2$, 10 mM β-mercaptoethanol, 10 µM dATP, 10 µM dGTP, 10 µM dTTP and 0.67 µM α-^{32}P dCTP at 14 °C. The reaction was monitored by spotting 1 µl of the reaction mixture onto a small square of nitrocellulose, drying it under a lamp and measuring the radioactivity present as Cerenkov radiation with the tritium channel of a scintillation counter. The nitrocellulose square was then washed for 10 min. with a cold 5% solution of trichloroacetic acid (TCA) in which any unincorporated nucleotides would dissolve leaving the TCA-insoluble material on the nitrocellulose. The residual radioactivity was then measured and the specific activity of the probe calculated from these values. A maximum of 10^6 to 10^7 Bq μg^{-1} of DNA was usually attained after 3 to 4 hours of incubation.

The second method which was adopted to produce radioactively labelled DNA involved *de novo* synthesis of DNA rather than the replacement of DNA which occurs in the nick translation procedure. Approximately 100 ng of DNA was denatured in 10 µl of water by heating it to 100 °C for 90 s. Supercoiled plasmid DNA was first linearised by digestion with a restriction enzyme to allow efficient denaturation procedure would not result in single stranded DNA. After cooling to room temperature, the DNA was brought to 0.1 mM dATP, 0.1 mM dGTP, 0.1 mM dTTP, 25 mM $MgCl_2$, 250 mM tris (pH 8.0), 50 mM β-mercaptoethanol, 1 M HEPES and 0.3 mM mixed hexadeoxyribonucleotides (Pharmacia) by adding a tenth of the final volume (usually 3 µl in a total of 30 µl) of a ten fold concentrated stock. 1 unit of DNA Polymerase (large fragment) otherwise known as Klenow enzyme (Boerhinger-Mannheim) was added and then 15 µl of the radioactively labelled dCTP.

The reaction was left to proceed at room temperature and monitored in exactly the same way as described before. Specific activities in excess of 10^8 Bq μg^{-1} were attained after two to three hours of incubation. The reaction could be left at room temperature overnight with no obvious adverse effects. Probe DNA prepared in this way was used in exactly the same way as nick translated probes.

Unincorporated nucleotides were removed by the spun Sephadex column method (Maniatis *et al.* 1982). The probe was denatured by making the solution 0.3 M with NaOH and neutralised by bringing it to 0.5 M tris (pH 8.0) before introducing it into the hybridisation bag.

2.14. DNA used as probes

Four molecular clones of the 87A7 heat shock region were used to screen the third-chromosome samples. The plasmids 56H8/C (Leigh Brown and Ish-Horowicz 1981), 87A/5 (from A. Udvardy) and BF17 (from A.J. Leigh Brown) were used initially. To improve the chance of revealing recombinant phage in the genomic library which carried DNA containing the three insertion events most proximal to the centromere (see results) a small *Eco* RI - *Sal* I fragment derived from λ903a (Leigh Brown 1983), a clone extending proximally from the 87A7 heat shock genes, was subcloned into the plasmid pUC8 (Vieira *et al.* 1982). This fragment extends from the left most *Eco* RI site of λ903a, which remains from the multiple cloning site of the vector phage, to the *Sal* I site of the inserted DNA most proximal to the centromere on the genomic map. The resultant plasmid, p903a was used only in the library screening procedure.

A series of 10 bacteriophage clones of the *achaete-scute* complex (Carramolino *et al.* 1982, Campuzano *et al.* 1985) were kindly provided by Dr J. Modellel. For convenience, several plasmid clones were constructed from these phage. The subcloning method used to construct both these plasmids and p903a were the same.

Approximately 10 µg of phage DNA were digested to completion in a total volume of 100 µl with a restriction enzyme. The enzyme was removed by extraction with phenol-chloroform and the digested DNA recovered by ethanol precipitation. In the case of λ903a DNA it was resuspended in TE and digested to completion with the second enzyme and purified as before. The digested DNA was then resuspended in 20 µl of TE. 20 µg of pUC8 DNA were digested and purified as the phage DNA had been, and resuspended in a total volume of 20 µl. 1 µg of digested pUC8 DNA and 2.5 µg of digested phage DNA were allowed to ligate at 4 °C overnight. 1 µl of a 1/10 dilution of this ligation was used to transform 100 µl of competent JM103 cells and of the $> 10^3$ colonies resistant to ampicillin, about 5% could not produce a blue colour in the presence of X-gal.

DNA was prepared from overnight cultures of several white colonies from each plate. This DNA was digested with *Eco* RI, *Pst* I, *Bam* HI or both *Eco* RI and *Sal* I and analysed on a 0.5% agarose gel alongside similar digests of the original phage DNA. By comparison of the fragments liberated from the plasmid with the fragments of the phage, the identity of each recombinant plasmid was established. The plasmid clones were named as follows: each A-SC subclone was called pASC, followed by the number of the parent phage and the enzyme used in the subcloning procedure and then the number of the phage fragment in order from nearest the centromere towards the telomere. As an example, the plasmid containing the A-SC *Eco* RI fragment extending from position 46.9 to 51.0 on the map of Campuzano *et al.* (1985) was called pASC94R4.

The plasmids produced in this way were pASC53R1, pASC31P4, pASC17B1, pASC64R3, pASC94R1, R2, R3, R4, pASC101R5, R7 and pASC133R1. λ DNA fragments acting as markers were revealed on the autoradiographs by using 5 ng of λ DNA as a probe.

Isolated restriction fragments of DNA were sometimes used as probes during the isolation of genomic λ clones. Fragments produced by digestion of approximately 10 µg of substrate DNA were initially separated by agarose gel electrophoresis. The desired fragment was then recovered by rotating the gel through 90 degrees before continuing electrophoresis and collecting the DNA on small squares of DEAE-cellulose (Schleicher and Schuell) inserted into the gel directly adjacent to the band.

Once the DNA was bound to the paper it was rinsed several times with TE before placing the paper in 50 µl of 1 M NaCl in TE and incubating at 65 °C for 30 min. By this time the DNA had been released from the paper and was used immediately for radiolabelling.

2.15. Cloning fragments of genomic DNA

In order to analyse sections of DNA more thoroughly than is possible with Southern transfer of genomic DNA fragments, genomic DNA which had been partially digested with *Sau* 3A was molecularly cloned into λEMBL3 (Frischauf *et al.* 1983). Phage particles containing DNA which shared homology with specific previously cloned fragments were purified.

100 µg of λEMBL3 DNA were digested to completion with *Bam* HI, extracted with phenol-chloroform and ethanol precipitated. After drying thoroughly, the DNA was resuspended in 20 µl of TE. 30 µl of genomic DNA were digested with 0.1 U of *Sau* 3A for 1 hour, 10 U of alkaline phosphatase (Boerhinger-Mannheim) were added and the reaction incubated for a further 20 min. Ethyleneglycol-bis-(2-amino-ethyl ether)N,N'-tetraacetic acid (EGTA) was added to a final concentration of 20 mM and the reaction heated to 70 °C for 15 min. to inactivate the phosphatase. After extracting with phenol-chloroform, the DNA was recovered by ethanol precipitation and resuspended in 20 µl of TE. This treatment of the genomic DNA produced predominantly fragments of 10 kb to 20 kb as determined by electrophoresis through a 0.3% agarose gel.

5 µg of the digested λEMBL3 DNA and 3 µl (approximately 2.5 µg) of genomic DNA partially digested with *Sau* 3A were made 50 mM tris (pH 7.5), 10 mM $MgCl_2$, 10 mM dithiothreitol, 10 mM ATP in a final volume of 10 µl and incubated at 4 °C overnight with 5 U of T4 DNA ligase (Boerhinger-Mannheim). Half of this ligation reaction was packaged into infectious bacteriophage particles as follows. Packaging extracts, freeze-thaw lysate (FTL) and sonic extract (SE) were prepared as described by Arber *et al.* (1983). 7 µl of buffer A, 5 µl of ligation reaction, 2 µl of buffer M1, 15 µl of SE and 10 µl of FTL were added in this order to a sterile eppendorf tube, mixed thoroughly and incubated at room temperature for 1 hour. The reaction was terminated by adding PSB to a final volume of 200 µl and the reaction stored at 4 °C.

By titrating this packaging reaction on both WL87 and WL95 plating cells the total number of viable phage particles and the number which displayed the *spi* $^-$ phenotype was determined. Between 10^4 and 10^6 *spi* $^-$ phage particles were usually obtained. These phage constituted the library of cloned fragments from which the desired phage were purified. The original library was plated out as described earlier on 14 cm L-agar plates supplemented with 0.1 % (w/v) glucose with 9 ml of L-top agarose at a concentration of approximately 10^4 phage per plate using the restrictive host WL95. When the plaques were approaching confluence the plates were cooled at 4 °C for an hour and nitrocellulose lifts were taken from each plate as described by Maniatis *et al.* (1982) (after Benton and Davis 1977). These were hybridised to ^{32}P-labelled probes in exactly the same way as the Southern transfer filters.

In order to position the autoradiographs of each disc, holes made by puncturing the nitrocellulose with a hypodermic needle were marked with a 20 ng μl^{-1} solution of denatured pUC8 DNA by dotting the solution onto the filter with a drawn-out pasteur pipette. The probes used to hybridise to each filter were produced from the plasmid clones described earlier and therefore shared homology with the pUC8 DNA markers. Positive plaques which hybridised to the probe DNA were picked by removing a plug of agar surrounding the positive plaque from the plate into 0.5 ml of PSB. 1 µl serial dilutions of 10^{-3} and 10^{-4} of these suspensions of phage were plated onto 9 cm agar plates as before and new nitrocellulose lifts taken of these. Plaques hybridising to the probe DNA which were well isolated from neighbouring plaques were picked into 0.2 ml of PSB.

A plate stock and a 10 ml liquid lysate, produced from a single plaque from the plate stock were prepared as described earlier. Small scale DNA preparations from the liquid lysates were used to identify each phage clone by analysis of restriction fragments.

2.16. Calculation of DNA Fragment Lengths

The lengths of DNA fragments separated by electrophoresis were estimated using fragments of known length produced from lambda DNA by digestion with various restriction enzymes. These lambda markers were introduced at each side of a gel prior to electrophoresis and allowed a standard curve to be constructed of $\log_{10}$ length against mobility with which unknown fragment sizes may be determined (Bishop *et al.* 1967, Peacock and Dingman 1968). The programme used to produce these size estimates is detailed in Chapter 3.

Throughout this work the fragment length estimates given were obtained as the mean of five length estimates. The standard error was calculated from (3.5).

CHAPTER 3

CALCULATING DNA FRAGMENT LENGTHS

3.1. Background to the Programme

Electrophoresis of DNA molecules through agarose gels separates them according to their lengths. Bishop *et al.* (1967) have shown that with DNA and RNA viral genomes of known molecular weight their rate of migration through polyacrylamide gels was approximately proportional to $\log_{10}$ length; Peacock and Dingman (1968) found later that this was also true for the migration of RNA through polyacrylamide and agarose gels. Further studies indicate that this is a general property of nucleic acid molecules (e.g. Helling *et al.* 1974).

Before the advent of DNA sequencing, the lengths of DNA fragments were estimated by assuming that the sum of all the fragments produced from a given substrate molecule was equal to the length of the original. This was a rather inaccurate procedure, since some of the fragments of DNA may not have been observed on the gels. The ultimate calibration of these estimates depended on the molecular weight of the parent molecule, estimated from analysis of sedimentation coefficients. After fragments of known length could be used as size standards, accuracy improved.

The lengths of DNA fragments are now commonly estimated by producing a standard curve of $\log_{10}$ length of the known fragments against mobility and constructing, freehand, either a straight line through the points, or a curve allowing for the non-linear portions of the graph. Southern (1979) was the first to introduce a more objective method than this of producing length estimates. He assumed a linear relationship between the length of DNA fragments and the inverse of their mobilities and calculated the lengths of DNA fragments directly. Any curvature of the standard curve observed was corrected for by determination of m_o. This was the position of a virtual origin from which the DNA molecules would have reached their observed positions if there truly was an inverse relationship between sequence length and mobility.

A recent attempt to assess the accuracy of sizing DNA fragments (Gough and Gough 1984), compared the relative merits of four different approaches to constructing standard curves. These methods were: hand drawn curves, linear interpolation between each pair of points, sequences of orthogonal polynomial

curves and splined cubic curves where the interval between each pair of points was assumed to follow the arc of a cubic. For most of these the data were transformed for a semi-log scale. From the analysis of their somewhat limited data they concluded that the use of a splined cubic curve gave the most accurate estimates of size although it seems hard to distinguish between the performance of each.

Since it remains unknown exactly how DNA migrates through a gel matrix, which therefore precludes the formulation of an exact function describing mobility with length, an approximation to this presumed function is necessary in order to calculate the size of DNA fragments directly. The choice of a splined curve seems very attractive because it conforms closely to the idea of a hand drawn curve. The mathematical theory involved stems from the desire to model the behaviour of a draftsman's 'spline', a long, thin strip of wood used to construct arbitrarily a smooth curve through a set of points. The use of splined curves is extensive throughout mathematics and they are widely applied to problems of numerical analysis, especially interpolation, of complicated functions (Ahlberg *et al.* 1967).

3.2. The 'Spline' Programme

The programme to calculate DNA fragment lengths was written in standard Pascal modified for the 'p-system' environment in which it was used on an Apricot microcomputer (Applied Computer Techniques) attached to a graf/bar digitizer (Science Accessories Corporation). The digitizer has a stylus which, when depressed onto a surface, emits a click produced by a small static discharge. The co-ordinates of the stylus position, relative to an arbitrary origin are determined from the time lag between the discharge and the sound being received by two small microphones on either side of the digitizer. These co-ordinates are transmitted to the computer as 4 digit integers, giving values to the nearest 0.1 mm.

To use the programme the operator first relocates the origin to above and left of the gel photograph. If only one marker lane was used on the gel it is assumed that the photograph is straight and that each lane has run perpendicular to the digitizer. If more than one marker lane was used (there can be up to 4 marker lanes at present) then the operator depresses the stylus (digitises) on one band of the same molecular weight from the furthest left

and right marker lanes. The line between these points is assumed to be parallel to the origin of the gel and subsequent X co-ordinates are then rotated about the digitizer origin to account for this. A similar procedure is repeated, digitising the top and bottom of any one lane of the gel to allow correction of the subsequent Y co-ordinates.

To set up a standard curve for the gel, each band of each marker track is then digitised and the distance through which each fragment has migrated is calculated as the averaged, relative mobility over all marker lanes. These distances are then associated with the $\log_{10}$ of the known fragment lengths stored in the computer. A splined cubic curve, passing through each mobility, $\log_{10}$ length co-ordinate is constructed as follows (after Ahlberg *et al.* 1967).

Let N+1 be the number of coordinates. In this case there are N+1 different mobilities (x_0, x_1, x_2, . . . , x_{N-1}, x_N) which are associated with a set of $\log_{10}$ size (y_0, y_1, y_2 , . . . , y_{N-1}, y_N). We have at any point x_j (j = 1, 2, . . , N−2, N−1),

$$\lambda_j m_{j-1} + 2m_j + \mu_j m_{j+1} = \frac{3\lambda_j(y_j - y_{j-1})}{h_j} + \frac{3\mu_j(y_{j+1} - y_j)}{h_{j+1}} \qquad (3.1)$$

where m_j is the slope of the curve at the jth coordinate, $h_j = x_j - x_{j-1}$, $\lambda_j = h_{j+1} / (h_j + h_{j+1})$ and $\mu_j = 1 - \lambda_j$.

The equation system for all the points in the curve can be expressed in the form,

$$\begin{bmatrix} 2 & \mu_0 & 0 & \cdots & 0 & 0 & 0 \\ \lambda_1 & 2 & \mu_1 & \cdots & 0 & 0 & 0 \\ 0 & \lambda_2 & 2 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 2 & \mu_{N-2} & 0 \\ 0 & 0 & 0 & \cdots & \lambda_{N-1} & 2 & \mu_{N-1} \\ 0 & 0 & 0 & \cdots & 0 & \lambda_N & 2 \end{bmatrix} \times \begin{bmatrix} m_0 \\ m_1 \\ m_2 \\ \vdots \\ m_{N-2} \\ m_{N-1} \\ m_N \end{bmatrix} = \begin{bmatrix} c_0 \\ c_1 \\ c_2 \\ \vdots \\ c_{N-2} \\ c_{N-1} \\ c_N \end{bmatrix}$$

Where c_j (j = 1, 2 , . . , N−1) represents the right hand side of (3.1) and the end conditions employed are,

$$2\mu_0 + \mu_0 m_1 = c_0$$
$$\lambda_N m_{N-1} + 2m_N = c_N$$

From this equation system a second is formed (k = 0, 1 , . . , N),

$$\begin{aligned} p_k &= \lambda_k q_{k-1} + 2 \\ q_k &= -\mu_k / p_k \\ u_k &= (c_k - \lambda_k u_{k-1})/p_k \end{aligned} \qquad (3.2)$$

with $q_{-1} = 0$ and $u_{-1} = 0$. From this we have the equivalent system (k = 0, 1 , . . , N),

$$m_k = q_k m_{k+1} + u_k \qquad (3.3)$$

with $m_{N+1} = 0$.

The values of m_j (j = 0, 1 , . . , N) are determined from (3.2) and (3.3).

The spline function at x on $[x_{j-1}, x_j]$ may be expressed in terms of the slopes of the curve (m_j) as,

$$\begin{aligned} S\Delta(x) = {} & \frac{m_{j-1}(x_j - x)^2(x - x_{j-1})}{h_j^2} - \frac{m_j(x - x_{j-1})^2(x_j - x)}{h_j^2} \\ & + \frac{y_{j-1}(x_j - x)2[(x - x_{j-1}) + h_j]}{h_j^3} \\ & + \frac{y_j(x - x_{j-1})2[2(x_j - x) + h_j]}{h_j^3} \end{aligned} \qquad (3.4)$$

As each unknown band is digitised, the length is calculated from its mobility (x) as the antilog of (3.4).

3.3. Performance Test of the 'Spline' Programme

In order to assess the performance of the 'spline' programme, the reproducibility and the accuracy of the length estimates it produces were examined. For this purpose four autoradiographs of 1% agarose gels produced in the X chromosome survey were chosen. Under the conditions of electrophoresis used, DNA fragments of 5 kb or more have a mobility which is

not directly proportional to $\log_{10}$ length. With 0.3% agarose gels the proportionality extends over a wider size range as shown in figure 3.1. This aspect of the 1% gels is useful here since the performance of the programme may be investigated over an almost linear and a more greatly curved portion of the standard curve from a single gel.

Each gel contained four marker lanes which were the two left- and right-most pairs of lanes, λ*Hind* III fragments in a lane next to λ*Pst* I fragments. In the test the λ*Hind* III fragments were used to construct a standard curve for each gel and the lengths of eight fragments in the λ*Pst* I lanes between the largest and smallest λ*Hind* III fragments were estimated five times each. This procedure was repeated and the roles of the different markers then reversed, five λ*Hind* III fragments estimated from the λ*Pst* I fragments. From the four gels, two trials, two marker lanes on each gel and five estimates from each fragment, a total of 80 estimates were produced. Extrapolation was avoided because in each case where length estimates were required in the chromosome surveys the unknown bands lay within the span of the marker fragments used. Also extrapolation tends to be much more inaccurate than does interpolation (Gough and Gough 1984).

3.3.1. Accuracy of the Digitizer

The primary scource of error in the DNA length estimates was inaccuracy in the distance estimates produced by the digitizer. A systematic error would only be introduced between different fragment sizes if there were a correlation between the distance estimates produced by the digitizer and its associated error. The corrected distances used to generate the length estimates as described above were produced in groups of five from a common origin. Since the origin was relocated between trials the mean values of the 208 groups of five estimates should be approximately normally distributed, as should be their associated variances.

A scatter diagram of mean distance against variance for the values from this survey suggests no real correlation between the two. Analysis of variance (ANOVA) of least squares linear regression (Sokal and Rohlf 1969, page 421) yield the figures given in table 3.1. The F value for the regression with 1 and 206 degrees of freedom in the numerator and denominator respectively is 4×10^{-4}. This is less than the critical value at the 5% level with one and an

infinite number of degrees of freedom which is 3.84. The null hypothesis, that there was no correlation between the two is thus accepted. The gradient of the least squares regression line through these points is not then significantly different from zero, distance estimates from anywhere within the region tested are equally inaccurate. Combining the sums of squares over all 208 groups gives an overall variance estimate of 0.0459 mm^2. The 95% confidence limits of any single distance estimate are thus ± 0.42 mm.

3.3.2. Accuracy of the DNA Fragment Length Estimates

It has been assumed that any effects on the accuracy of the length estimates produced by the 'spline' programme were not systematic.

The variance of a group of five length estimates about the group mean is an estimate of the error variance, the inherent variance of the measurements. The variance of these group means about the mean for any one trial, multiplied by the number of length estimates in a group is another variance estimate. This term includes the error variance and if estimates taken from the same lane, or two different lanes within a trial are not significantly different, then this is also an estimate of the error variance. If estimates from different lanes are more different from each other than estimates within a lane, then there must be some error introduced between lanes not present within a lane included in this second variance estimate.

Combining estimates within any trial the variance may be expressed as:

$$S^2 \simeq \sigma^2 + d\sigma_L^2$$

where $d\sigma_L^2$ represents the variance component introduced between lanes above and beyond σ^2, the error variance.

This argument holds when estimates from different trials within any one gel or the data from all gels are combined. The variance estimates produced by each are,

$$S^2 \simeq \sigma^2 + d\sigma_L^2 + cd\sigma_T^2$$

or

$$S^2 \simeq \sigma^2 + d\sigma_L^2 + cd\sigma_T^2 + bcd\sigma_G^2$$

respectively. The values a, b, c and d are used to represent the number of estimates within a lane, number of lanes, trials and gels respectively. The subscripts L, T and G denote lanes, trials or gels.

This type of analysis is termed nested model II ANOVA. The term nested indicates that the estimates of each band are grouped within a lane, pairs of lanes are grouped within a trial and trials are grouped within gels.

The calculation of sums of squares and mean squares was performed by a programme, written in standard Pascal for the EMAS mainframe computer at Edinburgh. The programme was based on the methodology suggested by Sokal and Rohlf (1969). Fragments of different known lengths were treated independently. Although the distance estimates all fall within the same distribution, the errors are multiplied by the gradient of the standard curve, which is not constant (figure 3.1). This invalidates any pooling of values from different sized fragments.

Tables 3.2 and 3.3 show examples of two of the thirteen ANOVA tables produce by the programme. Each mean square divided by the mean square of the level immediately below it gives an F value. If this value is less than the critical value, the null hypothesis that both mean square values are estimates of the same quantity is accepted. If greater than the critical value, the alternative hypothesis which postulates the presence of a variance component at the higher level of subdivision, not present at the lower one, is accepted.

In cases where a particular mean square value was not significantly greater than the value immediately below it, the next lowest, significant mean square value was taken to be the denominator. The reason for this was to avoid pooling data without further knowledge about the nature of the variation. This second procedure yields an F value with which to test for combined variance components introduced between the two levels of subdivision chosen. As can be seen from tables 3.4 and 3.5, only the values of $d\sigma_L^2$ are significant at the 5% level. These values along with the error variance and the percentage contribution to the total variance between marker lanes are given in tables 3.6 and 3.7.

These results suggest that the variance between marker lane means accounts for about 90% of the variance within a trial and that this is the only

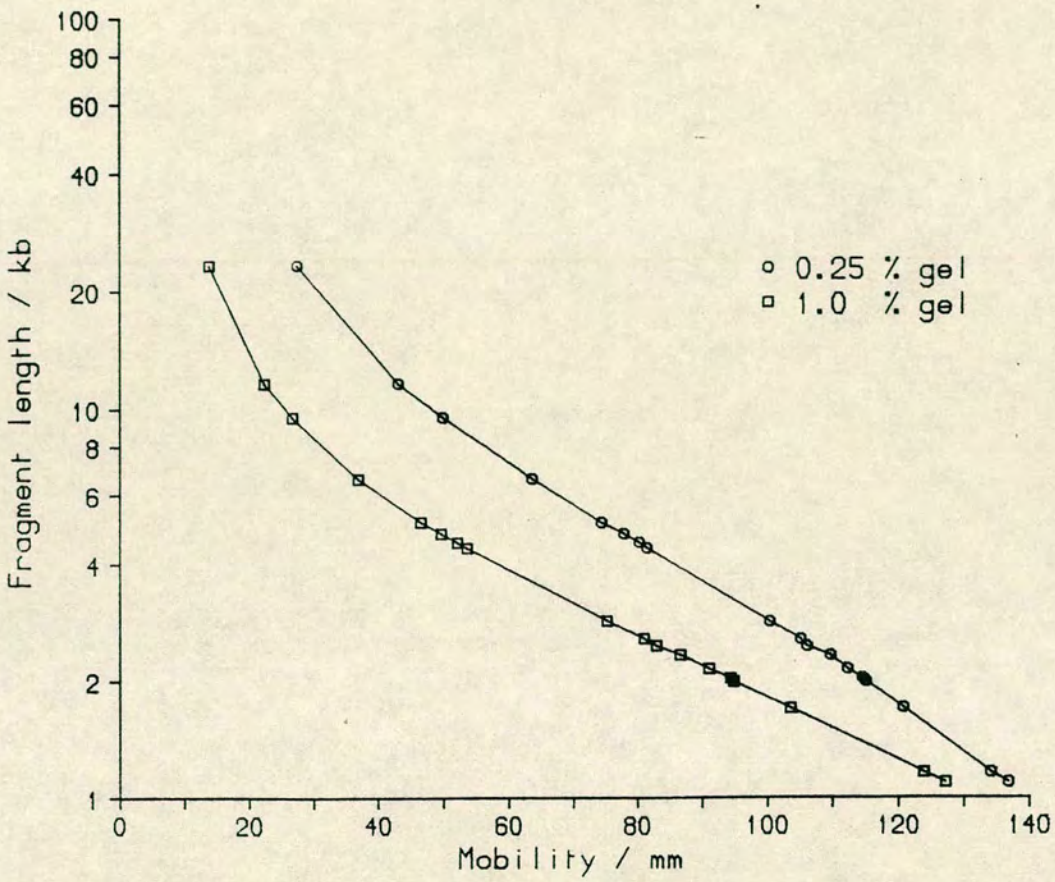
Opposite page 38

Figure 3.1

(Top) Graph showing how the mobility of DNA fragments varies with $\log_{10}$ fragment length through gels of two different agarose concentrations. The data are from λ *Hind* III and *Pst* I fragment markers used in the AS-C locus survey. The points on the upper curve are those from the 0.25% gel. Those on the lower curve are from the 1.0% gel.

Table 3.1

(Bottom) Analysis of variance table for the least squares linear regression of variance in the mean distance estimates produced by the digitizer. **Table**



Source of Variation	d.f.	Sums of Squares	Mean Squares
Regression	1	1.07×10^{-4}	5.17×10^{-7}
Residual	206	2.67×10^{-1}	2.67×10^{-1}
Total	207	2.67×10^{-1}	1.28×10^{-3}

Table 3.2

(Top) ANOVA table for the length estimates produced by the digitizer of the 2.838 kb λ *Pst* I fragment.

Table 3.3

(Bottom) ANOVA table for the length estimates produced by the digitizer of the 9.416 kb λ *Hind* III fragment.

Source of Variation	d.f.	Sums of Squares	Mean Squares
Between gels	3	5.68×10^{-3}	1.89×10^{-3}
Between trials	4	2.22×10^{-2}	5.55×10^{-3}
Between lanes	8	3.44×10^{-2}	4.30×10^{-3}
Error	64	6.72×10^{-3}	1.05×10^{-4}
Total	79	6.91×10^{-2}	8.74×10^{-4}

Source of Variation	d.f.	Sums of Squares	Mean Squares
Between gels	3	7.33×10^{-1}	2.44×10^{-1}
Between trials	4	1.62×10^{-1}	4.04×10^{-2}
Between lanes	8	3.79×10^{0}	4.74×10^{-1}
Error	64	5.18×10^{-1}	8.09×10^{-3}
Total	79	5.21×10^{0}	6.59×10^{-2}

Opposite page 40

Table 3.4

(Top) F ratios for the λ *Pst* I fragments used to test for the presence of the variance component given in brackets.

Table 3.5

(Bottom) F ratios for the λ *Hind* III fragments used to test for the presence of the variance component given in brackets.

Band number	1	2	3	4	5	6	7	8
$F[8,64]$ $(d\ \sigma_L^2)$	44.31	50.42	39.00	31.87	40.91	42.88	39.51	45.02
$F[4,8]$ $(cd\ \sigma_T^2)$	0.33	0.05	0.14	0.29	1.29	1.26	0.99	0.16
$F[3,8]$ $(cd\ \sigma_T^2+bcd\ \sigma_6^2)$	0.07	0.80	0.77	0.93	0.44	0.62	0.24	0.44

Band number	1	2	3	4	5
$F[8,64]$ $(d\ \sigma_L^2)$	58.61	52.91	40.21	86.72	104.05
$F[4,8]$ $(cd\ \sigma_T^2)$	0.09	0.14	0.40	0.04	0.03
$F[3,8]$ $(cd\ \sigma_T^2+bcd\ \sigma_6^2)$	0.52	1.15	0.58	0.26	0.17

Opposite page 41

Table 3.6

(Top) Estimates of the error variance (σ^2) and the variance between lanes (σ_L^2 for the λ*Pst* I fragments and their relative contributions to the total variance between lanes, given as a percentage.

Table 3.7

(Bottom) Estimates of the error variance (σ^2) and the variance between lanes (σ_L^2 for the λ*Hind* III fragments and their relative contributions to the total variance between lanes, given as a percentage.

Band Number	σ^2	σ_L^2	$\%\sigma^2$	$\%\sigma_L^2$
1	4.76×10^{-2}	4.13×10^{-1}	10.35	89.65
2	5.50×10^{-4}	5.43×10^{-3}	9.19	90.91
3	5.38×10^{-4}	4.09×10^{-3}	11.63	88.37
4	3.68×10^{-4}	2.27×10^{-3}	13.94	86.06
5	1.05×10^{-4}	8.40×10^{-4}	11.13	88.87
6	7.80×10^{-5}	6.53×10^{-4}	10.66	89.34
7	1.03×10^{-4}	7.92×10^{-4}	11.49	88.51
8	7.52×10^{-5}	6.63×10^{-4}	10.20	89.80

Band Number	σ^2	σ_L^2	$\%\sigma^2$	$\%\sigma_L^2$
1	8.09×10^{-3}	9.32×10^{-2}	7.99	92.01
2	2.87×10^{-3}	2.98×10^{-2}	8.79	91.21
3	4.39×10^{-4}	3.44×10^{-3}	11.31	88.69
4	8.97×10^{-5}	1.54×10^{-3}	5.51	94.49
5	8.16×10^{-5}	1.68×10^{-3}	4.63	95.37

Table 3.8

(Top) The known lengths of the λ *Pst* I fragments in this survey together with the mean length estimates and their associated standard errors. Values calculated from an arbitrarily chosen group are indicated by $^{+}$.

Table 3.9

(Bottom) The known lengths of the λ *Hind* III fragments in this survey together with the mean length estimates and their associated standard errors. Values calculated from an arbitrarily chosen group are indicated by $^{+}$.

Band number	1	2	3	4	5	6	7	8
Fragment length/kb	11.501	5.077	4.749	4.507	2.838	2.556	2.451	2.140
Mean length /kb	12.277	5.110	4.749	4.483	2.847	2.568	2.482	2.159
Standard error /kb	0.542	0.065	0.058	0.046	0.030	0.026	0.027	0.022
Mean length*/kb	12.773	5.224	4.787	4.475	2.850	2.557	2.521	2.196
Standard error*/kb	1.109	0.099	0.063	0.030	0.015	0.017	0.032	0.032

Band number	1	2	3	4	5
Fragment length/kb	9.416	6.557	4.361	2.322	2.027
Mean length /kb	9.860	6.753	4.397	2.298	2.002
Standard error /kb	0.257	0.159	0.054	0.031	0.032
Mean length*/kb	9.851	6.524	4.386	2.324	1.954
Standard error*/kb	0.275	0.126	0.075	0.027	0.044

significant component variance other than the error variance. Most of the variability between two length estimates from a given fragment thus derives from the variation in mobility across a single gel. The mean squares value between marker lanes is then the most reasonable choice for the variance associated with any single length estimate.

The acceptance of the hypothesis that the mean squares between and within trials estimate the same quantity may be interpreted in the following way. The use of two separate standard curves when estimating the length of a given fragment does not cause a significant increase in the variance of such estimates above the variability already present due to the physical properties of the gel electrophoresis system. In this sense the programme performs as well as could be expected. Improvements could be made but would involve complicated and lengthy methods of correcting the distance estimates returned by the digitizer. Effects such as 'smiling' gels and other deviations from fragments running in parallel would need to be accounted for.

As might be expected, the error variance of different fragments increases with fragment length. It is surprising however that the percentage contribution of the variance component, $d\sigma^2_L$, to the total variance between lanes is remarkably similar between fragments of different lengths. This property may be used to provide a simple method of approximation to the total variance between lanes.

For any band appearing on a gel several, say n, length estimates from the same standard curve will have an associated variance, the error variance term of the above analysis. Approximately nine times this value is an estimate of σ^2_L, the variance component between lanes. The variance associated with the mean length of these n estimates may be calculated as,

$$S^2 \simeq \sigma/n + \sigma^2_L \simeq \sigma^2/n + 9\sigma^2 \qquad (3.5)$$

For example the five length estimates of the *Bgl* II fragment disclosed with pASC53R1 as a probe in the AS-C survey were 6.823, 6.824, 6.805, 6.760 and 6.781 kb. The length of this fragment is therefore given as 6.80 ±0.08 kb. Tables 3.8 and 3.9 show the known length of each fragment in this survey, its estimated size and the standard error of this estimate, together with values

obtained from a group of five estimates, chosen arbitrarily from the sixteen such groups.

The use of a computer in the estimation of DNA fragment lengths greatly reduces the time and effort required, while at the same time producing more reproducible results than might be expected from hand drawn curves. The estimates produced, except for very large fragments, all have standard errors of about 1% of the true length. The method of evaluating these standard errors, although only approximate, seems to be adequate. With a 95% confidence limit the frequency of estimates which are expected to be more than 1.96 times the standard error from the true value is 0.05. One value from the last two rows of tables 3.8 and 3.9 is outside this range, a frequency of 0.077.

The provision of an error for a fragment length estimate gives an idea of how sensitive the electrophoresis conditions were for that particular gel. From this it may more easily be seen what size of change might or might not be detectable, if none has been. Previously this kind of judgement could only be made on prior experience of the electrophoresis conditions and was highly subjective. It is usually possible to see differences of about 0.5 mm between bands in adjacent lanes. The standard error of mobility estimates from the digitizer was about 0.4 mm. This distance error is transformed to a length error of the order of the standard error given by (3.5) and so differences of this size and greater should be detectable by visual inspection.

It should be noted here that this approach of approximation to a standard curve could equally well be applied to untransformed data or data transformed to any scale, such as length *versus* the inverse of mobility. The correction method suggested by Southern (1979) for this last type of transformation would only hold if the standard curve produced was uniformly concave or convex. As may be seen from figure 3.1, for lower percentage agarose gels, both lower and higher molecular weight DNA migrates faster than expected, generating a point of inflection. A splined cubic curve approximation, which would take this into account is therefore more widely applicable.

CHAPTER 4

RESULTS

4.1. Cytotype Tests

The phenomena associated with the two known types of hybrid dysgenesis in *Drosophila* include increased rates of transposition of the P element (Simmons and Lim 1980; Rubin *et al.* 1982; Engels 1983) or the I factor (Bucheton 1984). It has also been suggested that the members of other repetitive DNA families have increased transposition frequencies under these conditions (Gerasimova *et al.* 1984). There is, however, some evidence against this (Eanes W.F., personal communication).

To obtain chromosomes which are representative of those found in natural populations it is essential that the treatment of samples, once collected, does not induce any significant change from their natural state. A main part of this project was the study of insertion-deletion variation, such as that known to be associated with transposable elements (Leigh Brown 1983; Aquadro *et al.* 1986). It is important to know if the extraction procedures used to prepare lines homogeneous for a single chromosome were of the kind which might induce the movement of transposable elements.

Allowing virgin females from a stock of unknown cytotype to mate with males from a known P strain will induce hybrid dysgenesis in the offspring if the female does not have the P cytotype. It has been shown that the sn^w mutation, which is due to the insertion of P element sequences, is highly unstable under conditions of P-M hybrid dysgenesis (Engels 1979, 1984). The P strain males used in the test crosses here carry the sn^w allele. The mutation rate of this allele in the F_1 germ cells is a measure of the cytotype of the female tested. A high mutation rate indicates that hybrid dysgenesis has occurred, a low one that the cytoplasm of the test female ova prevented the increased transposition of P elements and therefore is of the P cytotype.

The mating scheme used to perform these tests is detailed in Chapter 2. Stocks tested were *mwh,e, TM6B,Tb,e* ; π_2, *TM6B,Tb,e / mwh,e, C(i)DX,y,f* and isofemale lines CM10, NC3, 10, 30 and 35. It was already known that the *TM6B,Tb,e;* π_2 and *C(i)DX,y,f* stocks were of the P cytotype (Engels 1979; Engels and Preston 1979) so these were included as controls. Isofemale lines

were derived from single wild-caught females and had never mated to other laboratory stocks. These lines should therefore have retained their original cytotype with respect to the two systems of hybrid dysgenesis. The numbers of sn^w and sn^p males in the F_2 progeny were recorded and are given below in table 4.1.

With both balancer chromosome stocks known to be P strains, *C(i)DX,y,f* and *TM6B,Tb,e*, π_2, no sn^p progeny were observed. The *TM6B,Tb,e / mwh,e* stock produced only a single sn^p male, as did the isofemale line CM10. None of the F_2 male progeny derived from the NC isofemale lines were sn^p. With 95% confidence limits all of these lines had mutation rates significantly lower than 5%. The *mwh,e* stock however had an estimated mutation rate of 28% which is significantly greater than 5%. All the lines, except the *mwh,e* stock, have thus been classified as having the P cytotype. This is consistent with recent surveys of flies throughout the world (Kidwell 1983; Anxolabehere *et al.* 1985).

The *mwh* mutation was first isolated in a wild-caught Moltrasio stock (Di Pasquale 1951) and ebony was discovered in the 1920's (Lindsley and Grell 1967). Strains of flies isolated from the wild before 1950 are commonly M strains (Kidwell 1983) so the *mwh,e* stock was expected to be an M strain. The high mutation rate of sn^w to sn^p in the progeny of the test cross involving this stock supports this.

In these test crosses, a known P strain male was used in the initial cross. Each strain tested, with the exception of the *mwh,e* stock, did not enhance transposition of the P elements at the *singed* locus of the sn^w stock. Crosses between any two of the lines should not be expected to mobilize P elements within the genome above the background level of transposition within a P strain (Engels 1979). Most natural and laboratory strains of *Drosophila melanogaster* which have been found to be of the P cytotype are also of the I cytotype (Kidwell 1979). It has been assumed therefore that this is also true here and that the strains found to have the P cytotype are also I strains.

The mating procedures employed between obtaining flies from the wild and preparing DNA from the extracted chromosome lines were not P-M dysgenic and almost certainly not I-R dysgenic, except for the final cross in preparing DNA from the third chromosome survey. This cross probably was

dysgenic but since DNA was prepared from adult F_1 flies the small quantity of DNA in the germ cells which may have undergone some rearrangement would not be significant. The DNA samples should accurately reflect the natural state of chromosomes in the wild.

4.2. Wild-Derived Chromosome Surveys

Genomic DNA was prepared from a series of lines, each containing DNA from only a single chromosome sampled from the wild in the desired region of study. Digestion of these DNA samples with restriction enzymes produced fragments which could be identified by Southern hybridization (Southern 1975) after separation by electrophoresis through an agarose gel (see Chapter 2). The pattern of fragments produced was interpreted as a reflection of the state of specific regions of the chromosomes in the wild population.

The lengths of fragments were determined as described in Chapter 3. From these sizes, and knowledge of the restriction map expected, the identity of each fragment was established. Where differences occurred between different DNA samples, the reason for the difference was determined. If a single nucleotide had changed in one sample, to either introduce or remove a single restriction site, then the fragments produced by this enzyme would be altered, usually in such a way that the sizes of the fragments which do appear sum to the same length as fragments from the other, unchanged lines. If however the fragments produced by more than one enzyme were affected, usually with a concomitant alteration in the sum of fragment lengths, then the underlying reason was probably the insertion or deletion of DNA.

In order to try to establish the nature of the insertion-deletion events found, genomic DNA libraries were constructed in bacteriophage λEMBL3 (Frischauf *et al.* 1983) from lines which appeared to contain these (see Chapter 2). Using the process of hybridization of radioactive probes to the DNA from these phage, clones containing sequences homologous to the probe were purified. Subsequent analysis of these purified phage enabled more detailed study of the cloned sequence than was possible with Southern analysis of genomic DNA fragments.

The estimates of length produced in the surveys described below are given in tables 4.2 to 4.6. These tables show the probes used to detect the

Opposite page 47

Table 4.1

Number of sn^w and sn^e male progeny in the F_2 offspring from the cytotype test crosses (see text). The number in brackets in column 5 indicates the percentage value used to determine the upper confidence limit given in the last column.

Stock Tested	No. sn^{v} Males	No. sn$^{\bullet}$ Males	Total No. Males	Mutation Rate(%)	95% Limits
TM6B, Tb, e π_2	203	0	203	0.0	0 - 1.80
TM6B, Tb, e mwh, e	233	1	234	0.4(1)	0 - 3.27
C(i)DX, y, f	284	0	284	0.0	0 - 1.29
Isofemale CM10	151	1	152	0.7(1)	0 - 4.16
Isofemale NC3	336	0	336	0.0	0 - 1.09
Isofemale NC10	373	0	373	0.0	0 - 0.99
Isofemale NC30	337	0	337	0.0	0 - 1.09
Isofemale NC35	196	0	196	0.0	0 - 1.87
mwh, e	103	40	143	28.0	20.95 - 35.99

Opposite page 48

Table 4.2

Estimated fragment sizes and standard errors for the fragments revealed by the e probes used to survey the 87A7 heat shock region. The size estimates were generated by the 'spline' programme (see Chapter 3).

Restriction Enzyme	Probe	Fly Line Scored	Fragment Number	Fragment Size/kb	Standard Error/kb
EcoRI	p56H8/C	CM19*	1	10.95	0.13
		CM16	1	15.17	0.30
		CM22	1	10.55	0.08
XhoI	p87A/5	CM25*	1	10.16	0.23
		CM21	1	5.39	0.10
XbaI	p87A/5	CM17*	1	16.61	0.59
		CM4	1	20.24	0.79
		CM21	1	21.74	0.63
		CM24	1	20.15	0.62
BamHI	p87A/5	CM26	1	3.33	0.04
BamHI	pBF17	CM19*	1	2.54	0.02
		CM22	1	2.39	0.04
PstI	p56H8/C	CM18*	1	3.49	0.09
		CM26	1	4.11	0.09
PstI	p87A/5	CM20	1	1.72	0.01

fragments, the length estimates and standard errors produced from the 'spline' programme (see Chapter 3). In nearly all cases DNA samples produced the same size fragments as each other. A single line which displayed the common fragment pattern was chosen arbitrarily and length estimates were made using this line. With fragments which differed in size from the common pattern, estimates of length were made in each case. The details of each autoradiograph are given in Appendix I.

The fragment number given in these tables denotes only the order of fragments in a particular autoradiograph, from largest to smallest. Where the appearance of more than one band coincided with the disappearance of one from the most common pattern, the fragments were also given a letter to denote the different fragments, presumed to be derived from the one absent. An example of this would be the replacement of the largest *Bgl* II fragment, fragment 1 from line NC17, revealed with the two plasmids pASC94R1 and pASC94R4 by the fragments 1a and 1b in line FV2 (table 4.3).

4.2.1. 87A7 Heat Shock locus

Five enzymes were used to digest the DNA in this survey. *Xho* I and *Pst* I were included to survey the known polymorphic restriction sites (Leigh Brown 1983). *Eco* RI and *Xba* I were included because they gave relatively large restriction fragments in this region which facilitated the search for insertion-deletion variation. *Bam* HI was used to screen for the small insertion-deletion events found in the previous survey between the two heat shock transcripts (Leigh Brown 1983).

The probes for this region were p56H8/C, pBF17 and p87A/5 (figure 4.23) which hybridize to the distal, spacer and proximal flanking regions of the two transcripts respectively. Fragments from about 25 kb of DNA surrounding the 87A7 HS locus could be surveyed with these. The estimated sizes of the fragments produced by the various enzymes and probes used are given in table 4.2.

As described above it is possible to interpret the pattern of fragments produced by more than one restriction enzyme with the use of a known restriction map. Figure 4.23 shows a molecular map of the 87A7 heat shock region compiled from various sources (Ish-Horowicz and Pinchin 1980; Leigh

Brown and Ish-Horowicz 1981; Leigh Brown 1983 and personal communication). The Co-ordinates of this map have been assigned, arbitrarily, as distances in kilobases from the distal *Eco* RI site. The restriction sites screened in this survey are shown above the map together with sites observed in some of the bacteriophage clones obtained. Plasmid clones used to survey this region are shown below the map and the variation due to insertion-deletion events indicated as triangles. Dashed lines represent the region of uncertainty for the point of insertion. A line arrow indicates the observed polymorphic *Pst* I site and the open arrows represent the two known transcripts in this region.

Co-ordinates 0.0 to 10.9

31 lines produced visible *Eco* RI fragments with p56H8/C. Of these most were about 10.95 kb, as estimated from line CM19. Size estimates were only possible with one gel because no bands were visible in the λ DNA fragment marker lanes in the other two. With this gel CM16 was estimated to have a fragment of 15.17 kb and CM22 one of 10.55 kb (figure 4.1). These represent an increase of 4.22 kb and a decrease of 0.4 kb respectively. With line CM22 the difference in size was not significant on the basis of the standard error estimates but was visible on this gel. On the remaining gels CM13 and 30 both appeared to have fragments at least 1 kb smaller than the common type. These lines were not on the same gel so it was not possible to say if either was larger than the other.

To resolve this point four samples were chosen as representatives of the size differences found above. Each sample was digested with either *Eco* RI or *Bam* HI separately. The fragments homologous to pBF17 were examined (figure 4.2). A single line of bands of the same size throughout either digest were observed. These fragments could not be explained on the basis of the molecular map of the 87A7 region and were presumed to be an artefact. One explanation, which seems likely, is that the loading buffer added to each sample following digestion, immediately before loading on the gel, could have been contaminated with a small amount of plasmid DNA.

Ignoring these fragments a single *Eco* RI fragment was observed in each of the digests. Line CM16 again possessed a larger fragment than the common type represented by CM21. The fragments of lines CM13 and 30 co-migrated

and were smaller than the fragment of line CM21.

Co-ordinates 2.8 to 7.3

Two *Bam* HI fragments in each line were observed with the above gel (figure 4.2). The lower of these did not differ between lines and probably corresponded to the fragment expected from position 5.4 to 7.3 (figure 4.23). With the larger fragment line CM21 was assumed to represent the most common size. In line CM16 this fragment appeared larger and in lines CM13 and 30 to have increased in size further still, almost to the same size as the common *Eco* RI fragment of CM21. Since the fragments hybridizing to pBF17 from lines CM13, 16 and 30 were unusual with both *Eco* RI and *Bam* HI it would appear that the cause was a change in the total amount of DNA present. The two lines CM13 and 30 could not be distinguished with either enzyme.

Insertion I

A single clone homologous to pBF17 was isolated from line CM16 (λ6CM16). Although no restriction map was produced for this phage, the DNA was digested with *Bam* HI, *Sal* I, *Eco* RI, both *Bam* HI and *Sal* I and both *Sal* I and *Eco* RI. The fragments were transferred onto nitrocellulose after separation through a 0.4% agarose gel. Two *Bam* HI fragments were disclosed by pBF17 of 6.58 ±0.29 kb and 2.00 ±0.04 kb. With *Sal* I a single band of 5.23 ±0.16 kb and in the double digests one of 4.74 ±0.09 kb was observed (figure 4.6). *Eco* RI did not appear to cleave the DNA at any point. These fragments are consistent with an insertion, relative to the most common pattern, between the *Bam* HI site at position 5.4 and *Sal* I site at 4.3. The insertion does not appear to contain sites for *Bam* HI, *Sal* I or *Eco* RI and is approximately 4 kb long, estimated from the difference between the *Eco* RI fragment homologous to p56H8/C in lines CM16 and 19.

The phage λ6CM16 was unusual in that the fragments seen with several digests were not always repeatable. Purification of this clone by plaque hybridization (see Chapter 2) also repeatedly produced clones not homologous to pBF17. The reason for this may be that the inverted repeats of the heat shock gene are not replicated efficiently in the *E. coli* strain used. ~~Recombination between them would produce a deletion of the spacer region and therefore remove the~~

~~pBF17 homologous sequence.~~ This phenomenon has been observed previously when trying to clone the more common arrangement of the heat shock genes at 87A7, with only about 1 kb separating the two transcripts (Leigh Brown personal communication).

Insertions IIa and IIb

Both lines CM13 and 30 differed from the other lines in the same way as each other. In both, the *Eco* RI fragment homologous to p56H8/C was about 2 kb shorter and the larger of the two *Bam* HI fragments homologous to pBF17 was about 6 kb larger than expected. These sizes were estimated without λ DNA marker fragments. Although in neither case was a genomic clone of the region purified (see Appendix I) it appears that there has been an insertion of about 6 kb of DNA into the distal *Bam* HI fragment central to the two HS coding sequences. This DNA probably contains no *Bam* HI sites but has at least one *Eco* RI site.

Co-ordinates 8.6 to 24.3

For 27 lines an *Xho* I fragment homologous to p87A5 was observed. The size was estimated to be 10.16 kb from line CM25. One variant line was found: line CM21 (figure 4.3). In this line the observed fragment was 5.39 kb, a decrease of 4.71 kb. With *Xba* I and p87A/5 three lines, out of 31, were different from the common pattern of a single 16.61 kb fragment, as estimated from line CM17 (figure 4.4). The size of the fragment in each of lines CM4, 21 and 24 was estimated to be 20.24 kb, 21.74 kb and 20.15 kb respectively. Thus the difference in line CM21 appears to be an insertion event since both *Xho* I and *Xba* I fragments were affected.

Insertion III

A single phage clone, λ1CM21, homologous to p903a was isolated from line CM21 and a restriction map deduced. Hybridization of DNA fragments to p903a indicated that the DNA represented in this phage extended distally from the p903a homologous region only about 2 kb. Since the insertion responsible for the change in line CM21 reduces the *Xho* I fragment homologous to p87A/5 from about 10 kb to about 5.5 kb it must lie within 5.5 kb proximal to the *Xho* I site at position 8.6 (see figure 4.23). The single clone isolated does

not extend this far so the insertion was not represented in this phage.

Insertion IV

Two clones homologous to p903a were purified for line CM24. One of these, λ1CM24, was radioactively labelled and hybridized to one of the filters of *Bgl* II digested genomic DNA produced in the AS-C locus survey below. Many bands of a range of sizes from below 2 kb to more than 20 kb were produced. Most bands appeared to be the same size throughout the different lines, while a few were specific to different lines. Such a pattern is expected to be produced by sequences homologous to a mobile dispersed gene family. The insertion event proximal to the *Xba* I site at position 24.2 is therefore presumed to be due to a transposable element.

Insertion V

It is unlikely that the pattern produced with CM4 could be explained by the loss of a restriction site at either end of the altered *Xba* I fragment. Distal to this fragment the next *Xba* I site is within 2 kb, and proximally is at least 7 kb away, as estimated from the restriction map of λ1CM21. The similarity between lines CM4 and 24 suggests that both lines could contain similar inserts.

Co-ordinates 2.8 to 8.8

A *Bam* HI fragment homologous to p87A/5 was visible in 30 lines. This fragment was the same size in each line, 3.33 kb as estimated from line CM26. Since this fragment appeared normal in line CM21 the insertion in this line must be proximal to the *Bam* HI site at position 12.1 kb. In order to give the observed reduction in size of the *Xho* I fragment there must be an *Xho* I site in the insertion within 2 kb of this *Bam* HI site. Only 11 lines could be scored reliably for the larger of the two fragments expected to hybridize with pBF17. 10 lines appeared to have a fragment of 2.54 kb as estimated from line CM19. As with the *Eco* RI fragment homologous to p56H8/C the fragment from line CM22 appeared smaller, with a size of 2.39 kb, a decrease of 0.15 kb. This event must therefore be due to a small insertion-deletion event between the *Bam* HI sites at positions 2.8 and 5.4. This event is likely to be the same as insertion-deletion variant B observed by Leigh Brown (1983) and also by

lsh-Horowicz and Pinchin (1980).

Polymorphic *Pst* I Site

The *Pst* I fragment which hybridized to p56H8/C could be seen in 29 lines. In 8 lines the size was estimated to be 4.11 kb from line CM26. In the remaining lines the size was 3.49 kb, as estimated from line CM18 (figure 4.5). No variation in *Eco* RI fragments which correlated with the difference in the *Pst* I fragments was observed so insertion-deletion variation proximal to the *Eco* RI site at position 0.0 was not responsible. The difference of 0.62 kb agrees with the difference expected from a *Pst* I site polymorphism, reported previously, at position -0.8 (Leigh Brown 1983).

4.2.1.1. Summary

Thirty two third chromosomes were surveyed from the *Xba* I site at position 24.3 to the *Pst* I site at position -1.2 on the map shown in figure 4.23, a distance of about 25 kb. Within this region eight differences from the most common sequence arrangement were found in certain lines. Only one of these could be explained as the result of a restriction site change. The other seven differences appear to be due to the insertion, or deletion, of DNA within the region surveyed.

Eight lines, out of twenty nine surveyed, lacked the *Pst* I site at position -0.8. This had been expected because a previous survey (Leigh Brown 1983) found this site to be highly polymorphic in a different population. The other polymorphic restriction site observed in the previous survey was at a relatively low frequency. This site, the *Xho* I site at position 18.8, was present in all twenty seven lines examined in the current survey.

Line CM22 appeared to contain a deletion, relative to the other lines, of about 150 bp between the *Bam* HI sites at position 2.8 and 5.4, labelled B in figure 4.23. This is likely to be an example of the insertion-deletion variant B revealed by Leigh Brown (1983) previously. This sequence is known to be homologous to a similar sequence at the 87C1 heat shock locus (Mason *et al.* 1982). Only eleven lines were surveyed in sufficient detail to reveal the presence or absence of this sequence unambiguously.

Line CM16 contained an insertion, labelled I in figure 4.23 between the

Sal I site at position 4.2 and the *Bam* HI site at position 5.4. This insertion contained no sites at which the enzymes *Bam* HI, *Sal* I or *Eco* RI cut and was about 4 kb in size. Lines CM13 and 30 also contained insertions of about 6 kb of DNA within this region (IIa and IIb in figure 4.23). These insertions could not be distinguished and were between the two *Bam* HI sites at positions 2.8 and 5.4. The sequence inserted must contain at least one *Eco* RI site but probably no *Bam* HI site.

Insertion III, shown in figure 4.23 was found in line CM21. This event must lie between the *Bam* HI site at position 12.1 and 5.5 kb proximal to the *Xho* I site at position 8.6 and contain at least one *Xho* I site but probably no *Xba* I site. The insertions in lines CM4 and 24 could not be distinguished by size of the *Xba* I fragments in which they were found (IV and V in figure 4.23). Although no clones were obtained for line CM4, the insertion in line CM24 was shown to be homologous to a mobile, dispersed, repetitive family of sequences. These insertions must lie between the *Xho* I site at position 18.8 and the *Xba* I site at position 24.4

4.2.2. AS-C Locus

For the AS-C locus survey four enzymes were chosen with which to digest the DNA. *Xba* I and *Xho* I were known to cut infrequently, giving rise to relatively large, overlapping fragments. These two enzymes were useful in rapidly screening most of the region for insertion-deletion events. The enzymes *Bam* HI and *Bgl* II were chosen because they restrict the DNA of this region comparatively frequently. By using the four enzymes together on the same region, the insertion-deletion events found with the *Xba* I and *Xho* I fragments could be positioned more accurately with the *Bam* HI and *Bgl* II fragments. The smaller fragments could also reveal the presence of smaller insertion-deletion events than was possible with the larger fragments.

The positions of the sites at which these enzymes cut in the region surveyed are shown in figure 4.24. This map also shows the position of the restriction site variants found in the survey (indicated by line arrows) together with the two small, putative insertion-deletion events (triangles labelled A and B). The co-ordinates, in kilobases, have been retained from the map of Campuzano *et al.* (1985) from which this map was compiled (also from Parkhurst and Corces 1986; Cabrera C.V. personal communication). Transcripts

Figure 4.1

(Top) *Eco* RI digested DNA from the heat shock locus survey probed with p56H8/C. Line numbers are given above the photograph. The central smear is degraded λ marker fragment DNA.

Figure 4.2

(Centre right) The four left-most lanes are *Eco* RI digested DNA from the lines CM13, 30, 16 and 21 from left to right. The four left-most tracks are *Bam* HI digests of DNA from the same lines in the same order. The fragments were revealed by pBF17. The arrows to the left indicate the three different positions to which *Eco* RI fragments in the different lines migrated. The four left arrows show the positions of the *Bam* HI fragments.

Figure 4.3

(Centre-left) *Xho* I digested DNA from the heat shock locus survey probed with p87A5.

Figure 4.4

(Bottom-left and right) *Xba* I digested DNA from the heat shock survey probed with p87A/5.

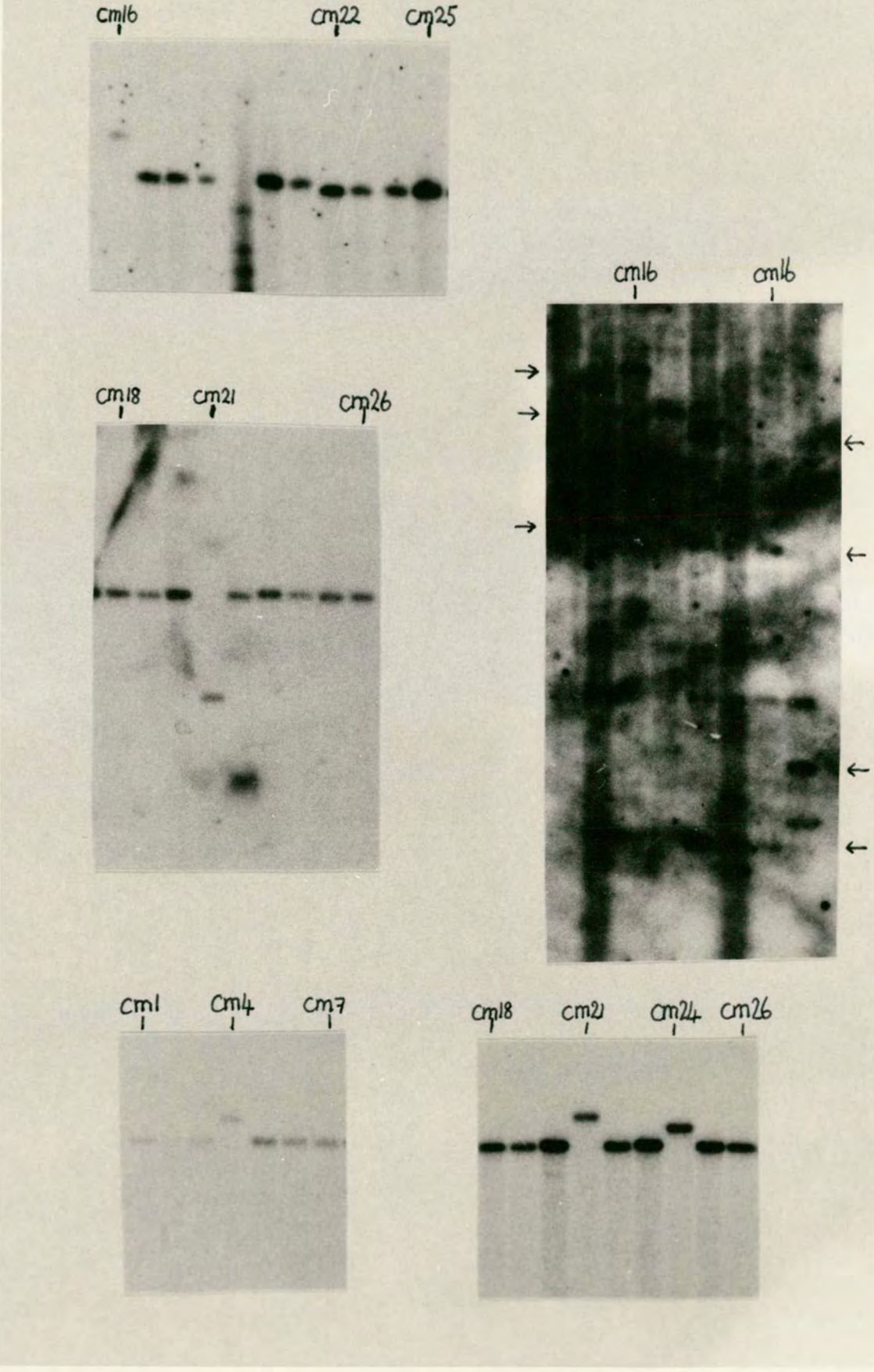

cm16
cm22
cm25
cm18
cm21
cm26
cm16
cm16
cm1
cm4
cm7
cm18
cm21
cm24
cm26

Figure 4.5

(Top-left) *Pst* I digested DNA from the heat shock locus survey probed with p56H8/C.

Figure 4.6

(Centre-right) λ6CM16 DNA digested with *Bam* HI, *Bam* HI and *Sal* I, *Sal* I, *Sal* I and *Eco* RI and *Eco* RI. The filter was probed with pBF17.

Figure 4.7

(Centre) *Bgl* II digested DNA from the AS-C locus survey probed with pASC53R1.

Figure 4.8

(Bottom) *Bgl* II digested DNA from the AS-C locus survey which was probed with pASC31P4. Only fragments 2 and 3 are shown.

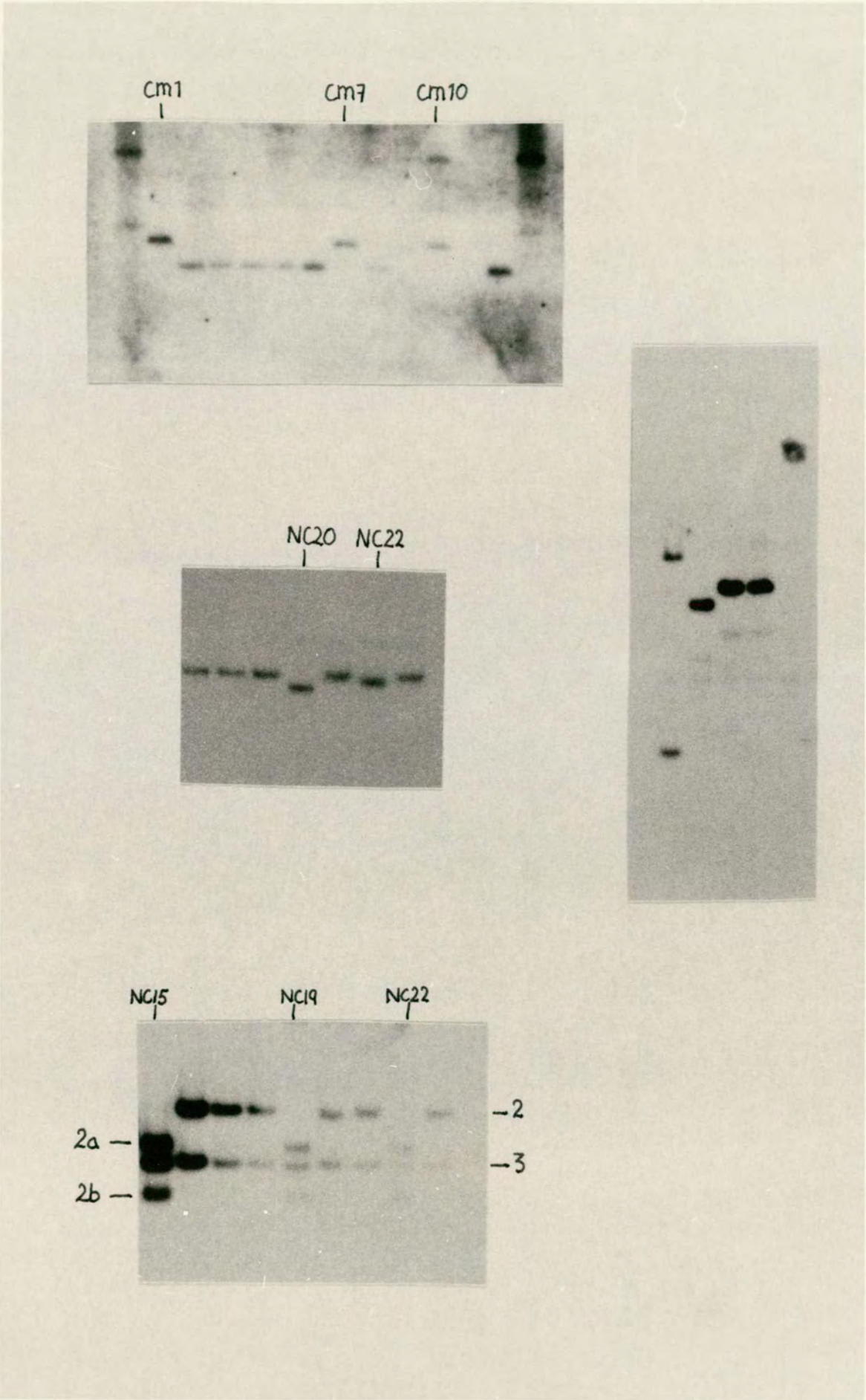

Cm1
Cm7
Cm10
NC20
NC22
NC15
NC19
NC22
2a
2b
2
3

Figure 4.9

(Left-centre and right) *Bgl* II digested DNA from the AS-C locus survey probed with pASC94R1 and pASC94R4. Only bands 1 and 2 are shown.

Figure 4.10

(Centre) AS above but showing fragments 3, 4 and 5.

Figure 4.11

(Bottom-left) *Bgl* II digested DNA from the AS-C locus survey probed with λsc112 showing fragments 4 and 5.

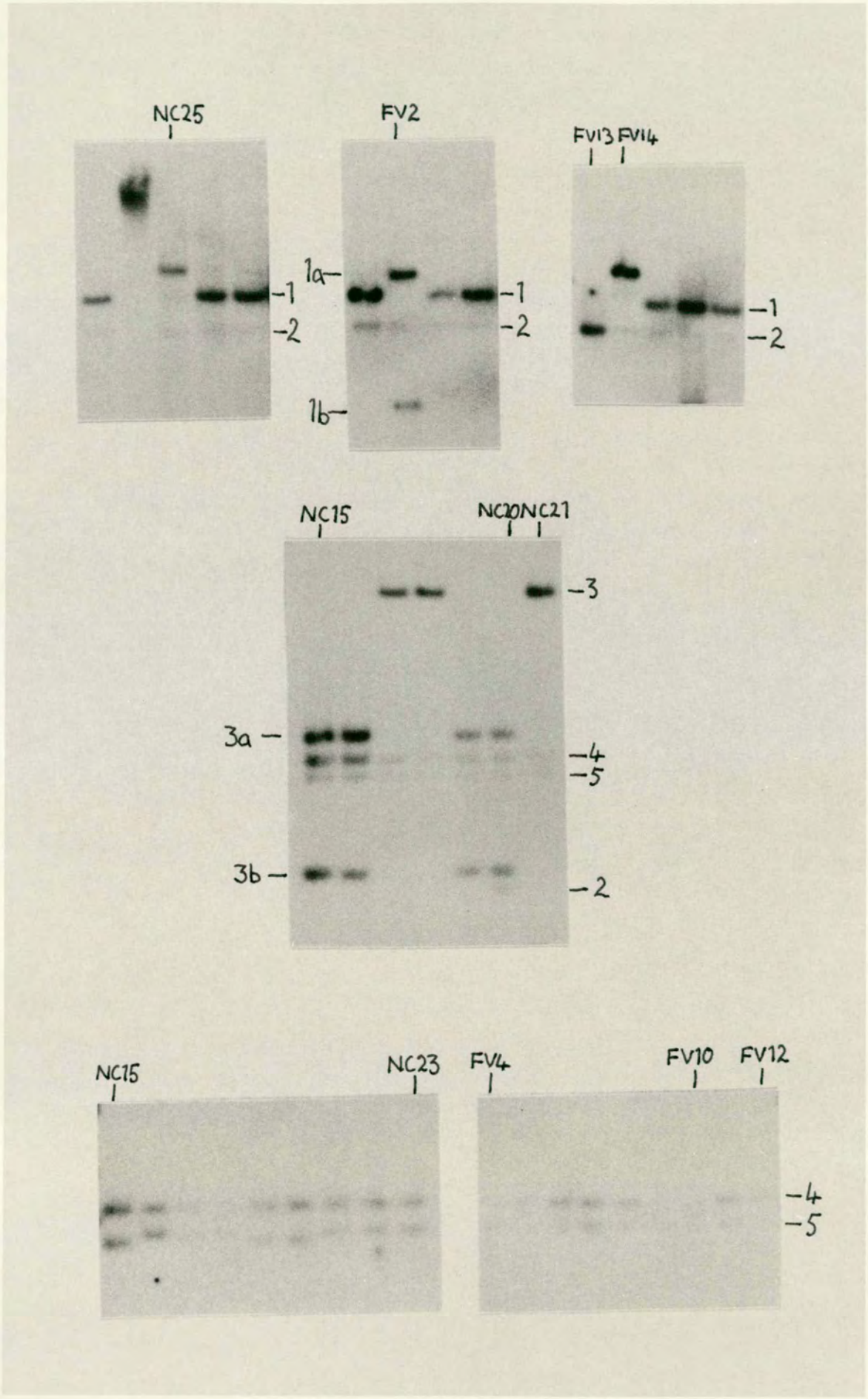

NC25
FV2
FV13
FV14
1a
1b
-1
-2
NC15
NC20
NC21
-3
3a
3b
-4
-5
-2
NC15
NC23
FV4
FV10
FV12
-4
-5

Figure 4.12

(Top-left and right) *Bam* HI digested genomic DNA from the AS-C locus survey which was probed with pASC53R1.

Figure 4.13

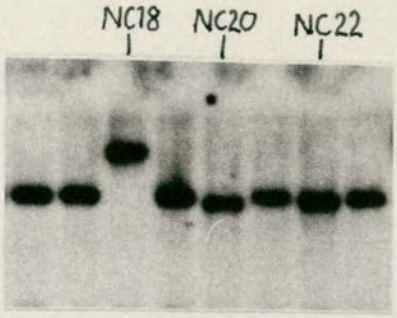
(Centre) *Bam* HI digested genomic DNA from the AS-C locus survey probed with λsc31 and λsc112.

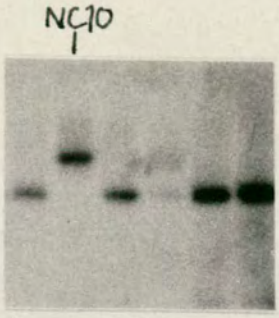
Figure 4.14

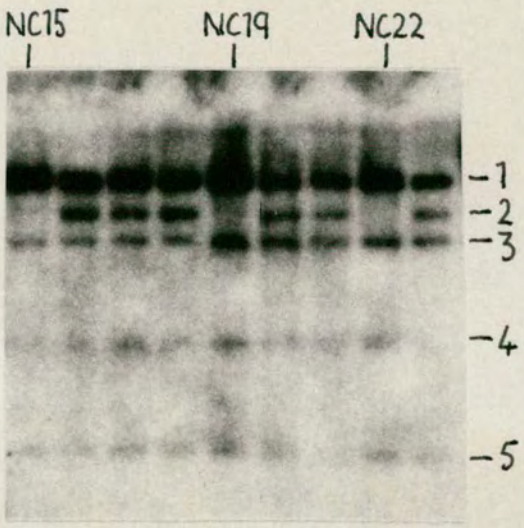
(Bottom-left) *Bam* HI digested genomic DNA which was probed with pASC94R1 and pASC94R4.

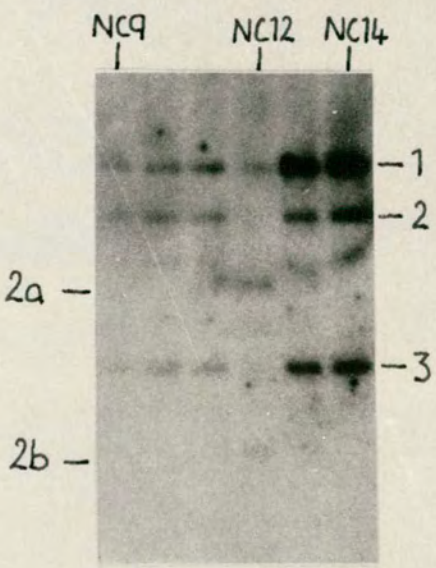
Figure 4.15

(Bottom-right) *Xba* I digested genomic DNA probed with λsc53 and λsc17.


NC18
NC20
NC22


NC10


NC15
NC19
NC22
1
2
3
4
5


NC9
NC12
NC14
1
2
2a
3
2b

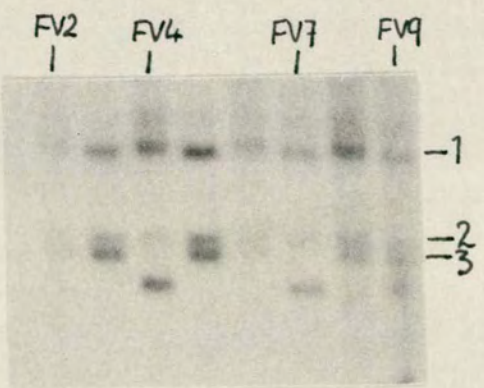

FV2
FV4
FV7
FV9
1
2
3

Figure 4.16

(Top-left) *Xba* I digested genomic DNA probed with λsc53 and λsc22.

Figure 4.17

(Top-right) *Xba* I digested DNA probed with pASC94R1 and pASC94R4.

Figure 4.18

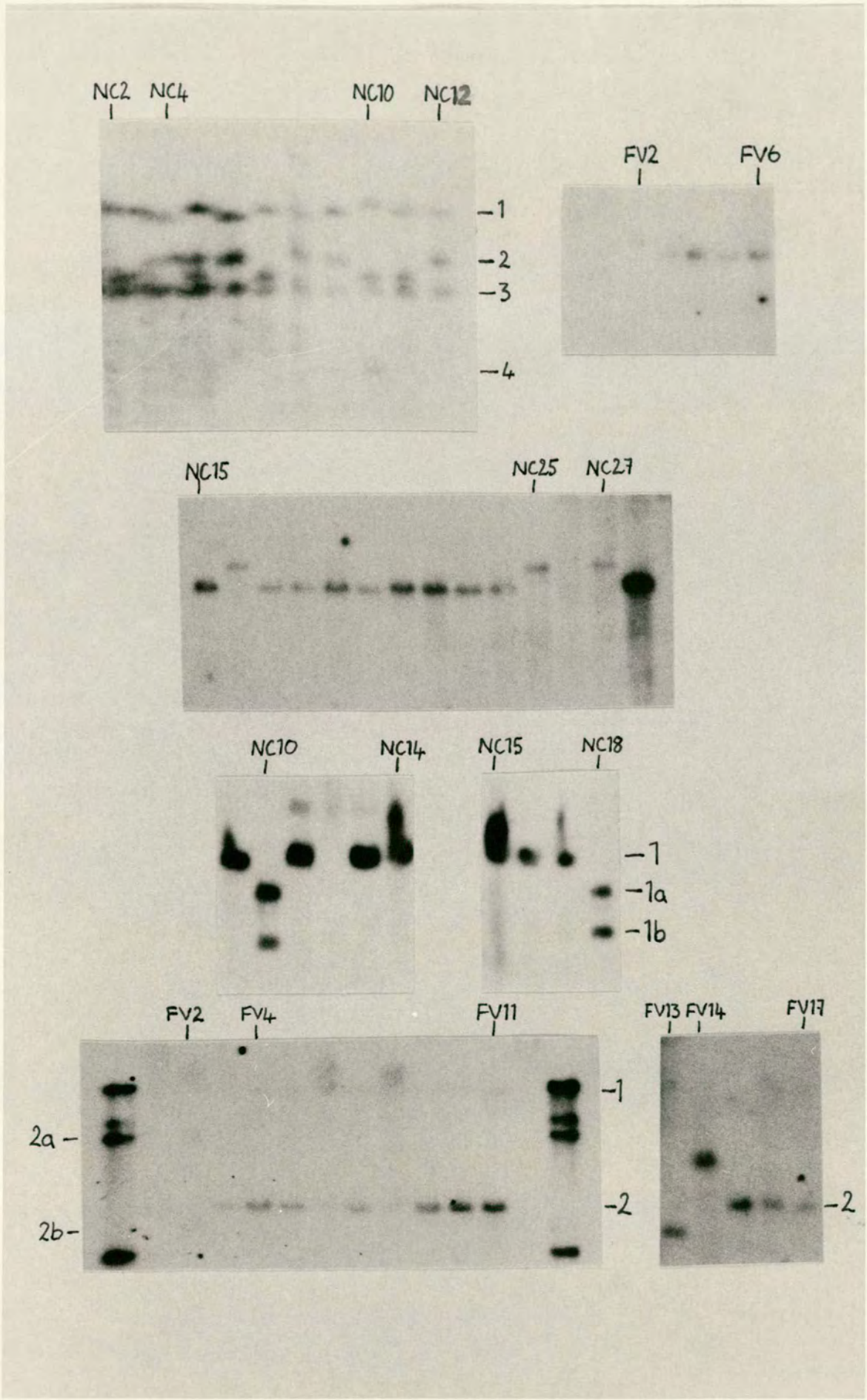
(Above-centre) *Xba* I digested DNA probed with pASC101R7.

Figure 4.19

(Below-centre, left and right) *Xho* I digested DNA probed with pASC53R1.

Figure 4.20

(Bottom-left and right) *Xho* I digested genomic DNA probed with pASC94R1 and pASC94R4.


NC2
NC4
NC10
NC12
-1
-2
-3
-4
FV2
FV6
NC15
NC25
NC27
NC10
NC14
NC15
NC18
-1
-1a
-1b
FV2
FV4
FV11
FV13
FV14
FV17
2a-
2b-
-1
-2
-2

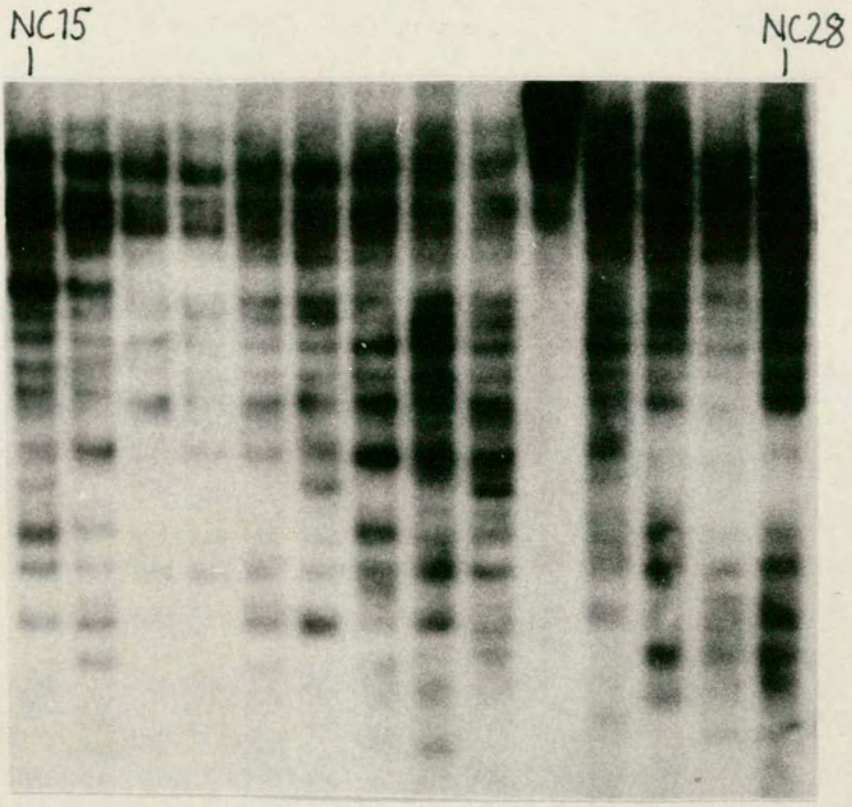
Opposite page 61

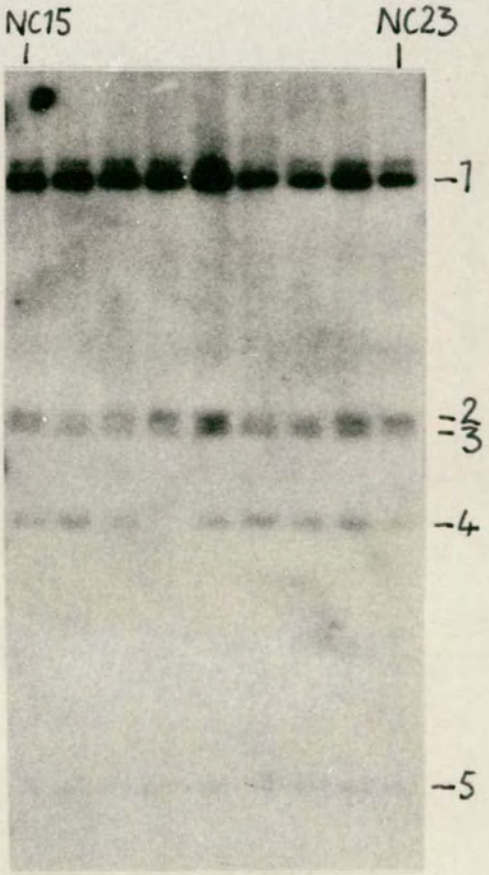
Figure 4.21

(Top) *Bgl* II digested genomic DNA probed with λ3NC22. The many bands seen indicate that the sequence homologous to the probe lie at different sites throughout the genome. Bands are seen to vary between lines in a manner characteristic of transposable elements.

Figure 4.22

(Bottom) *Bam* HI digested genomic DNA from the AS-C locus survey probed with λ1FV2.

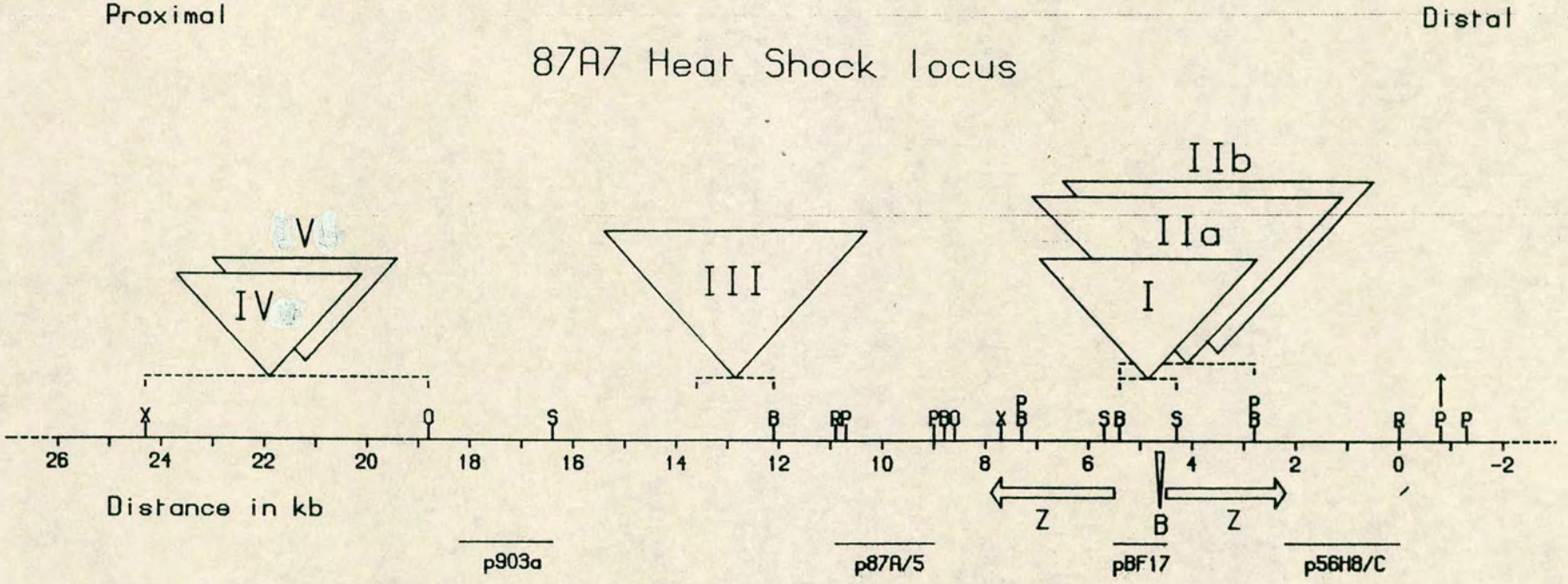

NC15
NC28


NC15
NC23
-1
=2/3
-4
-5

Opposite page 62

Figure 4.23

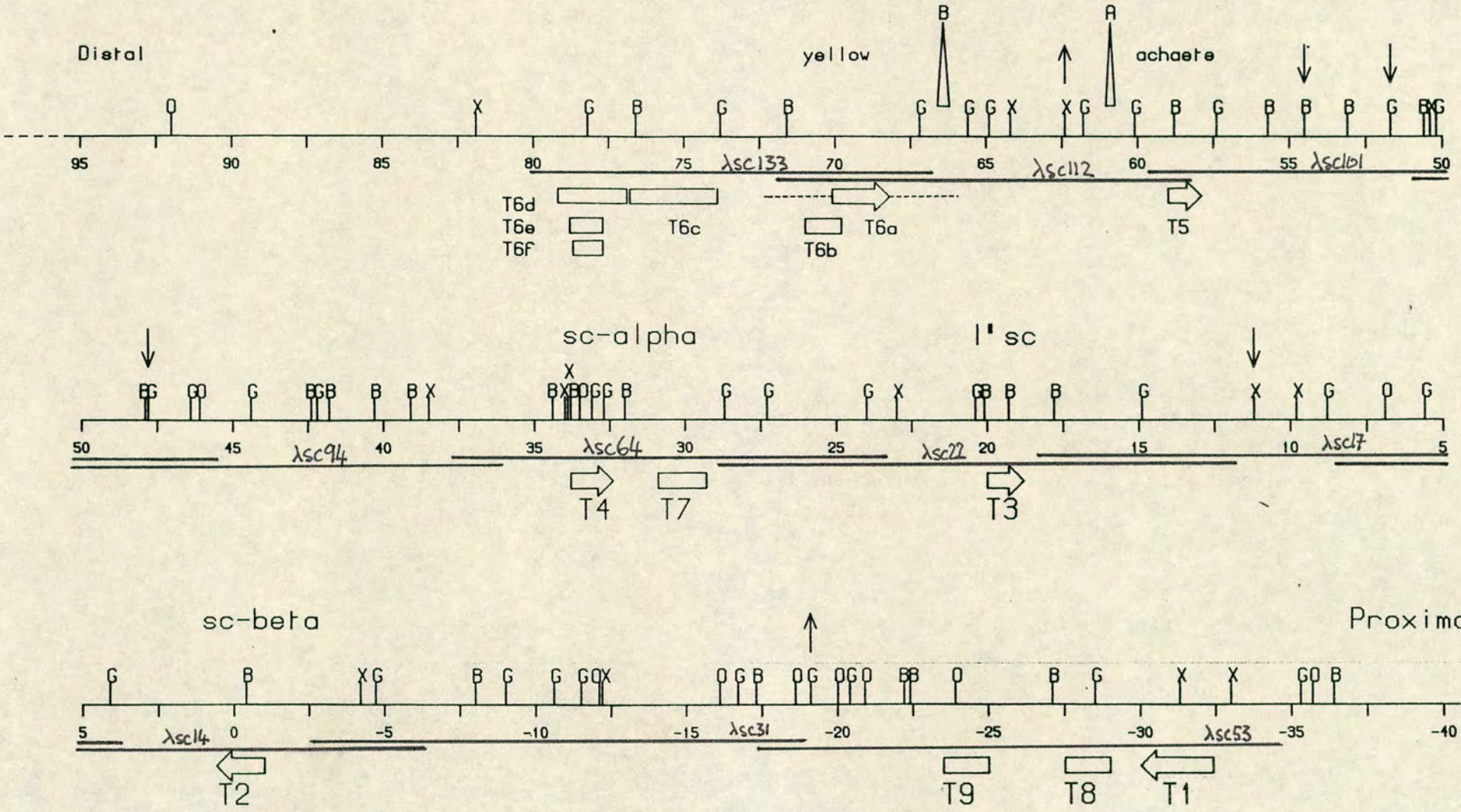
Molecular map of the 87A7 heat shock locus compiled from Ish-Horowicz and Pinchin (1980); Leigh Brown and Ish-Horowicz (1981); Leigh Brown (1983) and personal communication. The arrows labelled 'Z' represent the transcripts at this locus. The solid lines below the map indicate the positions of the four probes used to survey this region. Restriction sites are lettered as follows, B : *Bam* HI, O : *Xho* I, P : *Pst* I, R : *Eco* RI, S : *Sal* I and X : *Xba* I. Six insertion events are shown above the map with the uncertainty of their exact location indicated by the dashed lines, the small insertion- deletion event is labelled 'B'. The *Pst* I site at position -0.8 kb was polymorphic.


Proximal
Distal
87A7 Heat Shock locus
V
IV
III
IIb
IIa
I
X
O
S
B
R
P
26
24
22
20
18
16
14
12
10
8
6
4
2
0
-2
Distance in kb
Z
B
Z
p903a
p87A/5
pBF17
p56H8/C

Opposite page 63

Figure 4.24

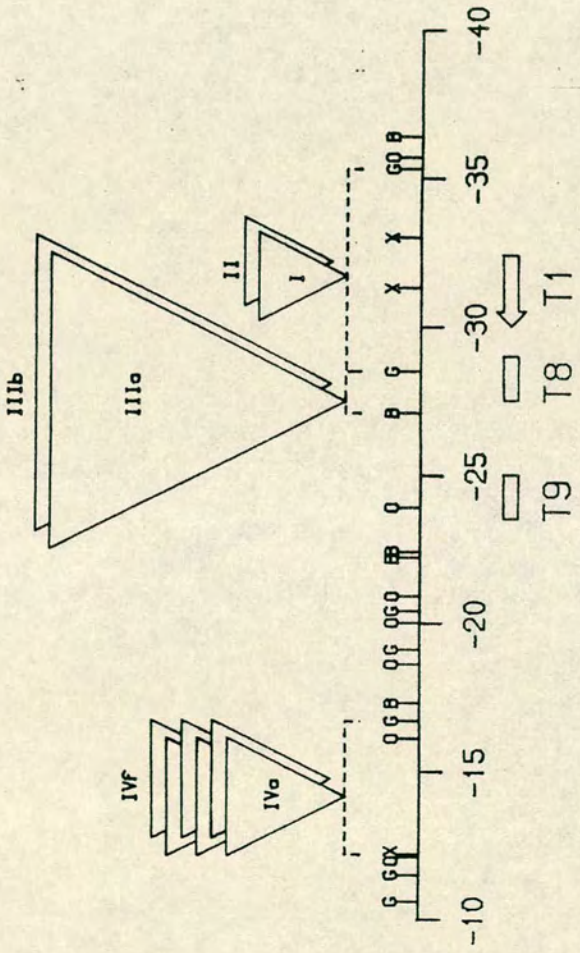
Restriction map of the AS-C locus compiled from Campuzano *et al.* 1985; Parkhurst and Corces 1986; Cabrera C.V. personal communication. restriction sites are lettered as follows B : *Bam* HI, G : *Bgl* II, O : *Xho* I and X : *Xba* I. transcripts are shown as open arrows to indicate the direction of transcription or when this is not known as open boxes. The transcripts are labelled after Campuzano *et al.* 1986. The position of the genetically identified loci are given above the map. The line arrows indicate where a restriction site was gained or lost relative to the map of Campuzano *et al.* 1985. The two triangles labelled A and B indicate the position of the small insertion-deletion events found.


Distal
yellow
achaete
95
90
85
80
75
λsc133
70
65
λsc112
60
55
λsc101
50
T6d
T6e
T6f
T6c
T6b
T6a
T5
sc-alpha
1' sc
50
45
λsc94
40
35
λsc64
30
25
λsc22
20
15
10
λsc17
5
T4
T7
T3
sc-beta
Proximal
5
λsc14
0
-5
-10
-15
λsc31
-20
-25
-30
λsc53
-35
-40
T2
T9
T8
T1

Opposite page 64

Figure 4.25

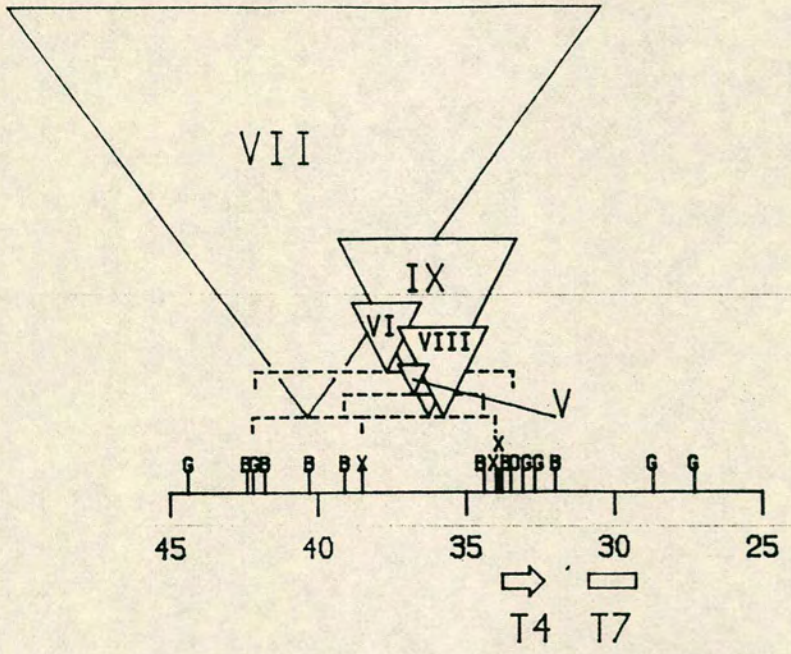
Molecular map of the proximal part of the AS-C locus region showing the positions of the large insertion events (open triangles). The transcripts and restriction sites are as given in figure 4.24. The dotted line indicates the region of uncertainty for the point of insertion.


IIIb
IIIa
II
I
IVf
IVa
-10
-15
-20
-25
-30
-35
-40
T9
T8
T1

Opposite page 65

Figure 4.26

Molecular map of the distal part of the AS-C locus region showing the positions of the large insertion events (open triangles). The transcripts and restriction sites are as given in figure 4.24. The dotted line indicates the region of uncertainty for the point of insertion.


VII
IX
VI
VIII
V
X
G
B
X
45
40
35
30
25
T4
T7

are represented by open arrows showing the direction of transcription, or boxes when the direction is unknown.

The transcripts T1 to T6 are those of Campuzano *et al.* (1985). The transcript T6a has been shown to be the *yellow* gene product (Parkhurst and Corces 1986). Transcripts T7 to T9 are expressed only in larvae, unlike the other transcripts (Cabrera personal communication). Full details have not yet been published. Several of the transcripts share homology with each other, these are T3, T4, T5, T8 and T9, (Cabrera personal communication) which are thought to be involved in functions required for the correct expression of the *achaete-scute* wild type micro- and macrochaetae phenotype. The approximate positions along the map of the genetically identified loci are indicated by the name of those loci above the map. The large insertion-deletion events found are shown in figures 4.25 and 4.26 as triangles above the map. The co-ordinates are the same as those in figure 4.24. Dashed lines represent the region of uncertainty for the points of insertion.

Only a brief account of the variation found is presented in this section. A more detailed description of each of the autoradiographs is presented in Appendix I.

4.2.2.1. Genomic *Bgl* II Fragments

The estimated sizes and standard errors for the fragments hybridizing to the various probes are given in table 4.3.

Co-ordinates -35.3 to -16.7

A single fragment of 6.80 kb, as estimated from line NC8, was homologous to pASC53R1. This fragment in lines NC20 and 22 was slightly smaller than this at 6.17 kb and 6.39 kb respectively (figure 4.7). With λsc53 two large fragments were visible. The second largest fragment was the same as that which had hybridized to pASC53R1. The largest fragment was the next most distal, was 8.54 kb (line NC2). The size of this fragment in lines NC10 and 18 was 11.63 kb and 11.05 kb respectively. These sizes are not significantly different. Two smaller bands appeared which were never both present in the same line, suggesting that they may be mutually exclusive. In 10 lines the fragment was 3.86 kb (line NC3) and in 10 lines was 2.43 kb (line NC2). The

Opposite page 67

Table 4.3

Estimated fragment sizes and standard errors for the *Bgl* II fragments revealed by the è probes used to survey the AS-C locus region. The size estimates were generated by the 'spline' programme (see Chapter 3).

Probe	Fly Line Scored	Fragment Number	Fragment Size/kb	Standard Error/kb
pASC53R1	NC8*	1	6.80	0.08
	NC20	1	6.17	0.05
	NC22	1	6.39	0.12
λsc53	NC2*	1	8.54	0.29
		2	6.93	0.13
		3	2.43	0.04
	NC3	3	3.86	0.03
	NC10	1	11.63	0.24
	NC18	1	11.05	0.33
	NC20	2	6.18	0.23
	NC22	2	6.43	0.07
pASC31P4	FV16*	1	9.08	0.37
		2	5.32	0.10
		3	4.24	0.11
		4	1.82	0.05
		5	0.83	0.02
	NC4	2a	4.59	0.07
		2b	3.62	0.02
λsc22	NC1	1	6.05	0.13
		2	5.45	0.22
		3	3.79	0.05
		4	3.57	0.06
		5	1.15	0.03
pASC94R1 + pASC94R4	NC17*	1	9.18	0.14
		2	7.34	0.16
		3	3.76	0.05
		4	2.13	0.01
		5	2.03	0.09
	NC25	1	11.50	0.43
	FV14	1	13.84	0.59
	FV2	1a	11.18	0.29
		1b	5.00	0.10
	NC20	2	1.47	0.02
	NC16	3a	2.35	0.09
		3b	1.55	0.03
pASC101R7	NC1	1	2.86	0.05
λsc112	NC7*	1	6.64	0.07
		2	3.24	0.04
		3	2.84	0.04
		4	1.73	0.01
		5	1.64	0.04
	FV10	4	1.69	0.01
	FV18	4	1.69	0.03
	NC8	5	1.61	0.01
pASC133R1	NC25	1	6.61	0.07
		2	1.75	0.02

sizes of these would be consistent with the map shown in figure 4.24 if the *Bgl* II site at position -19.1 were absent in the lines with the larger fragment. This polymorphism had been reported previously by Campuzano *et al.* (1985).

Co-ordinates -16.7 to 4.1

With pASC31P4 the second largest of five fragments was 5.32 kb long (line FV16). In six of the NC lines this fragment was absent being replaced by two of 4.59 kb and 3.62 kb (line NC15, 19 and 22, see figure 4.8). The explanation seems to be an insertion of DNA in each of these lines within the 5.32 kb fragment since there is an increase in the total amount of DNA present. A restriction site change at the proximal end of this fragment is unlikely since one would affect the small fragments which hybridized to λsc53. The variation in these fragments appears to be independent of that in the fragment here. A second enzyme was required to resolve this. The other fragments were 9.08 kb, 4.24 kb, 1.82 kb and 0.83 kb (FV16).

Co-ordinates 8.8 to 27.3

All five fragments which were homologous to λsc22 were identical in size in each line surveyed, 6.05 kb, 5.45 kb, 3.79 kb, 3.57 kb and 1.15 kb estimated from line NC1.

Co-ordinates 33.1 to 57.4

The most variation was revealed by a mixture of pASC94R1 and pASC94R4. The five bands seen in the most common pattern were estimated to be 9.18 kb, 7.34 kb, 3.76 kb, 2.13 kb and 2.03 kb (line NC17). The largest fragments in lines NC25 and FV14 were larger than the rest with sizes of 11.50 kb and 13.34 kb respectively. In line FV2 this fragment was again larger than the rest, with a size of 11.18 kb, and an extra band of 5.00 kb appeared. The 9.18 kb fragment in line FV13 appeared to have co-migrated with the next largest fragment since that band appeared more intense than usual (figure 4.9). Restriction site polymorphisms within this fragment would not explain the variants found easily. It would seem that these represent four different insertion-deletion events.

In twelve lines (figure 4.10) the 3.76 kb fragment seemed to be replaced by two of 2.35 kb and 1.55 kb (line NC16). Since the sum of the two fragments

in line NC16 is approximately that of the single fragment in line NC17, the representative of the common type, the explanation for the difference in band pattern is probably a polymorphic site present in line NC16 at position 48 (figure 4.24) and absent from line NC17. In one line, NC20, the 7.34 kb band revealed by the mixture of probes used above was absent and a new, smaller band of 1.47 kb appeared.

Co-ordinates 57.4 to 60.1

With pASC101R7 none of the lines surveyed, including NC20, were unusual. The size of this fragment was 2.86 kb (line NC1). Since the next most proximal fragment to the missing one of line NC20 and the next most distal, the one homologous to pASC101R7, were both not unusual, the 1.47 kb fragment was probably derived from the 7.34 kb one by the appearance of a *Bgl* II site at position 51.7 (figure 4.24).

Co-ordinates 57.4 to 73.8

Five fragments in each line were homologous to λsc112 and had sizes of 6.64 kb, 3.24 kb, 2.84 kb, 1.73 kb and 1.64 kb (line NC7). In two FV lines the 1.73 kb fragment appeared to be slightly smaller, with a size of 1.69 kb (line FV10). In eight lines from the two populations the 1.64 kb fragment was slightly smaller at 1.61 kb (line NC8, see figure 4.11). If either of these changes were due to the introduction of a new restriction site, the extra fragments, approximately 40 bp and 30 bp long respectively, would migrate beyond the end of the gels used and therefore not be observed. The fragments of the other enzymes in this region were too large to reveal variation of this size. The possibility that each is due to a small insertion-deletion event can not be ruled out.

Co-ordinates 67.2 to 73.8

Finally, the two fragments homologous to pASC133R1 were the same size in each of the lines where they were observed. The size estimates were 6.61 kb and 1.75 kb (line NC25).

4.2.2.2. Genomic *Bam* HI Fragments

The fragment sizes revealed by each of the probes used, and their estimated standard errors are given in table 4.4.

Co-ordinates -36.4 to -28.5

With pASC53R1 the single *Bam* HI fragment disclosed was 9.34 kb long (line NC21). The four lines NC10, 18, 20 and 22 were unusual (figure 4.12) with sizes of 19.02 kb, 18.98 kb, 8.86 kb and 9.11 kb respectively. All four of these lines were variant in the *Bgl* II fragments disclosed in this region so the differences must be due to insertion-deletion events. The fragment sizes of lines NC10 and NC18 were not significantly different.

Co-ordinates -17.3 to -0.4 and 55.7 to 76.6

With a mixture of λsc31 and λsc112 five bands appeared of 13.51 kb, 9.45 kb, 7.95 kb, 5.01 kb and 3.65 kb. The only variation observed was in six lines where the second largest fragment appeared to co-migrate with the 13.51 kb fragment (figure 4.13). These six lines were those also observed to have unusual *Bgl* II fragments homologous to pASC31P4, the region the second largest fragment is expected to come from. The variation in these lines is then due to an insertion event in each line, rather than a change in a restriction site.

Co-ordinates 34.4 to 53.1

pASC94R1 and pASC94R4 produced a pattern of six bands. These bands were 5.86 kb, 4.71 kb, 2.81 kb, 1.36 kb, 1.24 kb and 0.85 kb (line NC16). Only a single line displayed a pattern different from this. In line NC12 the second largest fragment was replaced by two of 3.54 kb and 2.12 kb (figure 4.14). The sum of these is 0.9 kb larger than the 4.71 kb fragment. This difference is a significant one and so is probably due to an insertion event. The insertion events found in the *Bgl* II fragments from this region do not affect these *Bam* HI fragments and so must lie proximal to them. With pASC101R5 two fragments were observed of 3.39 kb and 2.58 kb (line NC11). In line NC3 the lower fragment was absent and apparently replaced by one of 1.16 kb [illegible] No difference in this region was seen with the *Bgl* II fragments so a novel *Bam* HI site seems to be present in this line.

Opposite page 71

Table 4.4

Estimated fragment sizes and standard errors for the *Bam* HI fragments revealed by the e probes used to survey the AS-C locus region. The size estimates were generated by the 'spline' programme (see Chapter 3).

Probe	Fly Line Scored	Fragment Number	Fragment Size/kb	Standard Error/kb
pASC53R1	NC21*	1	9.34	0.30
	NC10	1	19.02	1.28
	NC18	1	18.98	1.99
	NC20	1	8.86	0.20
	NC22	1	9.11	0.21
λsc31 + λsc112	NC18	1	13.51	0.85
		2	9.45	0.29
		3	7.95	0.19
		4	5.01	0.08
		5	3.65	0.05
pASC94R1 + pASC94R4	NC16*	1	5.86	0.14
		2	4.71	0.05
		3	2.81	0.03
		4	1.36	0.05
		5	1.24	0.01
		6	0.85	0.01
	NC12	2a	3.54	0.03
		2b	2.12	0.02
pASC101R5	NC11*	1	3.39	0.05
		2	2.58	0.03
	NC3	2	1.16	0.02

4.2.2.3. Genomic *Xba* I Fragments

The fragment sizes from these digests and their standard errors are given in table 4.5.

Co-ordinates -33.0 to -12.2

Two fragments were visible with pASC53R1 as a probe. The smaller fragment of 1.85 kb (line FV20) was invariant throughout the samples surveyed. The larger fragment was also homologous to λsc53 and pASC31P4. Using these three probes this fragment was surveyed in all the lines except for NC26. Three different mobilities were apparent. The size of this fragment in all the FV lines was estimated to be 19.68 kb (FV20). In the NC samples the most common form was assumed to be the same as this and had an estimated size of 20.46 kb (NC7). These two size estimates are not significantly different although no direct comparison could be made because lines representing each type were not present on the same gel. This fragment in lines NC10 and 18 (figure 4.16) was slightly larger with sizes of 21.43 kb and 22.17 kb, which are not significantly different from each other. In these lines new fragments of 7.17 kb and 7.50 kb appeared. These fragments are also not significantly different in size. In the six NC lines in which the *Bgl* II fragments from this region migrated unusually this *Xba* I fragment was slightly smaller than the common NC form with a size of 19.44 kb (NC4). The differences in this fragment can be interpreted with the same insertion events found in this region with the previous two enzymes. All eight insertions found in this region must contain at least one *Xba* I site.

Co-ordinates -31.3 to 33.9

With λsc53 and λsc17 as probes three bands appeared. The largest of these was the same as the larger of the two discussed above. The second largest was invariant throughout the lines surveyed but its size could not be estimated because no λ marker fragments could be seen in these filters. The smallest fragment was also homologous to pASC17B1 and λsc22. The most common size of this fragment was 14.29 kb (line NC4) but in eleven lines its size was estimated to be 12.72 kb (line NC10). Since no variation was observed in the *Bgl* II fragments from this region it appears that there was a polymorphic *Xba* I site about 1.5 kb from one end of the 14.29 kb fragment.

With λsc53 and λsc22, apart from two fragments which hybridized with others of the previous probes, a third, of 11.31 kb (line NC4) was observed. All the lines appeared to have the same form of this fragment. This was also true of the smaller of the two fragments homologous to pASC31P4 which had a size of 8.77 kb.

Co-ordinates 38.5 to 50.4

Of the single fragment surveyed with pASC94R1 and pASC94R4 only one line was unusual. In line FV2 this fragment was estimated to be 14.34 kb whereas in all the other lines it was 12.95 kb (line FV3). The reason for this difference was probably the same as for the unusual *Bgl* II fragment in this region. The insertion event responsible must be within both the *Bgl* II fragment and *Xba* I fragment affected. Since three of the other four insertion-deletion events in this region do not affect this *Xba* I fragment they must lie proximal to it. Line NC25 was not surveyed for this fragment.

Co-ordinates 50.4 to 64.2

The single fragment homologous to pASC101R7 was unusual in four lines. In these lines the size was estimated to be 14.72 kb (line NC16) instead of the more common 12.68 kb (line NC15). No changes in either the *Bam* HI or *Bgl* II fragments coincided with this difference, suggesting a restriction site change was responsible. The next most distal fragment to the one considered here was expected to be 1.8 kb. This would migrate beyond the end of the gels as the conditions of electrophoresis used were chosen for separation of large DNA fragments. The difference between the two forms of the fragment here is almost exactly the size of this smaller fragment. It seems likely then that the *Xba* I site at the distal end of the 12.68 kb fragment was absent in four of the lines.

Co-ordinates 64.2 to 81.9

The most distal *Xba* I fragment observed was homologous to both λsc112 and pASC133R1. This fragment was invariant throughout the lines surveyed with a size of 17.74 kb (line FV19).

Opposite page 74

Table 4.5

(Top) Estimated fragment sizes and standard errors for the *Xba* I fragments revealed by the e probes used to survey the AS-C locus region. The size estimates were generated by the 'spline' programme (see Chapter 3).

Table 4.6

(Bottom) Estimated fragment sizes and standard errors for the *Xho* I fragments revealed by the e probes used to survey the AS-C locus region. The size estimates were generated by the 'spline' programme (see Chapter 3).

Probe	Fly Line Scored	Fragment Number	Fragment Size/kb	Standard Error/kb
pASC53R1	NC7*	1	20.46	0.54
	FV20	1	19.68	0.52
		2	1.85	0.03
	NC4	1	19.44	0.25
	NC10	1	21.43	0.56
	NC18	1	22.17	0.76
λsc53 + λsc22	NC4*	2	14.29	0.47
		3	11.31	0.15
	NC10	2	12.72	0.32
		1b	7.17	0.28
	NC18	1b	7.50	0.23
pASC31P4	NC9	2	8.77	0.30
λsc94	FV3*	1	12.95	0.36
	FV2	1	14.34	0.24
pASC101R7	NC15*	1	12.68	0.18
	NC16	1	14.72	0.30
pASC133R1	FV19	1	17.74	0.25

Probe	Fly Line Scored	Fragment Number	Fragment Size/kb	Standard Error/kb
pASC53R1	NC7*	1	11.78	0.07
	NC10	1a	9.89	0.19
		1b	7.74	0.22
λsc53 + λsc17	NC9	1	26.34	0.36
		2	19.00	0.35
pASC94R1 + pASC94R4	FV4*	1	46.36	2.57
		2	12.32	0.20
	FV2	2a	22.61	1.90
		2b	10.95	0.22
	FV13	2	10.38	0.15
	FV14	2	18.23	0.78

4.2.2.4. Genomic *Xho* I Fragments

The sizes and estimated standard errors of these fragments are given in table 4.6.

Co-ordinates -35.7 to -23.9

A single fragment was homologous to pASC53R1, commonly with a size of 11.78 kb. The two lines NC18 and 33 were missing this fragment and had gained two of 9.89 kb and 7.74 kb (NC10). The two insertion events producing the alteration in both the *Bgl* II and *Xba* I must lie distal to the most proximal *Bgl* II fragment surveyed since this fragment was unaffected. Because two *Xho* I fragments appear in these lines the probe pASC53R1 must span one of the *Xho* I sites producing these. This probe lies entirely within the proximal *Bgl* II fragment so the novel *Xho* I site must have arisen at a distance from the insertion which was into the next most distal *Bgl* II fragment.

Co-ordinates -35.7 to -23.9 and -12.1 to 33.5

With both λsc17 and λsc53 as probes three fragments appeared, the lower one of which was that which hybridized to pASC53R1. The upper two fragments were invariant in size where surveyed, with sizes of 26.43 kb and 19.00 kb (line NC9).

Co-ordinates 33.5 to 92.0

Of the two fragments hybridizing to pASC94R1 and pASC94R4 the larger was estimated to be 46.36 kb (line FV4). The lower fragment was estimated to be 12.32 kb (line FV4) in the most common form. Three lines were different from this. In line FV2 two fragments of 10.85 kb and 22.61 kb appeared in place of the 12.32 kb one representing an increase of 21 kb. This change was probably caused by the same insertion responsible for the alteration of the *Bgl* II and *Xba* I fragments. In line FV13 this fragment was smaller with a size of 10.38 kb, a decrease of about 2 kb. The *Bgl* II fragment affected in this region was shorter than usual by about 1.8 kb. These decreases are not significantly different. Since these are the only two fragments affected by this lesion it is unclear whether the change is a deletion of about 2 kb or an insertion containing restriction sites for both *Bgl* II and *Xho* I close to each other and giving fragments 2 kb shorter than expected. Deletions of more than

a few hundred nucleotides have not been reported in this kind of survey (Langley *et al.* 1982; Leigh Brown 1983; Aquadro *et al.* 1986) so it would seem more likely to be an insertion event. Finally line FV14 displayed a fragment of 18.23 kb instead of 12.32 kb. This change appeared to be due to the same insertion event as the difference in *Bgl* II fragments observed in this region. Line NC25 was not surveyed for these fragments.

4.2.2.5. Cloned Genomic Sequences

To investigate the insertions found in the AS-C region libraries of cloned genomic DNA were prepared from each line. The details of these and the purification of specific clones are detailed in Appendix I.

The two lines NC10 and NC18 contained insertions into the second most proximal *Bgl* II fragments screened (IIIa and IIIb figure 4.26). These insertions could not be distinguished on the basis of the genomic DNA fragments analysed. A λ clone from each of these lines was used to probe filters from *Bgl* II digested genomic DNA gels. With λ2NC10 three fragments were disclosed with sizes of 14.38 kb, 3.39 kb and 2.99 kb. λ1NC18 revealed two fragments which corresponded exactly to the larger of these two bands. These fragments did not correspond to any expected from the AS-C locus map (figure 4.24) and so seem to have come from elsewhere in the genome. The similarity of the genomic fragments homologous to both clones suggests that the same sequence is inserted in both lines. Neither of these clones contained an *Xho* I site.

No λ clone was successfully purified from line NC20. One of the clones from line NC22, λ3NC22, was used to probe a *Bgl* II filter as with the clones above. The pattern of homologous fragments (figure 4.21) is that expected to be produced by a member of a mobile, dispersed, repetitive sequence family. The similarity of the genomic fragments suggests that line NC20 may contain a similar insertion (I in figure 4.26).

Line NC22 was also one of the six lines with an insertion at about position -15 (figure 4.24). Unfortunately the phage clone purified from this region (λ4NC22) did not extend far distally enough to contain the insertion. Two clones containing DNA from the insertion revealed by pASC94R1 and pASC94R4 in line FV2 were isolated successfully. One of these, λ1FV2 was

homologous to five genomic *Bam* HI fragments. The largest of these, which did not correspond to any known fragment from the AS-C region, appeared to be five to ten times the intensity of the smaller, presumably single copy fragments. Such a pattern might be produced if the DNA inserted into line FV2 contained *Bam* HI sites towards either end or represented an element such as the G elements which are usually found as tandem repeats (See Chapter 6). The size of this insertion (20 kb) is unusually large. The reason for this could be that this represents an unusually large single element or that it represents more than one insertion, possibly one into another. Digestion of a repeated sequence such as this would give proportionally more copies of a given sized fragment. Neither of the insertion-deletion events associated with lines FV13 or 14 were cloned and isolated successfully.

4.2.2.6. Summary

Figure 4.24 shows a restriction map of the AS-C region surveyed. The variation found in this region may be divided into two types. One class of variant could be ascribed to a change in a single nucleotide which resulted in the creation or destruction of a site at which a particular restriction enzyme cut the DNA. The second class could not be explained by this type of event and requires the insertion or deletion of DNA from the region as an explanation. The insertion-deletion variants revealed are shown in figures 4.25 and 4.26.

The *Bgl* II restriction site at position -19.1, given by Campuzano *et al.* (1985), was not present in ten of the twenty lines surveyed. In twelve lines a *Bgl* II site at about position 47.8 was present, and absent from 36 lines. In a single line, NC20, a new *Bgl* II site appeared at position 51.7. Again a single line, this time NC3, appeared to possess a new *Bam* HI site at position 54.5. In eleven lines an *Xba* I site had been created at either position 11.2 or 21.6. This site has been shown in figure 4.24 at position 11.2. The *Xba* I site at position 62.4, given by Campuzano *et al.* (1985) appeared to have been lost in four lines.

Fifteen differences of the fragments from the AS-C region, from the common pattern, required an insertion or a deletion of more than 0.5 kb of DNA, compared to the most common type as an explanation.

The two lines NC20 and NC22 contained an insertion event in the most proximal *Bgl* II fragment surveyed (I and II in figure 4.26). Even though both lines possessed smaller *Bam* HI and *Bgl* II fragments in this region, and so could be due to deletion events, the pattern of genomic *Bgl* II fragments homologous to λ3NC22 suggests that a transposable element was probably inserted which contains both *Bgl* II and *Bam* HI sites. Although the event in line NC20 could not be characterized as fully as this, its similarity of fragment pattern with line NC22 suggests that it too may contain a similar insertion.

Lines NC10 and 18 each contained an insertion of about 10 kb, estimated from the difference between the sum of the two *Xba* I fragments disclosed with λsc53 and the fragment from which they were presumed to be derived (IIa and IIb in figure 4.26). These insertions could not be distinguished on the basis of the genomic fragments observed or the fragments homologous to the cloned sequences. They both appear to occur within the most proximal *Bam* HI and next most proximal *Bgl* II fragments surveyed and must therefore lie between the *Bam* HI site at position −27.1 and the *Bgl* II site at position −28.5. It would seem that these insertions arose in a chromosome containing an *Xho* I site at about position −32 which was not seen in any of the other chromosomes surveyed. These insertions are labelled IIa and IIb in figure 4.25.

The insertions labelled IVa − f in figure 4.26 were found in lines NC4, 6, 8, 15, 19 and 22. These insertions appeared to be identical at the restriction fragment level. The size was estimated to be about 4 kb from the difference in length of the *Bam* HI fragments observed. From the fragments which were affected by these insertions they may be localized between the *Xba* I site at position −12.2 and the *Bgl* II site at position −16.7.

Five insertion events were found just distal to transcript 4 (figure 4.26). In line NC12 the *Bam* HI fragments appeared to represent DNA which was 0.9 kb longer than the most common fragments. This insertion (V in figure 4.26) was presumed to lie between the *Bam* HI sites at positions 33.8 and 39.1. In line NC25 an insertion of at least 2.3 kb of DNA had occurred between the *Bgl* II sites at position 33.1 and 42.2 (VI in figure 4.26). In line FV2 there was an insertion of DNA some 20 kb long, judged from the difference in the sum of the *Xba* I fragments in this line from the most common size. The λ clone, presumed to contain at least part of this insertion, hybridized to genomic

Bam HI fragments which appeared to be repeated elsewhere in the genome. This insertion (VII in figure 4.26) may then represent the insertion of a repetitive element between the *Xba* I site at position 38.5 and the *Bgl* II site at position 42.2. Insertions of this size have not been observed before (Langley *et al.* 1982; Leigh Brown 1983; Aquadro *et al.* 1986). The large size might be due to more than one insertion event.

In both lines FV13 and 14 there appeared to be a change in the quantity of DNA present between the *Xba* I sites at position 34.0 and 38.5. In line FV13 the difference was either due to a deletion of some 2 kb of DNA, or the insertion of DNA containing restriction sites for both *Bam* HI and *Bgl* II. Since such a large deletion event has not been reported previously in this type of survey, and therefore appears to be relatively rare, this has been shown as an insertion (VIII in figure 4.26). The insertion IX occurred in line FV14 appeared to be at least 6 kb long from the difference in size of the *Xho* I fragment in this line from the most common pattern.

The cause of two further events could not be established precisely. In two FV lines there appeared to be a small deletion, of 40 bp, at about position 66 (A of figure 4.24) and eight further lines from the two population samples appeared to contain a deletion of 30 bp or so at about position 61 (B of figure 4.24). Deletions of this size have been observed before at the *Adh* locus (Kreitman 1983). Either of these events could also have been due to the presence of a new restriction site 40 bp or 30 bp from the end of each fragment. More detailed analysis would be needed to resolve this.

For events which were found to occur more than twice in the AS-C locus survey table 4.7 presents the various haplotypes and which fly lines these were present in. The variable restriction sites with frequencies of more than twice in the sample were the *Bgl* II sites at positions -19.1 (G -19.1) and 47.8 (G 47.8) and the *Xba* I sites at positions 11.2 (X 11.2) and 62.4 (X 62.4). Insertion event IV and the putative insertion-deletion event B were the only two variants at any appreciable frequency apparently not due to a single nucleotide substitution.

Opposite page 80

Table 4.7

Table of haplotypes determined for the chromosomes surveyed at the AS-C locus region. The restriction sites were the *Bgl* II sites at positions -19.1 (G -19.1) and 47.8 (G 47.8) and the *Xba* I sites at positions 11.2 (X 11.2) and 63.4 (X 63.4). IV represents the six insertion events IVa to IVf (see figure 4.25) and B the putative insertion-deletion event at about position 61 (see figure 4.24). A + denotes the presence of a site or insertion and a - the reverse. A ? indicates that the polymorphic event was not scored for that haplotype.

Line Number	G -19.1	IV	X 11.2	G 47.8	A 61	X 62.4
NC16, FV24	–	–	–	+	–	–
NC27	?	–	–	+	–	–
NC26	?	–	?	+	–	–
NC12	?	–	–	+	–	?
NC20	–	–	–	+	+	+
NC15, 19	–	+	–	+	+	+
NC6, 8	–	+	–	+	+	?
NC4	–	+	–	+	–	?
NC22	?	+	–	–	–	+
FV1	+	–	–	+	+	?
NC1	?	–	+	–	–	?
NC2	–	–	+	–	–	?
NC7, 14	+	–	+	–	–	?
NC10, 11	?	–	+	–	–	?
NC18, 23, FV4, 7, 20	?	–	+	–	–	+
FV2, 13 to 18	+	–	–	–	–	+
NC17, 21, FV3, 5, 6, 8 to 12	?	–	–	–	–	+
NC3, 5	?	–	–	–	–	?
NC9	–	–	–	–	–	?
NC25	?	–	–	–	–	–
NC13	?	–	?	–	–	?
NC24	?	–	–	?	?	+

CHAPTER 5

ANALYSIS OF RESULTS

5.1. Introduction

The variation observed at the DNA level between chromosomes may be classified into two different types. The first involves the substitution of a single nucleotide of a sequence for another. This variation appears to be responsible for the structural differences in proteins between different species and has been studied extensively. The second class has only been studied more recently. The differences between individuals here are due to the gain or loss of DNA rather than substitution. Comparisons between different species indicate that such variation may play little part in the long term evolution of specific genes. To understand the importance of the survey results presented here it is necessary to interpret them in the light of population genetics theory and compare them with other experimental results obtained using other techniques.

The theory behind the population genetics of sequence variation is extensive. The generation of such variation is understood at the molecular level and its distribution in populations has been the subject of much theoretical work. The method used to study nucleotide substitutions of the DNA level relies upon the fact that restriction enzymes which cleave the DNA recognize a specific nucleotide sequence. The alteration of a single base in this sequence will prevent recognition, and hence also cleavage, by the enzyme. Sites may also be created by a change producing the recognition sequence from a closely related one. If the enzymes are chosen in such a way that they form a representative sample of the DNA then inferences can be made about the entire DNA sequence.

The study of insertion-deletion variation has only expanded in recent years, following the discovery of mobile sequences within prokaryotic and eukaryotic genomes and speculation about the potentially parasitic nature of repetitive DNA (Doolittle and Sapienza 1980; Orgel and Crick 1980). The theoretical aspect of variation due to the insertion of DNA sequences has grown quickly, based mainly on mechanisms of the generation of insertions and observations of differences between laboratory strains of *Drosophila*. Analysis of natural populations for such variation has lagged behind the

theoretical work but more data are being rapidly accumulated.

Models proposed for the population dynamics of mobile elements have for the most part been based on the assumption that populations have reached equilibrium with respect of the numbers and the distribution of elements. Two opposing mechanisms by which such an equilibrium may be attained have been suggested. Transposition of elements is generally assumed to be duplicative, giving rise to an increased number of elements. This increase is opposed by a decrease due to the effect of finite population size where elements are lost by chance from the population, the spontaneous deletion of elements at a constant rate from the genome and the action of selection to systematically remove elements.

The model proposed by Langley *et al.* (1983) and Charlesworth and Charlesworth (1983) supposes that the rate of duplicative transposition decreases with copy number such that at equilibrium the number of new copies generated is balanced by the number lost due to random drift and spontaneous deletion. The presence of transposable elements in the genome is assumed to have no effect on the fitness of the host. The stable copy number of elements is thus reached by a decreasing rate at which new copies are generated. A second model has also been put forward by Charlesworth and Charlesworth (1983) in which the generation of new elements by duplicative transposition is balanced at equilibrium by a decrease in host fitness, with a resulting increase in the probability of selective removal of individuals with increasing numbers of elements. The rate at which elements are lost from the population will ultimately increase until an equilibrium is reached.

For both types of variation, insertion-deletion and nucleotide substitutions, there are two reasons that one might expect a reduction in variability at X-linked loci. Firstly, alleles not completely dominant respond more rapidly to directional selection on the X chromosome than autosomes (Morton 1971). Secondly the effective population size of X chromosomes is usually smaller than that of the autosomes. The ratio of the effective population size for X linked loci to that of autosomal loci is given by

$$9(F + M) / 8(F + 2M)$$

where F and M are the numbers of homo- and heterogametic types (Crozier 1976). The effects of random drift are greater for the X chromosomes if the ratio of the number of homo- to heterogametic types is less then 7.

5.2. Restriction Site Variation

5.2.1. The Level of Nucleotide Variability

The way in which the enzymes were chosen for the 87A7 heat shock survey almost certainly introduced some bias into the sample of restriction sites surveyed. Two enzymes were included purely because a previous survey had shown that one site for each enzyme was polymorphic. Because of this little can be said about the general variability of the DNA sequence at this locus

At the heat shock locus 21 out of 29 chromosomes possessed the *Pst* I site at position -0.8 (figure 4.23), a frequency of 0.72. In the previous survey of Leigh Brown (1983) this site was present in 19 out of 29 chromosomes, or a frequency of 0.66. These numbers are not significantly different at the 95% probability level. The chi-square value was 0.10 with a critical value of 3.84. In the previous survey two chromosomes out of 29 surveyed lacked the *Xho* I site at position 18.8 (figure 4.23). Although no chromosomes lacked this site in the 27 chromosomes examined here, the percentage frequencies of this site in the two samples are not significantly different with a 95% probability. The two populations sampled at the 87A7 heat shock locus are not significantly different from each other for the allele frequencies at the two polymorphic restriction sites known to exist.

The AS-C locus survey is more informative about the general sequence variability since the enzymes used were chosen solely with a view to facilitating the search for insertion-deletion variation, rather than nucleotide polymorphism. Because the sample of restriction sites surveyed was expected to be a random one with respect to their potential polymorphism, the nucleotides surveyed with these enzymes should be a representative sample of the DNA sequence. Inferences can be made about variation at the nucleotide level with this sample which would not have been valid with the sample taken from the 87A7 heat shock locus.

The positions where the four chosen enzymes (*Bgl* II, *Bam* HI, *Xba* I and *Xho* I) restricted the DNA of Canton-S and Oregon-R strains of *Drosophila melanogaster* had been mapped previously (Campuzano *et al.* 1985). Some degree of bias against polymorphic sites may have been introduced because all sites for these enzymes were fixed in the two strains used to generate the restriction map, although this absence was not expected to be a significant bias. A single *Bgl* II site was absent from one strain mapped at the molecular level which was mutant at the AS-C locus. It was not clear if this was due to the fixation of a previously existing polymorphic variant from a natural population or a novel mutation in that strain (Campuzano *et al.* 1985). This *Bgl* II site was found to be polymorphic in the two samples of natural populations here.

In all, the enzymes used for the AS-C locus survey restricted the DNA at 69 sites in the North Carolina sample and 67 sites in the Spanish one. The difference was due to two sites which were only cleaved once in the NC sample. The proportion of polymorphic restriction sites in the NC sample was 6 out of 69 or 0.087. In the FV sample this proportion was 4 out of 67 or 0.060. Each of the enzymes used recognized a sequence of six nucleotides. Assuming that the frequency of nucleotide substitutions is low enough so that each change in a restriction site was due to a single nucleotide substitution the following calculation is possible. Let k be the number of restriction sites segregating for a variant and m be the number of restriction sites surveyed. The proportion of polymorphic nucleotides in the DNA, assuming that the restriction sites were a representative sample of the sequence, is given by

$$p = k/(2mj - k) \qquad (5.1)$$

(Ewens *et al.* 1981) where j is the number of nucleotides in the enzyme recognition sequence. For the NC sample the proportion of polymorphic nucleotides was $7.41 \times 10^{-3} \pm 3.02 \times 10^{-3}$ with $k = 6$, $m = 69$ and $j = 6$. This value was $5.08 \times 10^{-3} \pm 2.54 \times 10^{-3}$ for the FV sample with $k = 4$, $m = 67$ and $j = 6$. The standard errors were calculated from the approximate sampling variance expression for this quantity given by Hudson (1982).

With the same information, but the further assumption that the nucleotide substitutions are selectively neutral, an estimate of the

heterozygosity per nucleotide may be calculated as follows. The probability that a zygote formed from two randomly chosen chromosomes would be heterozygous for any given nucleotide is approximately equal to 4Nu (= θ) where N is the effective population size and u is the rate at which novel bases are generated at a given site. Also

$$\theta = p/\ln(n)$$

where p is given in equation (5.1) and n is the number of chromosomes surveyed (Ewens *et al.* 1981). The variance of this estimate for no linkage (complete linkage) may again be calculated from Hudson (1982). In the case of complete linkage the standard error of the estimate is larger because the variance of p contributes substantially to the variance of θ. The values of θ for the NC and FV populations were 2.39×10^{-3} $\pm 9.75 \times 10^{-3}$ ($\pm 1.99 \times 10^{-2}$) and 1.75×10^{-3} $\pm 8.75 \times 10^{-4}$ ($\pm 2.09 \times 10^{-2}$). Neither the estimates of the proportion of polymorphic nucleotides nor the heterozygosity per nucleotide for the two populations are significantly different, although the standard errors are of the same order of magnitude as the mean estimates. The two populations do not appear different for the variability of nucleotides.

5.2.2. The X – Autosome Comparison

Comparisons between the third and X chromosome variability, with respect to nucleotide substitutions, are not possible with the results of this study only. The reason for this is that restriction site variation was not studied in sufficient detail at the 87A7 heat shock locus. Comparisons between this and other studies are informative however. As stated above, the American and Spanish populations sampled are very similar in their levels of nucleotide polymorphism at the AS-C locus. Indeed all four restriction site variants which were not unique were represented in both populations. Combining results from other surveys to give estimates of the variability at the nucleotide level for different regions of the genome would seem reasonable.

In *Drosophila melanogaster* four studies of this type of variation have been carried out, three at the *Adh* locus (Langley *et al.* 1982; Kreitman 1983; Aquadro *et al.* 1986) and one at the 87A7 heat shock locus (Leigh Brown 1983). Estimates of the proportion of polymorphic nucleotides were 1.35×10^{-2},

1.61×10^{-2} and 2.7×10^{-2} for the *Adh* locus surveys respectively and 7×10^{-3} for the 87A7 heat shock locus. The values of θ produced for these loci were 6×10^{-3}, 6×10^{-3}, 7×10^{-3} and 2.4×10^{-3} respectively. These estimates suggest that the X linked AS-C locus is just as variable at the nucleotide level as the autosomal loci surveyed since they do not differ significantly from the same values determined from the AS-C locus.

5.3. Insertion deletion variation

5.3.1. The X - Autosome Comparison

At the 87A7 heat shock locus, in 25 kb of DNA, 6 out of 32 chromosomes contained a large insertion of some kind, a frequency of 0.19. In 120 kb on the X chromosome at the AS-C locus 15 out of 49 chromosomes differed by more than 0.5 kb from the most common sequence arrangement, a frequency of 0.31. It appears quite common that chromosomes differ not only by nucleotide substitutions but by the total quantity of DNA present. The cause for these differences in all cases except one was shown to be an insertion of DNA indeed in two cases was determined to be a mobile repetitive element. In one case the inserted DNA hybridized to DNA fragments which appeared to be repeated elsewhere in the genome and two further insertions appeared to contain DNA not homologous to the AS-C locus probes used in the survey. It is tempting to ascribe each insertion to some sort of transposable element, even though only in three cases has evidence favouring this been found. In other surveys of this kind which have been carried out so far (Leigh Brown 1983; Aquadro *et al.* 1986) all insertions of more than 0.5 kb which have been characterized proved to be homologous to some type of transposable element.

As has been said above, intuitively one would expect that if insertions of DNA produce deleterious mutations which are not completely dominant then, as with the null alleles at enzyme loci, their numbers should be reduced on the X chromosome. Two similar approaches may be used to determine if the number of insertions on the X is lower than should be expected. Method 1: By calculating the average number of observed insertion events per nucleotide at each locus surveyed an expected value for the numbers of insertion in each sample may be generated. The deviations of each locus from this value can be

tested with a chi-square statistic for homogeneity between loci.

Secondly method 2: the expected value can be generated by an method independent of the data being tested. The proportion of the *Drosophila* genome which is represented in the form of transposable elements has been estimated (see Spradling and Rubin 1981). Also the average length of such elements is known (Manning *et al.* 1975). From both these estimates the number of expected insertions of all transposable elements per nucleotide can be calculated, assuming a uniform distribution of element insertions over the whole genome. Thus the deviations of the numbers of insertions at each locus from the number of transposable elements expected can be tested for significance.

Two previous surveys have produced observations of the number of large insertions in a sample of wild-derived chromosomes. At the 87A7 heat shock locus Leigh Brown (1983) found four insertions of DNA within 25 kb of DNA in a sample of 29 chromosomes from a single North Carolina population. Three of these insertions were found to be transposable elements, the fourth was not tested. Aquadro *et al.* (1986) surveyed 13 kb around the *Adh* locus in a sample of 48 chromosomes taken from four populations. Eleven large insertions were found in this span and each was shown to be a transposable element.

The number of insertions found in the survey presented here at the heat shock locus (six) is not greatly different from the number (four) observed by Leigh Brown (1983) with a very similar sampling regime. This would suggest that the combination of four small samples from different populations (Aquadro *et al.* 1986) is not a serious problem, although the difference in the number of insertions between the NC (twelve) and FV (three) populations at the AS-C locus was found to be relatively large. Samples between populations at the same loci have been combined.

Method 1

The numbers of insertions at each of the three regions surveyed were 10, 11 and 15 for the HS, *adh* and AS-C loci respectively. Dividing the number of insertions by the product of the number of chromosomes and the number

of nucleotides surveyed gives the observed insertion frequency per nucleotide. These values are 6.56×10^{-6}, 1.76×10^{-5} and 2.55×10^{-6} for the three regions respectively. The weighted mean insertion frequency, across all three loci was 4.48×10^{-6}, giving an expected number of insertions in the HS, *adh* and AS-C regions, respectively, of 6.84, 2.80 and 26.36. The chi-square value with two degrees of freedom is 30.37. The critical value, with a 95% confidence limit, is 5.99. Thus the number of insertions found at these three loci differ significantly from each other.

Method 2

Manning *et al.* (1975) have estimated that about 12% of the *Drosophila melanogaster* genome is present in families of homologous sequences reiterated about 70 times each. Approximately half this amount is dispersed throughout the haploid genome and appears to be mobile. The average length of individual members of these families is 5.6 kb (Manning *et al.* 1975). The total length of the *D. melanogaster* genome has been estimated by Rasch *et al.* (1971) to be 1.65×10^{8} bp. From these figures the number of mobile sequences in the genome may be estimated to be 1.77×10^{3}. If these sequences were distributed uniformly over the entire genome there should be on average 1.07×10^{-5} elements per nucleotide.

The expected number of insertions at each locus with this estimate of insertion frequency are 16.34, 6.69 and 63.00. The chi-square values with one degree of freedom are 2.46, 2.78 and 36.57 for the HS, *Adh* and AS-C loci respectively. The critical value with the same confidence limit used above is 3.84. This shows that the number of insertions found at the autosomal loci are not significantly different from that expected to be found due to a uniform distribution of transposable elements throughout the genome. There is however a significant paucity of insertions at the AS-C locus in this survey. Transposable elements are seemingly removed from the population by the higher degree of selection on the X chromosome.

5.4. Distribution of Insertions in Relation to the DNA Sequence

The significance of the difference above depends on a uniform distribution of insertions throughout the DNA. Examination of the HS and *Adh* regions were only over comparatively short regions surrounding known transcripts. The difference of the AS-C region would disappear if a distance of only about 20 kb had been examined surrounding, for example, T4 of figure 4.25. It is difficult to determine if the pattern of insertions found at the AS-C locus is significantly non-uniform since many more insertion events would need to be analysed. All the insertions found appear to be grouped into two clusters close to transcripts T4 and T1. If this non-uniformity were a general phenomenon then the apparent inhomogeneity between the three loci could simply reflect that the AS-C survey provides a more representative sample than either of the other loci of the majority of the genome. In this case the chi-square test would not be valid and the differences may appear significant only because of sampling effects. The apparent lack of elements on the X chromosome compared with the number expected from the second method of estimation may also be explained by a non-uniform distribution of elements. Many mobile elements have been found in clusters of elements, not interspersed with unique sequence DNA (Manning *et al.* 1975). The effect of this would be to reduce the expected values for the second chi-square test. The number of insertions found in the HS and *Adh* surveys would then be too high because samples biased towards regions of increased insertion frequency were used.

5.4.1. The Level of Insertional Variability

As with nucleotide polymorphisms the heterozygosity per nucleotide may be calculated for the insertion events larger than 0.5 kb. This is a useful way in which to compare the two classes of variation. The method used here calculates only observed heterozygosity rather than an estimate based on a specific population model. The heterozygosity is calculated as the proportion of the observed haplotypes for which pairs would differ by an insertion event, multiplied by the number of such events by which they differ. Haplotypes paired with themselves are not included.

This observed heterozygosity may be formulated as:

$$H = 2\sum n_i n_j v_{ij} / n(n-1)$$

where v_{ij} is the number of events by which haplotypes i and j differ, n_i, n_j ($i \neq j$) are the number of times each haplotype occurs and n is the total number of haplotypes. For the HS locus survey here there were six insertion events, two of which could not be distinguished on the basis of their restriction fragments. The combination between these two were presumed not to differ with respect to their insertion events. In total there were two chromosomes with insertion II (figure 4.23), four single chromosomes each with a single, different insertion and 26 chromosomes with the observed minimum of DNA at this locus.

The observed heterozygosity was 0.371. Correcting for the length of DNA examined (25 kb), the observed heterozygosity per nucleotide was 1.48×10^{-5}. For the AS-C locus survey in the NC population there were five chromosomes with insertion IV (figure 4.26), two chromosomes with insertion III, one chromosome with both insertions II and IV and three single chromosomes each with a unique insertion in a sample of 27 chromosomes. The observed heterozygosity per nucleotide was 6.89×10^{-6}.

Similarly for the FV population sample with three different insertions, each on a different chromosome from a sample of 22, the observed heterozygosity per nucleotide was 2.27×10^{-6}. In the other two surveys, one at the 87A7 HS locus (Leigh Brown 1983) and one at *Adh* (Aquadro *et al.* 1986), this value may be calculated as 1.10×10^{-5} and 3.39×10^{-5}. It is difficult to give a standard error for these estimates but the values for autosomal loci are very similar while those at the AS-C locus are slightly lower. It would seem that this difference is due mainly to the lower number of insertions found on the X chromosome, rather than similar numbers, but fewer distinguishable insertion events.

Heterozygosity estimates for insertion-deletion variation are approximately fifty fold lower than for nucleotide substitutions. In general one may infer from the proportion of polymorphic nucleotides that a given position will be segregating for a substitution once in 200 nucleotides. For insertional variation a large insertion is expected to be found, relative to the most common type, once in 200,000 nucleotides.

5.5. Gametic Disequilibrium at the AS-C Locus

Using computer simulation models of the behaviour of transposable elements in a population Charlesworth and Charlesworth (1983) reach the conclusion that the effect of linkage between elements is negligible. The value of the recombination fraction between two adjacent element sites was taken to be 0.003 at the most extreme. One chromosome from the AS-C locus survey contained two insertions within 30 kb of each other. An upper limit to the recombination fraction between *achaete* and *scute* has been estimated by Dubinin *et al.* (1937) to be about 6.6×10^{-5}, about 50 fold lower than the simulation value. When elements from different families are considered together, the effects of linkage may be more significant than supposed in the simulation model.

The effect of low levels of recombination is to allow selection to act on a series of different loci as a group, rather than independently. If a new mutation occurs only once in a population then, because it is unique, it must be associated with only one combination of any segregating alleles. Over a period of time, if the new mutation persists in the population, recombination can produce new combinations of polymorphic alleles. Until this happens the haplotype into which the new mutation is introduced acts as a single unit.

Unique mutants are immediately present in linkage disequilibrium, if other polymorphic loci exist. This is also the case for polymorphic variants introduced into a population by a single, or very few, migrants. Polymorphic alleles only found in the immigrants are initially only represented in the configuration in which they arrive in the population. It has also been shown that when the population size is finite, and alleles become either lost or fixed due to random drift, the stochastic nature of the process can produce gametic disequilibrium. The effects of selection on linked loci can produce a systematic change in the haplotypes present in the population. The net effect of mutation, migration and drift will not produce such a systematic change.

Linkage will only have an effect on a change in gene frequencies at different loci if different haplotypes have different selective values. Under these circumstances disequilibrium will produce apparently synergistic effects between alleles which would otherwise appear independent. Such a synergistic effect has been found to be necessary to maintain a realistic copy number of

transposable elements by selection alone (Brookfield 1982; Charlesworth and Charlesworth 1983).

To determine whether pairs of segregating loci are in significant disequilibrium the deviation from the expected frequencies of the four different types can be calculated as

$$D = f_1 f_4 - f_2 f_3$$

where f_1 and f_4 are the frequencies of the classes where both or neither of a given pair of alleles are present and f_2 and f_3 the frequencies of the classes where only one or the other is observed. D can range from +1 to −1. To test for the significance of any observed disequilibrium the goodness of fit statistic

$$Q = Nr^2$$

is used. r^2 is the estimate of the squared correlation coefficient and N is the sample size with

$$r^2 = D^2 / (p_1 p_2 q_1 q_2)$$

where p_1, p_2, q_1, q_2 are the frequencies of the four alleles in the sample (Hill 1974). Providing the sample size is reasonably large, of the order of 25 or so, Q is distributed as a chi-square statistic with 1 degree of freedom.

The haplotypes for polymorphic differences in the AS-C sample which occur more than twice are given in table 4.7. Four of these differences are restriction site polymorphisms (G −19.1, G 47.8, X 11.2 and X 62.4), one is a large insertion event (IV) and the last is probably an insertion-deletion event of about 30 bp (B). Two of the restriction sites (G −19.1 and X 62.4) could not be scored in many of the lines and so the sample size of pairs involving these were too small to produce meaningful results. Insertion IV was not present in the FV population sample and so could not be included in the analysis of that population.

In the NC population sample, combinations between insertion IV and X −11.2, G 47.8 and B had Q values of 5.17, 6.64 and 6.26 respectively with

sample sizes of 21, 26 and 26. Combinations of X 11.2 with G 47.8 and B had Q values of 7.20 and 0.60 with sample sizes of 24 each. Between G 47.8 and B the value of Q was 4.40 with a sample size of 26. Combinations of X 11.2 with G 47.8, B and X 62.4 produced Q estimates of 0.39, 0.19 and 0.17 with sample sizes of 20, 20 and 22. Finally for combinations between G 47.8 and B and X 62.4 Q was 8.47 and 19.0 with sample sizes of 18 and 19 respectively. Of these eleven combinations, seven were in significant disequilibrium ie had values larger than the critical value of 3.84 at the 95% probability level. Three of these in the NC population sample involved insertion IV and five out of six throughout both populations involved G 47.8. There is therefore significant disequilibrium present at the AS-C locus.

CHAPTER 6

DISCUSSION

6.1. Introduction

The purpose of the survey presented here was to examine the pattern of insertion-deletion variation in samples of chromosomes from natural populations at two regions of the genome. The first, the 87A7 heat shock locus on the third chromosome, was chosen to compare the observed pattern at this locus with that found previously in a sample from a different population (Leigh Brown 1983). Secondly, a much longer contiguous stretch of DNA at the AS-C locus at 1B1-2 to 1B4-5 on the X chromosome was examined (Campuzano *et al.* 1985). The reasons for this second choice were two fold.

Knowledge of the effects produced by many insertions of transposable elements suggests that they cause generally deleterious mutations which are to some degree recessive (Shapiro 1983; Mackay 1986a; Fitzpatrick and Sved 1986). Alternatively, the distribution of certain elements on the chromosomes of wild- caught *Drosophila melanogaster* is accounted for adequately by a model for the expected distribution at equilibrium which assumes that each transposable element insertion is effectively, selectively neutral (Montgomery and Langley 1983; Kaplan and Brookfield 1983; Leigh Brown and Moss 1987). To distinguish between these two apparently contradictory observations a survey of insertional variation on the X chromosome would indicate if selection was affecting the distribution of insertions in natural populations. For alleles which are at least partly recessive, as the alleles produced by the insertion of a transposable element appear to be (Mackay 1986a), the effects of directional selection are more rapid on the X chromosome than the autosomes (Morton 1971). If selection does influence the distribution of elements in populations then there should be a detectable difference between the distribution on the X compared to the autosomes.

The second reason for examining the AS-C region was that the clones covering this region extended five times as far as the largest region studied previously in this way (Leigh Brown 1983). The pattern of transcription in this region has been studied and the positions of many transcripts have been mapped (Campuzano *et al.* 1985). These transcripts have also been placed in the context of the genetic map of this region, and in some cases the

transcript responsible for a particular genetic function has been identified. Analysis of both nucleotide and insertion-deletion variation over a long contiguous, well characterized, region of the genome would be informative about the intragenomic distribution of variation as well as the intergenomic pattern.

To understand the implications of the results presented here something must be said of how we know transposable elements behave at the molecular level and of the two main models which try to account for the observed distribution of transposable elements in natural populations.

6.2. The Properties of Transposable Elements

Analysis of genome structure has shown that in higher eukaryotes less than 5% of genomic DNA is present as single copy genes (Davidson and Hough 1973). Also closely related organisms, which presumably require similar amounts of genetic information, can contain widely different quantities of DNA (eg Rothfels *et al.* 1966). The difference appears to be due to changes in the quantity of repeated sequence DNA at localized points in the genome rather than a general increase of all sequences throughout the genome (Keyl 1965). The dispersion pattern of moderately repeated DNA within the genomes of higher eukaryotes has suggested to some that they may be involved in the regulation of gene action (Britten and Davidson 1969). More recently the idea that these sequences are present simply because they may promote their own replication within genomes has been put forward (Doolittle and Sapienza 1980 and Orgel and Crick 1980). Such 'selfish' replication it was claimed would ultimately result in a large proportion of the genome consisting of repeated sequences of this type.

Predictions of the population dynamics of these sequences were rather vague until more recently the theory of how mobile dispersed repetitive elements may evolve has expanded greatly (eg Ohta 1982, 1983, 1986; Langley *et al.* 1983; Charlesworth and Charlesworth 1983; Brookfield 1982, 1986; Kaplan *et al.* 1985; Slatkin 1985). These models have been based primarily on knowledge of the molecular mechanisms of transposition, determined in bacteria and yeast (eg. Shapiro 1979, 1983; Kleckner 1981; Grindley and Reed 1985; Craigie and Mizuuchi 1985, 1986; Bender and Kleckner 1986; Boeke *et al.* 1985 and Roeder and Fink 1982) and the observed cytological distribution of

certain transposable elements in the chromosomes of wild-caught *Drosophila melanogaster* (Montgomery and Langley 1983; Leigh Brown and Moss 1987).

6.2.1. The Sequence Organization of Mobile Elements in *Drosophila*

In *Drosophila* there are 10's to 100's of copies per genome of many families of sequences about 5 kb long, separated by over 13 kb of unique sequence DNA (Manning *et al.* 1975). Many such elements of this class have been identified in *Drosophila* which appear to be quite mobile in the genome (Wensink *et al.* 1974; Potter *et al.* 1979, 1980; Strobel *et al.* 1979; Finnegan *et al.* 1978; Ilyin *et al.* 1978; Young 1979). These sequences may be divided into several structurally distinct groups.

The *copia* class of transposable elements make up about 5% of the genome of *Drosophila melanogaster* (Rubin *et al.* 1980; Spradling and Rubin 1981). These elements are about 5 kb to 13 kb long bounded by direct repeats, or LTR's, of between 250 and 550 nucleotides. The sequence arrangement of the central region of *copia* appears to be very similar to the TY1 element of yeast (Cameron *et al.* 1979). These elements contained regions of homology at the DNA and protein level to retroviruses, although with a slightly different organization (Will *et al.* 1981; Mount and Rubin 1985). These elements appear to transpose by reverse transcription (Flavell and Ish-Horowicz 1983; Flavell 1984). The RNA intermediate may be encapsulated into a virus-like particle as part of the transposition process, although it has not been demonstrated that such particles can leave one host and infect another as the retroviruses do (Shiba and Saigo 1983).

The foldback (FB) elements consist of long inverted repeats of a simple tandemly repeated sequence up to several kb long (Potter *et al.* 1980). Elements of this class have been demonstrated to cause certain mutations, such as w^c, which mutate at rates of up to 10^{-3} per generation to other genetically distinct alleles (Green 1967; Collins and Rubin 1982, 1983). It is not clear how these elements move but they have been demonstrated to transpose large sections of the genome to other locations (Levis *et al.* 1982; Goldberg *et al.* 1982) suggesting that an RNA intermediate is not involved.

The F and G element families are not well characterized as yet but consist of sequences not internally repetitious (Dawid *et al.* 1981; Di Nocera

and Dawid 1983). These elements are almost certainly transposable although they tend to occur as tandem duplications, unlike other mobile elements, and their sites of integration are probably more stable (Pierce and Lucchesi 1981; Di Nocera *et al* 1986). One end of such elements contains a 30 bp to 40 bp poly-A sequence characteristic of processed pseudogenes. The similarity infers that, as with pseudogenes, new copies of the elements may be generated by reverse transcription of processed mRNA molecules (Sharp 1983).

Two different elements in *Drosophila melanogaster* can transpose at very high rates under certain conditions, the P element and the I factor (Kidwell *et al.* 1977; Bregliano *et al.* 1980; Kidwell 1982). The P element sequences are bounded by 31 bp inverted repeats and the complete P factor contains four open reading frames (O'Hare and Rubin 1983). Maturation of RNA transcribed from intact P factors initially involves the removal of two introns, the third intron is only spliced from the mRNA molecule in the germ cells. The mature mRNA is then translated to give a functional transposase (Laski *et al.* 1986; Rio *et al.* 1986). Transposition is therefore very much predominantly a germline, rather than a somatic event.

The mechanism of transposition of the P element probably does not proceed via an RNA intermediate and may involve the excision of one copy which then inserts elsewhere in the genome (Engels 1983). Transposition seems to generate many elements which, although internally deleted and can not therefore encode a functional transposase, remain mobile (Rubin and Spradling 1982; Engels 1984). Sequences homologous to the *Drosophila melanogaster* P factor have not been found in sibling species of *melanogaster* (Brookfield *et al.* 1984) but are present in more distantly related species (Lansman *et al* 1985). One theory to explain this pattern involves the horizontal transmission of P factor sequences, by infection, into *Drosophila melanogaster* from the more distantly related species.

The 5.4 kb I factor has a sequence organization somewhat similar to the F elements, having no internal repeats and a poly-TAA tract at one end (Fawcett *et al.* 1986). As with the P element, the transposition rate is greatly enhanced in crosses where sperm carrying complete elements fertilize eggs of females lacking them (Bregliano *et al.* 1980; Pelisson 1981; Bucheton *et al.* 1984). The transposition of the I element probably occurs via an RNA

intermediate, deduced from the observation that one of the two open reading frames may encode a protein which shows homology, at the protein level, to the reverse transcriptase genes of retroviruses and other mobile elements (Fawcett *et al.* 1986). I factor sequences have been found in other sibling species in the *Drosophila* subgroup (Bucheton *et al.* 1986; Moss and Leigh Brown personal communication). The distribution of these sequences is consistent with vertical transmission of the element sequences in *Drosophila* species. In all species where I factor sequences have been found there exist copies, usually in the centric heterochromatin, which carry 5' terminal deletions and are presumably stably integrated into the genome (Bucheton *et al.* 1984; Fawcett *et al.* 1986).

6.2.2. The Genetic Effects of Transposable Element Insertion

Elements from each of these families have been found inserted close, or in to, known transcription regions. These sequences are numerous and apparently quite mobile in the genome. Insertions of many of these elements have been shown to cause mutations. In fact most spontaneous mutations in *Drosophila* which have been analysed at the molecular level appear to be due to transposable element sequences. Five mutations at the *bithorax* locus, three *hairy-wing* mutants, eight mutants at the AS-C locus and three *white* alleles have all been shown to be caused by insertions of the elements 412, gypsy, *copia* or another uncharacterized mobile repetitive element (Bender *et al.* 1983; Levis *et al.* 1984; Pirrotta and Brockl 1984; Campuzano *et al.* 1985, 1986). Only one spontaneous mutant, y^1, was not associated with a change detectable by Southern analysis.

The mechanism by which insertions result in a mutant phenotype have been analysed and five classes have been found so far. These may be typified by $w^{hd80k17}$, w^a, y^2, Hw^{Ua} and w^{DZL}. $w^{hd80k17}$ results in a bleached white eye phenotype due to the complete absence of the wild type mRNA. The mRNA produced initiates normally but terminates within a P element inserted into the coding sequence of the *white* locus (Levis *et al.* 1984). The mutation is thus caused by an interruption of the normal protein coding sequence. The w^a allele contains an insertion of *copia* into an intron (Levis *et al.* 1984). Although many of the *white* transcripts terminate within the element a few continue right through the *white* gene. These complete transcripts may then be

processed and result in the removal, by RNA splicing, of the *copia* copy and the intron into which it is inserted, from the molecule giving a mature wild type transcript. The mutant phenotype is produced because of the much reduced levels of the wild type transcript (Levis *et al.* 1984).

The y^2 allele causes a yellow bodied phenotype by a reduction in the level of transcription from the *yellow* locus. The insertion of a complete *gypsy* element, about 500 bp 5′ to the *yellow* locus, causes this phenotype since y^2 revertants have lost the element (Parkhurst and Corces 1986). *Gypsy* is transcribed at the same time that the *yellow* product is expected to be required for normal pigmentation and appears to interfere with the transcription of *yellow* (Modellel *et al.* 1983; Parkhurst and Corces 1986). The fourth type of effect is observed with the Hw^{Ua} allele. Here a *copia* element is inserted into the transcript termed T4 at the AS-C locus (Campuzano *et al.* 1986, see figure 4.24). This gene does not appear to contain introns and its transcript in the mutant, which is terminated in the *copia* element, is produced at elevated levels resulting in a gain of function phenotype. The dominant mutant phenotype seems to be due to an over expression, relative to wild type, of the 5′ segment of the gene.

Each of these types of effect occur with the insertion of an element close to, or within the transcription region. In the case of the w^{DZL} allele the insertion responsible for the mutation is located 5 kb 5′ to the *white* gene (Levis *et al.* 1984). It is not clear how this insertion may cause a mutant phenotype although the presence of a segment of chromosome II, mobilized by two different FB elements, seems to be involved (Levis *et al.* 1982).

6.3. Distribution of Mobile Elements in the Chromosomes of *Drosophila*

Once a clone has been obtained for a given transposable element it becomes possible to localize homologous sequences in the chromosome arms of *Drosophila melanogaster* salivary gland polytene chromosomes by *in situ* hybridization. This involves labelling the cloned DNA in some way and hybridizing this to the intact chromosomes. The positions where homologous sequences are found in the chromosomes can be placed on the known morphological map (Bridges 1935).

The distribution in *Drosophila melanogaster* populations of *copia*, 412,

297, I element, Mdg1 and the P element have all been examined in this way (Montgomery and Langley 1983; Leigh Brown and Moss 1987; Biemont 1986; Ronsseray and Anxolabehere 1986; Ajioka and Eanes 1987). In each case the sites at which elements occur have been distributed throughout the chromosome arms in an approximately uniform manner. Any site occupied in any one chromosome is usually occupied in only a very few chromosomes and usually only one, ie is unique in the sample. This implies that sites of insertion which appear indistinguishable at the cytological level are not related by decent but rather appear identical because of the lack of resolution of the *in situ* technique.

The patterns of hybridization observed, and the very low frequency of any one site being occupied, are adequately explained by relatively rapid excision and insertion of these elements (Montgomery and Langley 1983). It is difficult to discount the effects of selection because the observed distribution of elements may also be explained by a model in which selective constraint is incorporated.

6.3.1. Equilibrium Models for the Distribution of Mobile Elements

The numbers of the elements 412, *copia* and 297 in the genomes of laboratory strains and wild-caught *Drosophila melanogaster* are very similar (Strobel *et al.* 1979; Montgomery and Langley 1983). In tissue culture cells, however, the number of copies of all three of these elements increase dramatically, although apparently not indefinitely (Potter *et al.* 1979). Some process must occur which limits the number of copies of a transposable element in the genome. The balance between gain and loss by which a stable number of elements per genome is achieved appears to be disturbed in tissue culture cells. Two reasons have been proposed for this. Either the mechanism by which the rate of duplicative transposition is regulated is altered or selection against high numbers of copies becomes relaxed in tissue culture. These ideas, together with the observations of transposable element behaviour above have been combined into two models which try to explain the observed distribution of transposable element sites in natural populations on the basis of an equilibrium between the generation and loss of elements.

Charlesworth and Charlesworth (1983) have proposed a model where the generation of new element sites by transposition and their loss by

spontaneous deletion occur at a constant rate. An equilibrium copy number is ultimately attained because it is assumed that each insertion of an element produces a deleterious effect on the host. The mean fitness of individuals decreases with increasing copy number to the point where the number of new copies entering the population by transposition is balanced by those lost by deletion and selective pressure. When such an equilibrium has been reached the rate of new element site production is balanced by a decrease in fitness of the population. It has been predicted that the mean fitness of an individual with the equilibrium number of copies of elements would be about 95.5% of the fitness of an individual containing no elements. This compares with a fitness of 91.3% for individuals with a number of copies 2 standard deviations above the mean (Charlesworth and Charlesworth 1983).

The form of the distribution of occupied sites predicted by such a model, where the number of sites is large and an equilibrium is reached without the saturation of these sites, is approximated by a Poisson distribution. By computer simulation of the population dynamics of elements expected from the assumptions made for the theoretical equations it was found that the analytical predictions were close to the results of the simulations. Also by using different degrees of recombination between the sites of insertion it was shown that such linkage did not affect the simulated distribution of elements significantly (Charlesworth and Charlesworth 1983).

Langley *et al.* (1983) and Charlesworth and Charlesworth (1983) have suggested that the number of copies of a given element might be limited by some form of negative feedback reducing the rate of duplicative transposition with increasing copy number, rather than a reduction in fitness. There is a wide range of evidence to support this sort of self regulation of transposition in bacteria (eg. Robinson *et al.* 1977; Foster *et al.* 1981; Kleckner 1981). In *Drosophila* the P and I elements also appear to be able to regulate their own copy number (Engels 1983; Pelison and Bregliano 1987) Under these models the insertion of each element is assumed to have a negligible effect on the survival of the host organism. Again a stable equilibrium would be reached, this time by reducing the number of new elements produced each generation rather than increasing the number lost.

The statistical properties of such an equilibrium have been examined

(Kaplan and Brookfield 1983). In the model of Langley *et al.* (1983), with an infinite number of potential sites of integration for the mobile elements, the frequency spectrum of the element distribution is determined by two quantities, Λ : the expected number of elements per haploid genome and $\theta = 4Nu$: N is the effective population size and u is the spontaneous deletion rate of an integrated element. For the neutral model of Charlesworth and Charlesworth (1983) three parameters determine the distribution of occupied sites in the population. These are $\alpha = 4N\mu$: N is defined above and μ is the rate of insertion of an element into a given site, $\beta = 4Nv$: v is the rate of spontaneous deletion and n : the expected number of copies of a given element per genome at the equilibrium point. The two models are equivalent if there are an infinite number of potential sites of integration so that the chance of insertion into any one of these tends to zero, ie $\alpha \rightarrow 0$.

If $\theta > 1$ (Langley *et al.* 1983) then the number of deletion events per site occurring each generation will be more than one and the majority of insertions will have a very low frequency. If θ is low, however, ($\theta < 1$) then some sites of insertion are expected to drift upwards in frequency and eventually become fixed because the effects of spontaneous deletion are negligible. For low frequencies of site occupation the distribution is expected to approximate a Poisson distribution. This is also true for the model incorporating selective constraint.

6.3.2. Observations of Transposable Elements at the Cytological level

Analysing the data of Montgomery and Langley (1983), Kaplan and Brookfield (1983) estimate θ to be 16.72, 34.97 and 48.26 for 297, 412 and *copia* respectively in a North Carolina population. These values are all much greater than 1 which means that the effects of insertion and deletion predominate rather than random drift due to finite population size. The implication of this is that where sites appear to be occupied more than once in the sample they are not related to each other by descent from a common ancestor but rather represent cases where an element has inserted into a particular site more than once. Leigh Brown and Moss (1987) have estimated θ to be 21.5 and 16.9 for the I factor and *copia* respectively in a Spanish population. Again these values are much greater than unity. The difference of three fold in the values for *copia* in the two populations is probably mainly

due to a difference in copy number, rather than a difference in the distribution of elements (Leigh Brown and Moss 1987). θ values are all similar between different elements, considering that the errors associated with high θ values are large (Kaplan and Brookfield 1983). Either the properties peculiar to the host are more important than those of the mobile elements in determining the distribution of insertion sites, or elements of very different structure possess very similar properties (Leigh Brown and Moss 1987).

Analysing the data of Biemont (1986), Leigh Brown and Moss (1987) found that in a wild-derived culture of *Drosophila melanogaster*, mass mated in the laboratory for 18 months, the increased drift component due to limited population size reduced the observed θ value for the I factor to 4.6. Although this is about one fifth of the value found in a natural population sample, and so would seem much lower, this is almost certainly too high to be the value of θ at equilibrium in the mass mated culture. The transposition rate would need to be of the order of 10^{-2} excisions per site per generation, assuming a value for N of 100 (the culture was initiated with 50 inseminated females). During I-R hybrid dysgenesis the transposition of the I factor is increased markedly (Picard 1976). Even so the reversion frequency of the I factor induced *white* (w^{IR1}) mutation was less than 3×10^{-5}. This I factor is known to be genetically active (Pelison unpublished results, cited in Buchton *et al.* 1984). Although the restoration of wild-type eye colour may depend on precise excision of the I factor. There remains the possibility that the I homologous sequences can excise frequently so that they could not be detectable by the *in situ* technique while retaining the mutant phenotype. It does not appear that any of the alterations derived from the w^{IR1} mutant were due to I factor excision (Sang and Sved, cited in Bucheton *et al.* 1984). From this it must be concluded that the spontaneous deletion rate of the I factor is much lower than 10^{-2}.

Care may be needed in the use of equilibrium models for the interpretation of transposable element distributions in the genome. Models for populations which are not at equilibrium (eg Ohta 1986) may be needed.

6.4. Molecular Analysis of Wild-Derived Chromosome Segments

6.4.1. Comparisons Between Loci and Different Populations

The technique of *in situ* hybridization of mobile element clones to the salivary gland chromosomes of *Drosophila* provides a method of mapping finely the sites of insertion of an element to the nearest cytological band of the chromosome. Each of these bands may contain several tens of kb of DNA into which any element could integrate (eg Bossy *et al.* 1984). With the *in situ* technique it is not possible to resolve insertion events which may be relatively distant on a molecular map. Also only a limited set of elements may be examined in this way since the element must have been isolated previously. By using cloned fragments of single copy DNA, however, it is possible to examine insertions of any kind, in fragments of DNA only hundreds of nucleotides long. In this way the distribution of mobile elements may be examined with much greater resolution.

Two detailed examinations of population samples at the restriction fragment level have been carried out so far. In the first of these 25 kb surrounding the 87A7 heat shock genes were examined in 29 third chromosomes (Leigh Brown 1983). Four chromosomes were determined to contain a single insertion, each of which was larger than 0.5 kb. These different insertions were all into the 1 kb spacer region between the two divergent heat shock transcripts (see figure 4.23). Upon further analysis three of these insertions proved to be homologous to members of transposable element families while the fourth was not tested. In the second, more extensive survey, 13 kb of DNA surrounding the *Adh* locus on the second chromosome were analysed in 48 chromosomes taken from four separate wild populations (Aquadro *et al.* 1986). 11 insertions of more than 0.5 kb were found surrounding the single *Adh* transcript and each was shown to represent a transposable element family. It became clear from these results that variation between chromosomes due to transposable element insertion is a widespread phenomenon.

Initially it was surprising to find these insertions so close to known transcription regions, especially in the case of the heat shock locus since all the insertions were found within only 1 kb of the 5′ ends of both transcripts. A second point was that, although the second survey combined several

populations, the number of insertions into these two autosomal loci were similar, suggesting that elements would be found with similar frequencies over different regions of the autosomes.

For the first part of the survey presented here the 87A7 heat shock locus was examined in a way similar to that used by Leigh Brown (1983). Again 25 kb were surveyed in 32 chromosomes from a French population, whereas the previous work had been conducted with a sample taken from a population in North Carolina. Six insertions were found, one of which has been shown to be a mobile element, the others were not tested (see figure 4.23). This number of insertions is similar to that found previously. Variability due to large insertion events appears similar between different populations. Three of the six insertions were found between the two heat shock transcripts, bringing the total to seven out of ten insertions within 1 kb spacer between the heat shock transcripts. The distribution of insertions throughout the heat shock locus is also similar between populations.

The distribution of elements at the level of restriction fragment analysis appears to be general between different populations and between different loci throughout the genome. At the cytological level transposable element insertions appear to be uniformly distributed throughout the genome and unrelated by descent. With restriction fragment analysis this also seems to be the case. Each insertion in the heat shock survey presented here appeared different, except for the two insertions (IIa and IIb, figure 4.23) which could not be distinguished. Each of the insertions found in the other surveys have been found not to be related by descent. Five chromosomes in the *Adh* survey (Aquadro 1986) contained insertions of the same element, 1261, which were indistinguishable at the restriction fragment level. Sequencing experiments have shown each of these to be inserted between different nucleotides of the *adh* sequence and so must represent repeated integration rather than descent from a common ancestor (Aquadro C., Quattlebaum W., Billings D. and Langley C.H. unpublished results). Without sequencing the ends of insertion IIa and IIb (figure 4.23) the question of identity by descent must remain unanswered.

6.4.2. The X Chromosome – Autosome Comparison

As pointed out by several authors recently (Eanes *et al.* 1985; Mackay 1986a; Ajioka and Eanes 1987) a knowledge of the distribution of insertions on the X chromosome, and an analysis of their effects would be valuable since the effects of directional selection are expected to be more rapid against deleterious mutations on the X chromosome rather than on autosomes, simply because in males the X chromosome is hemizygous (Morton 1971). The effective degree of dominance for alleles at an X-linked locus is given by (2/3)h + 1/3, where h is the degree of dominance of the allele in the diploid state. For mutations with a low degree of dominance, such as novel P element insertions (Mackay 1986a), the observed difference should be greatest between X-linked and autosomal loci. The effective population size for X-linked alleles is also lower than that of autosomal loci if the number of males and females are approximately equal (Crozier 1976). According the the neutral theory (Kimura 1983) a reduction in effective population size would also produce a reduction in variability. The fact that such a reduction could be explained by selection or random drift should be borne in mind.

The variability of proteins at autosomal and X-linked loci has been studied in the three organisms: man, *Drosophila* and the red kangaroo (Cooper *et al.* 1979). No conclusive evidence for a reduction in variability on the X chromosome was produced. The heterozygosity of loci on either type of chromosome did not differ significantly but the proportion of polymorphic loci was, however, slightly lower on the X chromosome. Assuming that electrophoretic protein variants are selectively neutral (Kimura and Ohta 1971) it would appear that the reduced effective population size of the X chromosome in *Drosophila* produces little significant reduction in variability.

A large reduction in variability for null alleles of metabolic enzyme loci on the X chromosome has been observed in *Drosophila melanogaster* (Voelker *et al.* 1980; Langley *et al.* 1981). For over twenty autosomal loci in about 500 to 1000 chromosomes for each of two populations the mean frequency of null alleles was 2.5×10^{-3}. On the X chromosome, for the five loci surveyed, no null alleles were detected. The null alleles surveyed were expected to be deleterious and in some cases lethal as homo- or hemizygotes (Voelker *et al.* 1980; Langley *et al* 1981). These alleles have a low degree of dominance (Voelker *et al.* 1980) which is expected for such enzymes (Kacser and Burns

1981). The difference in the number of null enzyme alleles then appears to be due to the action of increased purifying selection on the X chromosome rather than random genetic drift.

In order to determine whether the insertions of transposable elements are generally deleterious in natural populations the AS-C region of the X chromosome was examined. In this region from two populations there were 15 insertions in 120 kb from 49 chromosomes. Insertion IV (figure 4.25) occurred six times. Again sequencing of the ends of these insertions may be required to determine if they could be related by descent. By comparing the frequencies of insertion at the different loci it has been shown that the number of insertions between the heat shock, *Adh* and AS-C loci are significantly different with a chi-square value of 30.37 and a critical value of 5.99 (see Chapter 5). Secondly it has been shown that the number of insertions at the *Adh* and heat shock loci are not significantly different from that expected to be due to transposable elements (chi-square values of 2.46 and 2.78 respectively with a critical value of 3.84) while the AS-C region contains significantly fewer (chi-square of 36.57 with a critical value of 3.84, see Chapter 5).

The proportion of polymorphic nucleotides found at the AS-C locus were 7.41×10^{-3} and 5.08×10^{-3} for the two populations surveyed. These values are not very different from those found at other loci on the autosomes of *Drosophila* (see Chapter 5). This type of variation behaves in a very similar fashion to the protein electrophoretic mobility variation examined between the X and autosomes. It would seem reasonable that these nucleotide substitutions are effectively neutral (Kimura and Ohta 1971) and that the similarity between the X and autosomal loci examined shows that the increased drift component for X-linked loci is not very significant in the American and Spanish populations surveyed. The paucity of insertions at the AS-C locus could therefore reflect the fact that mobile element insertions produce deleterious mutations which are predominantly recessive.

6.4.3. The Non-Uniform Distribution of Insertions

The expected numbers of insertions for the chi-square tests were based on the assumption that DNA sequences are inserted uniformly over the genome. The distribution of insertions found at the AS-C and HS loci, most

noticeably, is surprising in that most of the insertions are very close to transcripts (see figures 4.23, 4.25, 4.26 and Chapter 5). This non-uniformity may invalidate the chi-square test and mean that the reduction in number of insertions at the AS-C locus simply reflects that a larger span of DNA not close to transcription regions was surveyed. If this is true how can the distribution of insertion be explained?

A simple view of how insertions might cause deleterious effects entails the disruption of transcription regions in such a way that a gene product is altered, either in structure or regulation. If selection were the force responsible for the observed reduction in the number of insertions at the AS-C locus, and elements inserted themselves uniformly throughout the genome, then one might expect insertions close to transcripts to be removed more rapidly from the population (Morton 1971). Extant insertions would be expected to accumulate in the regions between transcripts. The opposite appears to be true for the insertions observed so far. Of the 36 insertion events of more than 0.5 kb observed in the three loci surveyed (HS: Leigh Brown (1983) and this presentation; *Adh*: Aquadro *et al.* 1986 and AS-C: this presentation) 28 are within 5 kb of a known transcript. One explanation for this pattern, which seems unlikely, could be that insertions not within 5 kb of a transcript have larger phenotypic effects than insertions elsewhere. Alternatively it may be argued that even insertions close to transcripts might not affect the phenotype significantly. If this were true then there would be no reason to expect a reduced frequency of insertions between transcripts if elements were inserted uniformly in the genome.

There is a growing volume of evidence which suggests that the insertion of transposable elements is not random with respect to the DNA sequence. Examples from other organisms, especially bacteria are numerous (eg Kleckner *et al.* 1979). In yeast the TY element mutations are clustered within the regulatory sequences rather than within coding sequences (Roeder and Fink 1983). In *Drosophila melanogaster* there is also evidence for site specificity of insertions of the P element (O'Hare and Rubin 1983), *gypsy* (Freund and Meselson 1984) and 297 (Young and Rubin, unpublished cited in Shapiro 1983).

Such preferential integration may not be so surprising since the insertion of DNA into the genome may be facilitated by a target site which has a more

open chromatin structure. The DNA surrounding and especially 5' to transcription regions has been shown to be more accessible to external agents such as DNase I (Weintraub and Groudine 1976, Stalder *et al.* 1980), S1 nuclease (Larsen and Weintraub 1982) and topoisomerase II (Udvardy *et al.* 1985; Rowe *et al.* 1986). In these last two papers the sensitivity of the DNA surrounding the 87A7 HS genes of *Drosophila melanogaster* has been examined. It was found that the 1 kb spacer region, where seven of the ten insertions have been found (Leigh Brown 1983 and this presentation), is unusually accessible to the DNA cleavage activity of topoisomerase II.

The simplest explanation of the pattern of insertions found at the molecular level is that DNA is inserted preferentially at positions which are presumably more accessible to the insertion process. Such non-uniformity would not be detectable by the *in situ* hybridization methods described above since the clustering of insertions occurs below the level of resolution of these techniques. Although the AS-C region may be exceptional from the point of view of the distribution of insertions, observations at the *white* and *notch* loci, also on the X chromosome, suggest that the small scale inhomogeneity of extant insertions is a general phenomenon in *Drosophila melanogaster* (Aquadro 1986). Despite the non-uniform pattern of insertions the number of insertions at the AS-C locus still appears lower than expected, since there are many known regions of transcription within the region while only two, or three, apparent 'hot spots' for insertion.

6.4.4. The Effects of Reduced Recombination

The theoretical models of Charlesworth and Charlesworth (1983) and Brookfield (1982) in which the number of elements is limited by decreasing host fitness with increasing copy number predict that the effects of each insertion must be more than simply the additive effect of each of the individual insertions. Such an increase in effect would be observed if each insertion was closely linked to other insertions with a similar effect (Charlesworth and Charlesworth 1983). In this paper, Charlesworth and Charlesworth conclude from Monte Carlo simulations that linkage has a negligible effect on the simulated distribution of insertion sites. This is almost certainly true when only one family of elements is taken in isolation. The figure for the expected number of insertions per nucleotide, as estimated from

the proportion of the *Drosophila* genome which is apparently mobile, would give an expected number of 63 insertions for the chromosome samples screened in the AS-C locus survey here. Since only 49 chromosomes were analysed it becomes obvious that there should be at least thirteen cases in which two insertions are found on the same chromosome ie genetically linked. One chromosome did contain two insertions (II and IV, see figure 4.25) within 20 kb of each other.

From both population samples examined at the AS-C locus it was found that for the eleven pairs of polymorphic events for which sample sizes were above 20, seven were in significant disequilibrium (see Chapter 5). An upper limit for the recombination fraction between *achaete* and *scute* has been given as 6.6×10^{-5}. These loci are separated by 25 kb. The high levels of disequilibrium found at the AS-C locus are therefore not unexpected and are probably due to random genetic drift (Hill and Robertson 1968) since the inverse of the recombination fraction is of a similar order to the estimates of the *Drosophila* effective population size in North Carolina (Mukai and Yamaguchi 1974; Kreitman 1983). The recombination fraction between sites of insertion of mobile elements in natural populations may be of the order of 100 fold lower than the figure of 0.003 used in the simulation procedure of Charlesworth and Charlesworth 1983.

Gametic disequilibrium could give rise to a higher degree of apparent additive genetic variance than the value of the genic variance suggests (Lande 1976). If alleles are linked in the coupling phase, ie on the same chromosome, then the disequilibrium will be positive and will increase the additive genetic variance component observed. This could lead to an enhancement of the differential selection on the X chromosome compared to the autosomes. Elements at different loci in the same zygote, each inducing an effect on fitness, would be exposed to a higher degree of selection than either would on its own. After one generation of random mating elements in *trans* would be separated. Elements in *cis* would be exposed to this higher degree of selection for the time until recombination separated the two, or either were lost from the population. It is this which gives rise to the apparently synergistic effects between two alleles linked in the coupling phase.

Assuming that each insertion arises once in the population, and therefore

has an initial frequency of s = 1/2N, where N is the population size, the great majority of new insertions will be lost due to the process of random drift. The average time to loss for a neutral allele may be calculated from

$$t_0(p) = -4N(p/(1-p))\log_e p$$

(Crow and Kimura 1970). With a population size of about 10^6 individuals (Mukai and Yamaguchi 1974 and Kreitman 1983) and an initial frequency of 1/2N the mean time to loss would be about 55 generations. This is not long for recombination to separate two very closely linked alleles. The vast majority of elements will therefore be lost from the population in the same haplotype in which they entered it.

It should be pointed out here that another experiment in which the numbers of transposable element insertions on the X chromosome have been compared to those on the autosomes has been conducted (Montgomery *et al.* 1987). By using the technique of *in situ* hybridization the numbers of the elements 297, 412 and B104 have been analysed in 20 larvae from individual wild-derived females. For all three elements the numbers of elements found on the chromosome arms were not significantly different. When the variance in number was examined, however, the values for the two elements 297 and 412 were approximately the same as the mean number. This is what would be expected if the pattern of insertions followed a Poisson distribution. For the element B104 the variance on the X chromosome was 3.73 with a mean copy number of 11.45. For the other, autosomal, chromosome arms the variance estimates were similar to the mean.

The number of copies of a transposable element in an individual can be regarded as a quantitative character with a heritability of one (Charlesworth and Charlesworth 1983). Under directional selection, as is assumed to occur in the model of Charlesworth and Charlesworth (1983), negative disequilibrium is expected to be generated (Bulmer 1971). Using computer simulation it has been shown that this theoretical result, assuming infinite population size, is a good approximation even when the population size is not infinite (Keightley and Hill 1987). Negative disequilibrium generated by selection of element insertions would be observed as a reduction in the variance of copy number and it may be that this is the explanation for the observed variance reduction

for the number of the element B104.

6.5. Summary

The number of insertions found in 120 kb of DNA at the AS-C locus in 49 X chromosomes was significantly lower than expected from a) the number of insertions found at the heat shock locus here and in two previous surveys at autosomal loci; b) the number of transposable elements expected per nucleotide, calculated from the proportion of the *Drosophila melanogaster* genome which is apparently mobile and the average length of such mobile elements. The reason for the low number of insertions appears partly to be due to a non-uniform distribution of insertions in the DNA. This non-uniformity cannot easily be explained as the result of selection but is probably due to preferential insertion of DNA close to transcripts where the chromatin is more accessible to external agents. Although more experiments need to be conducted to determine whether such a non-uniformity of insertions is general and exactly what form the non-uniformity takes, it still appears that the number of insertions at the AS-C locus is lower than the number found on the autosomes. This evidence suggests that although regulation of transposition rate probably does occur selection plays a significant role in determining the distribution of transposable element insertions on X chromosomes in natural populations.

The reason that the effects on fitness which reduce the number of insertions on the X chromosome are not easily detectable in wild populations can be ascribed to the high rate of spontaneous deletion of element insertions. According to the models of Langley *et al.* (1983) and Charlesworth and Charlesworth (1983) the observed low frequency of any one site being occupied results from a very high spontaneous deletion rate. The distribution of elements in the population is expected to be determined mainly by high transposition rate rather than selection or random drift. Selection may have little effect on autosomal insertions since each is rare and that stochastic effects of finite population size and spontaneous deletion predominate, especially if most of the selective effect of insertions is confined to homozygotes (Mackay 1986b). If the population size were drastically reduced then the value of θ could become small ($4Nu < 1$). In this situation, although transposition is expected to continue at the same rate the number of excisions

per site becomes less than one per generation and the mutations induced by the insertion will be apparently more stable. The frequency of homozygotes will increase and some sites of insertion may become fixed (Langley *et al.* 1983; Biemont 1986). Under these conditions the effects of insertions on fitness components would become significant.

This is the opposite to what might be expected for additive variation due to nucleotide substitutions where the action of selection is more effective in large populations than in small ones. Under the model of effectively neutral alleles (Kimura 1979) an allele may be considered effectively neutral if $s < 1/2N$, where s is the coefficient of selection and N is the effective population size. As population size increases the stochastic process of Wright-Fisher sampling decreases and the effect of selection may become significant. Under the same conditions, genetic variance due to transposable element insertion would appear unstable by comparison and would behave as if effectively neutral.

CHAPTER 7

CONCLUSIONS

The purpose of this project was to use restriction enzymes to compare the molecular maps of different wild-derived chromosomes. With the aid of known restriction maps for the two regions chosen it was possible to interpret any differences from the concensus map as either restriction site changes or as the result of an insertion-deletion event. By comparing the variation found on the X chromosome with that in the autosomes it was hoped that any effects of selection on the distribution of transposable element insertions would be detectable.

The two regions chosen for study were the third chromosomal 87A7 heat shock and X-linked *achaete-scute* complex loci. The heat shock locus survey was conducted along very similar lines to a previous one (Leigh Brown 1983) with the aim of comparing the two populations surveyed for their insertional variability. By comparing the results presented here with those from a second autosomal region - the second chromosomal *Adh* locus (Aquadro *et al.* 1986)the generality of variation due to insertion events over different regions of the genome could be determined.

With the AS-C region it was possible to examine a much longer, and more completely characterized, section of the chromosome than has been possible before. Secondly it was possible to compare the pattern of both nucleotide polymorphism and insertional differences between the X chromosomal and the autosomal loci studied. The majority of novel P element induced quantitative variation has been found to be deleterious and largely recessive (Mackay 1986a; Fitzpatrick and Sved 1986). If transposable element insertions generally produce effects of this type then the more rapid action of selection against these on the X chromosome should result in a reduction in their number.

7.1. Summary of the Survey

In the heat shock survey there were six insertions in 32 chromosomes within 25 kb of the two heat shock transcripts. One of these has been shown to hybridize to genomic DNA fragments characteristic of a mobile element. Of these six events three occurred within the central region of the two transcripts

(see figure 4.23). Two of these insertions could not be distinguished from each other. In the previous survey (Leigh Brown 1983) a similar pattern had been found: all four large insertions in 29 chromosomes were between the two transcripts. The two populations were not found to differ significantly in the observed frequency of insertions at this locus. Also for the two known polymorphic restriction sites in this region they did not differ significantly in the frequency of allelic types.

At the second autosomal locus eleven insertions in 13 kb from 48 chromosomes were found close to the *Adh* transcript. Again this number of insertions is similar to that found at the heat shock locus. All the insertions from the three surveys which have been analysed so far have proven to be homologous to transposable elements. It has been shown that the numbers of insertions in the surveys at each of these loci are not significantly different from those expected from a uniform distribution of mobile elements throughout the genome. The level of restriction site variation at the *Adh* locus also appears to be very similar to that found at the heat shock locus (Kreitman 1983; Aquadro *et al.* 1986). The similarity in the degree of variability for the two different classes of variation extends not only between populations but also between autosomal loci.

With the AS-C locus survey the numbers of insertions were much lower than expected. In 120 kb at this locus in 49 chromosomes only 15 insertion were found when the expected number of transposable element insertions was 63. Two of these insertions appear were homologous to DNA which is repeated in the genome. This difference is significant. Also the numbers of insertions throughout the three loci studied are significantly different from each other and this difference derives mainly from the low number of insertions on the X chromosome. The AS-C locus displayed very similar numbers of polymorphic nucleotides, and had very similar estimates of heterozyosity per nucleotide, to the autosomal loci studied. It is thought that the majority of nucleotide substitutions observed by the changes in pattern of restriction fragments are influenced mainly by the finite size of the population rather than by selective pressure. If this is true then there appears no difference between the X chromosome and autosomes caused by a reduced effective population size of the X chromosome. Differences in the degree of insertional variation must therefore be due to processes other than random

drift.

7.2. The Non-uniform Distribution of Insertions

The most notable feature of the pattern of insertions found so far is that the majority of insertions have been within short distances of known transcripts. For the heat shock and *Adh* surveys only short stretches of DNA were examined and so only variants close to the known transcripts could have been observed. At the AS-C locus, however, a larger contiguous stretch of the genome has been analysed. Again the insertions have been found close to known transcripts. The difference in the number of insertions between this region and the two autosomal loci surveyed is manifest as a paucity of insertions throughout most of the region but with the observed insertions all clustering closely together into two regions. The models which try to account for the distribution of transposable element insertions throughout the chromosomes of *Drosophila melanogaster* in natural populations assume a uniform distribution of sites at which elements may insert throughout the genome. The distance over which the non-uniformity of insertions has been observed at the AS-C locus is too small to be resolved by the *in situ* technique. This heterogeneity would therefore not be visible with this method of analysis, which is presumably why it has not been reported before.

If novel insertions are deleterious then the number of insertions on the X chromosome is expected to be lower than on the autosomes. Can the observed lower number and non-uniformity of insertions be explained as the action of selection? From the known properties of transposable elements it would seem reasonable to assume that, in general, insertions closer to transcripts should have a greater phenotypic effect. Since natural selection acts on the phenotype rather than the genotype, insertions which lie closer to transcripts should respond more rapidly to the action of selection and if deleterious be removed more quickly from the population. The result of such selection would be that extant insertions would tend to lie further, rather than closer, to transcription regions. This is the opposite to what seems to occur in nature as observed from the chromosome samples examined.

An alternative explanation for the observed distribution of insertions is that mobile DNA is inserted preferentially close to transcripts. The most plausable explanation for such a phenomenon is that the sequence of events

which must occur during the process of insertion are facilitated by the more accessible chromatin structure to be found around transcripts. This greater accessibility has been shown by using agents such as DNaseI and S1 nuclease (Weintraub and Groudine 1976; Larsen and Weintraub 1982). Seven of the ten insertions which have been found at the HS locus lie between the two known transcripts. It is this region which is most accessible to the action of the enzyme topoisomerase II (Udvardy *et al.* 1985: Rowe *et al.* 1986).

7.3. Transposable Elements and Evolution

All the insertions so far examined in natural populations at the restriction fragment level have been unique in the population sample. At the cytological level insertions have been found more than once at certain locations. The reason for this appears to be a lack of resolution of the *in situ* technique where the insertions were in fact distinct. It is not possible without further work to establish if the insertions IIa and IIb of the heat shock survey, IIIa and IIIb and IVa to IVf of the AS-C survey are related by descent or appear related because of repeated insertion to very closely separated sites. In the *Adh* survey one insertion apparently occurred five times in the sample. It has been established by DNA sequencing that the reason for this is repeated insertion since the events all occur at slightly different positions in the sequence (see Chapter 6).

Examination of the AS-C locus did not find any significant differences in sequence arrangement from the map of this region produced by Campuzano *et al.* (1985) other than the, individually rare, large insertions. In a comparison between several *Drosophila* species at the 87A7 and 87C1 heat shock loci no differences were found to be due to transposable element sequences. Specific insertions do not then appear to gain any appreciable frequency in populations. This would preclude them from inducing genetic changes which become incorporated into populations and so give rise to genetic differences between species. Nucleotide substitutions on the other hand are the cause of the majority of structural differences in proteins between different species. New nucleotide variants are continually created by mutation and some of these variants become fixed in populations where they may be observed at the restriction site level as the appearance or loss of a restriction site between populations or species.

From the available models for the population dynamics of transposable elements it would appear that in *Drosohpila* the very high mobility of these sequences is a major factor in the maintenance of the very low frequency of each insertion. Although this project failed to find any significant evidence for selection determining the distribution of insertions on the X chromosome it did show that the distribution of insertions was not uniform, as has been assumed previously, but that insertions are distributed in clusters which appear to be close to transcription regions.

The mutations due to single base changes are usually caused by lesions within the coding, splicing or regulatory regions of transcripts. Transposable elements on the other hand have been found to be able to cause mutations when they are inserted many kilobases away from transcripts, or into introns where the sequence of DNA does not appear to be significant. The proportion of the genome in which element insertions can cause mutations is large by comparison with that in which single base changes can. For this reason element insertions are more likely to effect a change in phenotype than nucleotide substitutions are. If elements do become inserted more frequently close to transcripts then they will presumably disrupt the sequences most necessary for the correct expression of genes. The effect of insertions on the phenotype would then be greater still than might be expected from a uniform distribution of elements. A second reason why insertions do not attain any real frequency could be that they induce greater deleterious effects than do nucleotide substitutions.

7.4. Suggestions for Further Work

The non-uniform distribution of insertions in the DNA sequence was unexpected and as a consequence the difference between the X chromosome and autosomes can not be attributed to the effects of selection. The distribution observed appears to be the opposite of what might be expected to be due to the action of selection. The heterogeneous pattern of insertion does appear to be a common feature of X-linked and possibly autosomal loci since this non-uniformity has been found when analysing the *white* and *notch* loci in *Drosophila melanogaster* natural population samples (Aquadro 1987). The distribution of insertions over longer stretches of DNA must be examined to determine the underlying cause of their clustering. Once this is understood the

role of selection in determining the distribution of insertions may be established.

Although study of the heat shock locus between species found no difference in sequence arrangement due to transposable elements, the 87C1 heat shock locus of *Drosophila melanogaster* appears to contain an insertion of a stretch of satellite DNA between the two heat shock transcripts (Leigh Brown and Ish-Horowicz 1981). The position of satellite sequences do not appear to vary significantly between strains of *Drosophila melanogaster* (Hennig *et al.* 1970). The spontaneous deletion rate of these sequences is consequently assumed to be low. Chance insertion of the αβ repeated satellite between the transcripts would therefore generate a relatively stable insertion which could then drift upward in frequency or indeed be selectively incorporated into the gene pool of the species. To determine if transposable elements can induce mutations which become incorporated into the gene pools of different species it would be informative to examine complementary regions of the genome in several different species. So far in studies of this kind only single samples representing a species have been used. Some estimation of the within species variation must be made before it is possible to understand the significance of any variation found.

Since it is known that mobile elements can produce small insertions and deletions upon excision (Voelker *et al.* 1984) the variation between individuals for these events should be examined. In both surveys presented here insertion-deletion events smaller than 500 nucleotides have been found, although it is difficult to establish precisely the nature of events A and B in the AS-C locus survey (figure 4. 24). Such small insertion-deletion events have also been found at the *Adh* locus ranging from a few nucleotides to several hundreds (Kreitman 1983; Aquadro *et al.* 1986). It has not been shown that the small insertion-deletion events observed are causally linked to mobile elements but the possibility deserves some investigation. By cloning small insertion-deletion events it may be possible to determine whether they show any homology to any of the mobile elements or if they have similar characteristics to the small changes produced when transposable elements excise. The excision of an FB element from the *white* gene, to produce a phenotypic reversion precisely removes the element (Collins and Rubin 1983). Excision of the element may occur imprecisely but such an event would not

necessarily alter the phenotype. There remains the possibility that although mutations caused by the insertion of a specific element may have no long term evolutionary future, the excision derivatives of these may retain the mutant phenotype while apparently losing the inserted element. Small insertion deletions may be the way in which transposable element induced variant could be incorporated into the gene pool of a species.

Small insertion-deletion events have been found at appreciable frequencies in populations and they may have similar dynamic properties in populations to nucleotide substitutions. If so it is expected that some such differences should become fixed in populations and so be responsible for some of the genetic differences between species. Again a comparison between the X chromosome and autosomes would be useful in the determination of the effect selection has on these.

If it is correct that element insertions in natural populations are effectively neutral because of their high degree of mobility it should be possible to measure the effects of insertions derived from wild chromosomes on phenotypic characteristics. The AS-C region would have provided an excellent choice for such a study. Using known genetic techniques available with *Drosophila melanogaster* it should be possible to recombine the telomeric section of the wild-derived chromosomes into otherwise genetically identical backgrounds. The wild-derived insertions could then be analysed genetically since any alteration in the characteristic bristle phenotype should mainly be due to variation at the AS-C locus. Unfortunately such a project is not possible with the variants revealed in this survey since many of the extracted X chromosome stocks containing insertions at the AS-C locus have become extinct.

Variation due to large element insertions in *Drosophila melanogaster* appears to be a common phenomenon between different regions of the genome and different populations. This variation appears to follow a novel population genetic process because of the very high instability of the insertions responsible. Although the variation observed to be due to novel transposable element insertions in the laboratory is extensive the significance of this in natural populations is unclear. The distribution of insertions at the AS-C locus, and when compared to that found at autosomal loci, does not

appear to be affected by selection greatly. The differences between the X chromosome and autosomes may be explained in terms of non-uniformity of integration of DNA sequences. If transposable element insertions are effectively neutral, and also never attain any appreciable frequency then variation induced by such insertions may be insignificant in evolutionary terms. More investigation into the non-uniform distribution of transposable elements at the DNA sequence level needs to be conducted before the significance of insertional variation can be ascertained.

I. Details of the chromosome surveys

Presented here are the details and observations of each of the genomic DNA fragment autoradiographs. The analysis of the λ clones are also given. This section is only intended as a reference for Chapter 4.

The lengths of fragments revealed by the various probes used in the genomic surveys were estimated using the 'spline' programme as described in Chapter 3, except where stated otherwise. A single line was chosen arbitrarily from those displaying the most common pattern as a representative and the fragment lengths were estimated from this line. The lengths of fragments migrating unusually were also estimated in each line where they occurred and the values obtained are given in tables 4.2 to 4.6.

I.I. 87A7 Heat shock locus

Genomic DNA fragments

Co-ordinates 0.0 to 10.9

Three different 0.3% agarose gels were used for the electrophoresis of *Eco* RI fragments of the CM line genomic DNA. The filters from each were probed with plasmid 56H8/C. A single fragment was disclosed in each line except line CM11 where the DNA did not appear to digest properly. λ *Hind* III and λ *Eco* RI marker fragments were only visible on one of the autoradiographs (figure 4.1), but there were at least three DNA samples in common between any two of the three gels which could be used as internal size standards.

Of the 31 lines in which a band was visible, three possessed a fragment markedly different in size from the most common type. The fragment from line CM16 had an estimated size of 15.17 kb. The most common fragment size was 10.95 kb. Lines CM13 and 30 both appeared to have a fragment about 2 kb smaller than the most common type, although the lack of marker fragments did not allow a direct calculation of size. The band in line CM22 appeared to migrate slightly faster than the bands in neighbouring lanes on the gel shown in figure 4.1 but not noticeably so on another. The estimated size of this fragment was 10.55 kb.

A single 0.4% agarose gel with *Eco* RI and *Bam* HI fragments from lines CM13, 16, 21 and 30 was transferred and probed with pBF17 (figure 4.2). The line of fragments present in all eight lanes can not be easily explained on the basis of the molecular map (figure 4.23). They could have been due to contamination of the samples with a very small amount of plasmid DNA. Apart from this a single *Eco* RI fragment was revealed in each line as expected from the known map of the region (figure 4.23). CM16 displayed a fragment larger than the fragment from line CM21 which in turn was larger than the fragment in both lines CM13 and 30. These latter two were in adjacent lanes on the gel and appeared to to be about 7 kb long.

Co-ordinates 2.8 to 7.3

In each of the four *Bam* HI digested samples two fragments were produced. The lower band was invariant in each of the four lanes and the larger, in lines CM13 and 30, appeared to co-migrate at about 9 kb. The upper fragment in line CM16 was intermediate between these last fragments and that seen with line CM21, with a size of roughly 6 kb. These size estimates were made without λ markers using the knowledge of the size of fragments estimated from other gels.

Co-ordinates 8.6 to 24.3

Two 0.3% agarose gels of *Xho* I digested DNA were transferred and probed with p87A/5. A single fragment was observed except for lines CM11, 12, 14, 29 and 32 where no band was visible. Only the fragment from line CM21 migrated unusually (figure 4.3), with an estimated size of 5.39 kb instead of 10.16 kb. Size markers of a mixture of λ *Hind* III and λ *Bgl* II fragments were used.

A single *Xba* I fragment hybridizing to p87A/5 was visible in every line, except CM32, with filters from a 0.3% gel. The most common size, estimated from λ *Bgl* II markers was 16.61 kb. Three lines, namely CM4, CM21 and CM24, appeared to have larger fragments than the most common type (figure 4.4). The size estimates of these were 20.24 kb, 21.74 kb and 20.15 kb respectively.

Co-ordinates 2.8 to 8.8

Bam HI digested genomic DNA was electrophoresed through two 1.0% agarose gels alongside λ *Pst* I DNA fragments. The filters from these gels were probed twice, first with p87A/5. A single fragment was visible with each line, except CM6, and each appeared to be of the same size. When these filters were rehybridized to pBF17 only faint bands were visible on one autoradiograph. Bands from lines CM17 to CM27 could be seen and all except CM22 appeared to be the same size, 2.54 kb. The fragment in line CM22 was estimated to be 2.39 kb long.

Polymorphic *Pst* I Site

Filters from three 0.6% agarose gels with genomic *Pst* I fragments and λ *Hind* III and λ *Bgl* II markers were probed with p56H8/C (figure 4.5). For the lines CM9, 11 and 25 no band was visible. Of the remaining lines, CM1, 7, 10, 22, 23, 26, 27 and 32 possessed a fragment of 4.11 kb. Fragments from the others appeared to migrate faster, with a size of 3.49 kb. Filters from two 1.0% gels with λ *Pst* I markers, probed with p87A/5, produced a single fragment invariant throughout the sample. No band could be scored for lines CM11, 14 and 16.

Cloned DNA Fragments

λ phage libraries were constructed of genomic DNA from lines CM4, 13, 16, 21 and 24 as described in Chapter 2. Too little DNA remained from the genomic survey to do the same with line CM30. Single plaques homologous to p903a were observed in the libraries from both lines CM4 and 13 respectively. In neither case were these phage represented in the sample subsequently screened on smaller plates.

Insertion I

A single clone homologous to pBF17 was isolated from line CM16 (λ6CM16). Although no restriction map was produced for this phage, the DNA was digested with *Bam* HI, *Sal* I, *Eco* RI, both *Bam* HI and *Sal* I and both *Sal* I and *Eco* RI. The fragments were transferred onto nitrocellulose after separation with a 0.4% agarose gel. Two *Bam* HI fragments were disclosed by pBF17 of

6.58 ±0.29 kb and 2.00 ±0.04 kb (figure 4.6). With *Sal* I a single band of 5.23 ±0.16 kb and in the double digest one of 4.74 ±0.09 kb were observed (figure 5.6). *Eco* RI did not appear to cleave the DNA at any point. These fragments agree with the molecular map (figure 4.23) with the inclusion of an insertion between positions 4.3 and 5.4.

Insertion III

A single phage clone, λ1CM21, homologous to p903a was isolated from line CM21 and a restriction map deduced. Hybridization of DNA fragments to p903a indicated that the DNA represented in this phage extended distally from the p903a homologous region only about 2 kb and therefore not far enough to include the insertion event in this line.

Insertion IV

Two clones homologous to p903a were purified for line CM24. One of these, λ1CM24, was radioactively labelled and hybridized to one of the filters of *Bgl* II digested genomic DNA produced in the AS-C locus survey below. Many bands of a range of sizes from below 2 kb to more than 20 kb were produced.

I.II. *Achaete-Scute* Complex Locus

Genomic DNA Fragments

Genomic *Bgl* II Fragments

Genomic *Bgl* II fragments were electrophoresed through 1.0% agarose gels with λ *Hind* III and λ *Pst* I fragments as size markers. Four gels were used for the 50 samples, two for the North Carolina lines and the *C(i)DX,y,f* stock and two for the Spanish lines. The first four gels were transferred onto nitrocellulose and the filters probed with λsc22 DNA. Four more gels were subsequently transferred onto 'Gene Screen Plus' (New England Nuclear) and these filters were hybridized with all the other probes used. These filters could be used many times more than nitrocellulose ones and even after at least fifteen rehybridizations there was no sign of deterioration of the transferred DNA. Sample NC24 did not digest properly on this second set of gels and so

only the *Bgl* II fragments hybridizing to λsc22 DNA were surveyed.

The sizes and standard error estimates for each fragment are given in table 4.3.

Co-ordinates -35.3 to -16.7

When probed with pASC53R1 a single band appeared in each line. This fragment was 6.80 kb long except in lines NC20 and NC22 (figure 4.7). The bands in these lines were 6.17 kb and 6.39 kb respectively.

With λsc53 there was a very high background of non-specific labelling which obscured parts of the autoradiographs. Two large fragments could be seen in all the lines, except NC24, which were invariant in most. The largest fragment in lines NC10 and 18 was 3.09 kb and 2.49 kb larger than the more common size of 8.54 kb. The second largest band varied in exactly the same manner as that produced by pASC53R1. Two smaller fragments could be seen in some of the samples. In lines NC2, 4, 6, 8, 9, 15, 16, 19, 20 and FV22 a fragment of 3.86 kb could be seen and in lines NC7, 14, FV1, 2 and 13 to 18 one of 2.43 kb

Co-ordinates -16.7 to 4.1

With pASC31P4 the most common pattern consisted of five bands. The sizes of these were 9.08 kb, 5.32 kb, 4.24 kb, 1.82 kb and 0.83 kb. The smallest of these was only seen with the FV line samples since this fragment had a sufficiently high mobility to have migrated beyond the end of the filters used to transfer the two NC line gels. Each band, except the second largest was invariant in size throughout the lines surveyed. In lines NC4, 6, 8, 15, 19 and 22 the 5.32 kb band was replaced by two new bands of 4.59 kb and 3.62 kb (figure 4.8).

Co-ordinates 8.8 to 27.3

The autoradiographs of the filters probed with λsc22 DNA were rather faint so several of the lines could not be scored for all the fragments seen with some of the lines. Each of the five fragments was invariably the same size between lanes where they could be seen. These sizes were 6.05 kb, 5.45 kb, 3.79 kb, 3.57 kb and 1.15 kb. The fragments not scored were the

largest fragment in lines NC14, 27 and FV7, second largest in lines NC14 and FV7, the three smallest in lines NC15 to NC18 and FV7 and the smallest fragment in lines NC19 to 23, FV7 and FV13 to FV22.

Co-ordinates 33.1 to 57.4

The most variation was revealed by pASC94R1 and pASC94R4. Of the five bands from the most common pattern, the largest band in lines NC25 and FV14 were larger than the rest with sizes of 11.50 kb and 13.84 kb respectively, rather than 9.18 kb. In line FV2 this fragment was also larger with a size of 11.18 kb and an extra band of 5.00 kb appeared. In line FV13 this largest band was absent but the next largest band was more intense than expected. It is likely that in this line fragments 1 and 2 co-migrated (figure 4.9). In line NC20 the 7.34 kb band was absent and coincided with the appearance of a new, smaller band of 1.47 kb. The 3.76 kb band was absent and two new bands of 2.35 kb and 1.55 kb appeared in lines NC4, 6, 8, 12, 15, 16, 19, 20, 26, 27, FV1 and 22 (figure 4.10).

Co-ordinates 57.4 to 73.8

Five bands were visualized by λsc112, the three largest of which were of a uniform size between lanes. In lines FV10 and FV18 the fourth largest fragment migrated slightly faster than usual with a size of 1.69 kb rather than 1.73 kb and in lines NC1, 2, 6, 8, 15, 19, 20 and FV1 the smallest fragment migrated slightly faster with a size of 1.61 kb rather than 1.64 kb (figure 4.11).

Co-ordinates 57.4 to 60.1 and 67.3 to 73.8

With pASC101R7 a single band, and with pASC133R1 two bands, in each lane were revealed which were invariant in size throughout the sample. The lower band from this second probe was not scored for lines FV9, 10 and 16 to 20.

Genomic *Bam* HI fragments

The size estimates and approximate standard errors are given in table 4.4.

A single set of filters from four 0.8% agarose gels with λ*Hind* III and λ*Pst* I marker fragments were used for the survey of genomic *Bam* HI fragments. In these gels the digests of lines FV1 and FV2 were lost and so were not surveyed with this enzyme.

Co-ordinates -36.4 to -28.5

With pASC53R1 a single fragment was disclosed in each line except FV13 where no band could be seen, probably because the bands on this autoradiograph were rather faint. The only lines which displayed a difference were lines NC10, 18, 20 and 22 (figure 4.12). In the first two lines the band visualized had a mobility rather less than the most common form with estimated sizes of 19.02 kb and 18.98 kb respectively. With lines NC20 and 22 this fragment had migrated just detectably faster than the rest, at 8.86 kb and 9.11 kb.

Co-ordinates -17.3 to -0.4 and 55.7 to 76.6

A set of five bands was revealed with a mixture of λsc31 and λsc112. These were 13.51 kb, 9.45 kb, 7.95 kb, 5.01 kb and 3.65 kb. The fourth largest fragment was visible in lines NC3, 8 to 11, 13 to 27 and the *C(i)DX,y,f* stock and lines FV5 to 12. The fifth largest was not visible in line NC1 and all of the FV lines. The reason for this was again probably because of rather faint bands. The only difference was that in lines NC4, 6, 8, 15, 19 and 22 the second largest fragment was absent and the largest band was more intense than expected (figure 4.13).

Co-ordinates 34.4 to 53.1

With a mixture of pASC94R1 and pASC94R4 a pattern of six bands appeared. The sizes of fragments in the common pattern were 5.86 kb, 4.71 kb, 2.81 kb, 1.36 kb, 1.24 kb and 0.85 kb. Not all these bands were visible in each line. The largest fragment was the only one seen for lines FV13 to FV22. The three smallest fragments for lines FV3 to FV9, the smallest from line NC12 and

all the FV lines and the third largest from line NC25 could not be scored. In line NC12 the second largest fragment of 4.71 kb was absent and two smaller fragments of 3.54 kb and 2.12 kb appeared (figure 4.14).

Co-ordinates 53.1 to 58.8

Two fragments were visualized with pASC101R5 with sizes of 3.39 kb, 2.58 kb. Both of these could not be scored in lines FV13 to FV22 and both were invariant except that in line NC3 the smaller 2.58 kb fragment was absent and a fragment of 1.16 kb appeared.

Genomic *Xba* I fragments

Table 4.5 gives the estimated sizes and standard errors for these fragments.

The gels of *Xba* I digested genomic DNA proved not to be as successful as those for the previous two enzymes and many of the DNA samples were not digested completely. The filters hybridized to the different probes were often transferred from gels electrophoresed on separate occasions.

Co-ordinates -33.0 to -12.2

Three 0.25% agarose gels with λ, λ*Xho* I and λ*Hind* III fragments as markers were transferred and probed with pASC53R1. Two bands could be seen in most of the lanes. The lower fragment, of 1.85 kb, was invariant but could not be scored for lines NC1, 2, 8, 11, 12, FV4 to 8 and FV11 to 13. The upper fragment was invariant throughout the Spanish sample but there appeared to be three types in the North Carolina one. The pattern is most easily explained by lines NC5, 7, 9, 13, 14, 16, 17 and 23 having the same size of fragment as the Spanish lines. The sizes were 20.46 kb in the former case and 19.68 kb in the latter, which are not significantly different. Lines NC4, 6 and 15 had a slightly faster migrating band of 19.44 kb and lines NC10 and 18 slightly slower with sizes of 21.43 kb and 22.17 kb respectively. The larger fragment was not scored for lines NC1 to 3, 8, 11, 12, 19 to 22, 24 to 27, *C(i)DX,y,f* and FV7 and 12.

Co-ordinates -31.3 to 33.9

Two filters of 0.3% gels of the Spanish line digests were probed with λsc53 and λsc17 (figure 4.15). Marker fragments did not appear on the autoradiographs. Three bands could be seen with each line except for line FV1 where nothing was visible and FV13 where only the largest fragment could be scored. Of these three bands, the two larger were invariant but the smallest migrated faster in lines FV4, 7 and 20 than in the other lines.

Three 0.25% agarose gels with λ*Hind* III markers were transferred and probed with a mixture of λsc53 and λsc22 DNA. Three bands were produced with most of the samples (figure 4.16). The upper band appeared to vary in the same way as the largest fragment revealed with pASC53R1. Several lines where the larger fragment could not be seen previously were scored with these autoradiographs. Lines NC1 to 3, 11, 12, FV7 and 12 appeared to have the same form as line NC7 and line NC8 co-migrated with line NC4. The middle band seemed to be the same as the smallest fragment visualized with λsc53 and λsc17 as probes. Lines NC13, 19 to 27, *C(i)DX,y,f* and FV14 could not be scored for this fragment. Lines NC1, 2, 7, 10, 11, 14 and 18 had the faster migrating form with a size of 12.72 kb rather than 14.29 kb. The smallest fragment, of 11.31 kb, appearing in most of the lines was invariant throughout but could not be seen in lines NC19 to 27, *C(i)DX,y,f*, FV5, 13 and 14. An extra band could be seen in lines NC10 and 18 with sizes of 7.17 kb and 7.50 kb respectively. These were not significantly different.

A single filter from a 0.3% agarose gel with lines NC15 to 27 and *C(i)DX,y,f* was probed with pASC31P4 and pASC17B1. No size markers were visible but four bands could be seen in each lane, except for line NC26 where the DNA had not digested properly. The upper band, although rather faint, was again the same as the larger band seen with pASC53R1. Lines NC15, 19 and 22 appeared to have the faster migrating form of this band and line NC18 the slower form. The second and fourth largest fragments were invariant in size while the third varied in the same way as the lowest band seen with λsc53 and λsc17. In lines NC18, 23 and the *C(i)DX,y,f* stock this fragment migrated faster.

The filters hybridized with pASC53R1 were rehybridized with pASC31P4 alone. Two fragments were visualized in this case and as for the previous

hybridizations the largest fragment visible seemed to be the same as the largest one seen with pASC53R1. This band was scored for all the FV line samples and for lines NC4 to 7, 9, 10, 13, 14 and 18. The lower band was scored in lines NC 4 to 7, 9, 10, 13 to 17 and all the FV line samples except for FV7 and 8.

Co-ordinates 38.5 to 50.4

The filters probed with λsc94 were those previously hybridized with λsc53 and λsc17. The only variant of the single band observed in each line was in line FV2 where the fragment was slightly larger than usual with a size of 14.34 kb rather than 12.95 kb. Line FV1 produced no visible band.

pASC94R1 and pASC94R4 were hybridized to the filters used with pASC53R1. A single band appeared in lines NC3 to 7, 9, 10, 13 to 18 and all the FV lines except FV12. The only size variant was again FV2 (figure 4.17). This fragment was the same as that hybridizing to the intact λ phage from which these two plasmids were derived.

Co-ordinates 50.4 to 64.2

The three filters transferred from 0.3% agarose gels were hybridized with pASC101R7 and a single band was produced on the autoradiographs (figure 4.18). This band could not be seen with lines NC1 to 14, 26 and FV1. In lines NC16, 25, 27 and FV22 the size of this band was estimated to be 14.72 kb whereas in the other cases it was 12.68 kb.

Co-ordinates 64.2 to 81.9

Three 0.2% agarose gels with λ*Hind* III markers were transferred and probed with λsc112 DNA. A single invariant band was seen in lines NC5 to 7, 9 to 11, 13 to 18, 23, FV1 to 10 and 12 to 19. The filters from the two 0.3% gels of the Spanish samples were hybridized for a fourth time to pASC133R1. As with λsc112, a single invariant fragment was disclosed for lines NC1 to 10, 17, 19 to 27, *C(i)DX,y,f* and FV1. This fragment appeared to be the same in both cases with an estimated size of 17.74 kb.

Genomic *Xho* I Fragments

The size and standard error estimates for these fragments is given in table 4.6.

Co-ordinates -35.7 to -23.9

Four 0.2% gels with λ, λ*Xho* I and λ*Hind* III DNA fragments as markers were transferred and probed with pASC53R1. A single fragment was revealed of 11.78 kb in lines NC3 to 7, 9, 11, 13 to 17, 23 and FV2 to 22. In lines NC10 and 18 two bands appeared in place of the more common one with sizes of 9.89 kb and 7.74 kb (figure 4.19). In the remaining lines no fragment could be scored.

Co-ordinates -35.7 to -23.9 and -12.1 to 33.5

When rehybridized with λsc53 and λsc17 these filters produced a pattern of three bands for each of the lines where a band was seen with the previous probe except line FV22 where none could be seen. The lowest band was the same as that seen with pASC53R1. Again lines NC10 and 18 displayed a difference. While the two slowest migrating bands were the same as the common pattern, the fastest varied in exactly the same way as the fragments seen with the previous probe. No fragments could be scored for the other samples.

Co-ordinates 33.5 to 92.0

With the same filters and λsc94 as a probe a single band was produced in the same lines as before except for lines FV2 and 12 as well this time. The fragment was uniform in size throughout the sample except in line FV13 where it migrated faster and FV14 where it migrated slower than usual.

Filters from four 0.3% gels with λ and λ*Hind* III fragments as markers were probed with pASC94R1 and pASC94R4, which are homologous to the same fragments as λsc94. A very large fragment of 46.36 kb was visible in lines NC11, 13, 15 to 18, FV3 to 11 and 14 to 22. A smaller fragment of 12.32 kb appeared in lines NC3 to 7, 10, 11, 13 to 18, FV3 to 11, 15 to 22. Lines FV13 and 14 both had fragments which differed in the same way as the fragment disclosed by λsc94. The sizes of these fragments were estimated to

be 10.38 kb and 18.23 kb respectively. In line FV2 two fragments of 22.61 kb and 10.95 kb were apparently derived from the smaller of the two most common fragments (figure 4.20).

Cloned DNA Fragments

Genomic λ phage libraries were constructed from lines NC10, 18, 20, 22, FV2, 13 and 14.

Insertions I and II

Two plaques homologous to pASC53R1 were observed with NC20, and three with NC22, in the initial library screening procedure. No positive plaques were observed in the second screening of the samples taken from the library of NC20 but all three were purified for NC22. λ3NC22 was used to probe a *Bgl* II genomic digest filter and produced a pattern of many bands as λ1CM31 had done (figure 4.21).

Insertions IIIa and IIIb

Six phage were purified from each of the libraries prepared from lines NC10 and 18. These phage were homologous to the 5 kb *Xho* I - *Bgl* II fragment isolated from λsc53. Only three of each were also homologous to pASC53R1. One of these, λ2NC10, was radioactively labelled and used to probe a *Bgl* II genomic digest filter from this survey. Three bands appeared in the autoradiograph of 14.38 kb, 3.39 kb and 2.99 kb.

λ1NC18 which also hybridized with both these probe DNA fragments was used similarly. This time only two bands were produced, which seemed to correspond exactly to the larger two of the three bands seen with λ2NC10. The DNA inserted into both these phage contained no *Xho* I sites and both produced a large fragment of about 9 kb with *Bgl* II which was only cleaved from the arm of the phage vector with *Sal* I, which is expected to cut the remaining poly-linker sequences of the phage. λ2NC10 also contained a single *Bam* HI site close to one end of the inserted DNA. No such site was present in λ1NC18.

Insertion IV

Four further plaques which were positive with the *Xho* I - *Bgl* II fragment used to screen the libraries of NC10 and 18 were observed initially with line NC22. Only one of these, λ4NC22, was purified successfully. λ4NC22 was also homologous to pASC31P4. When this phage was hybridized to a *Bam* HI genomic digest filter, the only fragments disclosed appeared to be those homologous to either pASC53R1 or pASC31P4.

Insertion VII

Six plaques from the FV2 line library were homologous to pASC94R1. Only two of these were successfully isolated, λ1FV2 and λ3FV2. The first of these contained a 9 kb *Bgl* II fragment attached to the larger of the two phage arms which hybridized to pASC94R1, the second an internal 5 kb *Bgl* II fragment which hybridized similarly. These clones seem to contain the distal and proximal regions of the insertion. λ1FV2 was used to probe a *Bam* HI genomic digest filter and produced a pattern of five bands (figure 5.23). The largest of these appeared to be five to ten times the intensity of the smaller bands. Fragment lengths were estimated to be 7.34 kb, 2.91 kb, 2.81 kb, 2.15 kb and 1.21 kb.

Insertion VIII

From the FV13 library, two phage were purified from more than twenty which appeared to be positive with pASC94R1. Both these phage were homologous to pASC94R4. The DNA cloned in λ2FV13 contained two *Sal* I sites 6.7 kb apart and no *Bam* HI or *Xho* I sites. λ5FV13 on the other hand contained a single *Sal* I site with two *Xho* I sites on one side of this and one on the other. These clones contain sequences which are very different from the known restriction map for this region. It would appear that they might not contain DNA which was originally a contiguous stretch of the genome. The reason for this could be that the phosphatase step of the DNA preparation procedure (see Chapter 2) was ineffective and that the digested genomic DNA was thus able to ligate to itself. Concatenation of short DNA fragments would generate scrambled pieces of a suitable size for cloning.

Insertion IX

Four from six plaques, initially positive with pASC94R1, were purified for line FV14. A single phage, λ6FV14, was also homologous to a 1 kb *Hind* III - *Xba* I fragment isolated from λsc94. Southern transfer analysis of restriction fragments from this phage indicated that this fragment hybridized only to the very end of the inserted DNA. This clone did not appear to extend far enough to include the insertion known to have occurred in this line.

II. Solutions and Media

Buffer H

10 mM tris
100 mM NaCl
10 mM EDTA
15 mM spermidine
15 mM putreceine
pH 7.5

Buffer I

200 mM tris
30 mM EDTA
2% SDS
1 mg ml^{-1} pronase-E (Sigma)
pH 9.0

TBE

89 mM Tris(hydroxymethyl)aminobutane (tris)
89 mM Boric Acid
3 mM Ethylenediaminetetraacetic acid (EDTA) pH 8.0

TE

10 mM Tris
1 mM EDTA pH 8.0

TES

mM tris
mM EDTA
% SDS

Tfb I

30 mM potassium acetate
100 mM RbCl
10 mM $CaCl_2$

50 mM $MnCl_2$
15% glycerol
pH 5.8 Sterilized by filtration.

Tfb II

10 mM PIPES
10 mM RbCl
75 mM $CaCl_2$
15% glycerol
pH 6.5 Sterlized by filtration.

TMN

10 mM Tris pH 8.0
100 mM NaCl
10 mM $MgCl_2$

PSB

TMN made 0.05% with gelatine.

Phenol and chloroform

Phenol was redistilled from solid and kept in the dark at -20°C before being melted at 65°C, saturated with water and made 0.1% w/v with 8-hydroxyquinoline. Extraction several times with 0.2 M tris (pH 8.0) brought the pH value of the phenol to 8.0. It was then stored at 4°C. Chloroform was made 4% v/v with 1-pentanol. Phenol-chloroform was a 50 - 50 mixture of phenol and chloroform as prepared above.

L-Broth

1% w/v Tryptone (oxoid)
0.5% w/v Yeast extract (oxoid)
100 mM NaCl

L-Agar

L-broth supplemented with 0.1% glucose and made 1% w/v with agar.

L-top Agar (Agarose)

L-broth made 10 mM $MgSO_4$ and 0.5% w/v with agar (agarose).

T-Broth

1% w/v Tryptone (oxoid)
100 mM NaCl

ψ-broth

0.5% w/v yeast extract (bacto)
2% tryptone (bacto)
10 mM $MgCl_2$

References

1. Ahlberg J.H., Nilson E.N. and Walsh J.L. 1967 The theory of splines and their application. Mathematics in science and engineering **38**. New York Academic Press.

2. Ajioka J.W. and Eanes W.F. 1987 Manuscript in preparation.

3. Aquadro C.H., Deese S.F., Bland M.M., Langley C.H. and Laurie-Ahlberg C.C. 1986 Genetics Molecular population genetics of the alcohol dehydrogenase gene region of *Drosophila melanogaster.* Genetics **114:** 1165-1190.

4. Aquadro C.H. 1987 Manuscript in preparation.

5. Arber W., Enquist L., Hohn B., Murray N.E. and Murray K. 1983 Experimental methods for use with Lambda. Lambda II ed. Hendrix R.W. Cold Spring Harbour monograph series no.13. Cold Spring Harbour, New York.

6. Arber W. and Linn S. 1969 DNA modification and restriction. Ann. Rev. Biochem. **38:** 467-500.

7. Artavanis-Tsakonas S., Schedl P., Tschudl C., Pirrotta V., Steward R. and Gehring W.J. 1977 The 5s genes of *Drosophila melanogaster.* Cell **12:** 1057-1067.

8. Avise J.C., Lansman R.A. and Shade R.O. 1979 The use of restriction endonucleases to measure mitochondrial DNA sequence relatedness in natural populations I. Population structure in the genus *Peromyscus*. Genetics **92**: 279-295.

9. Barnes S.R., Webb D.A. and Dover G. 1978 The distribution of satellite and main-band DNA components in the *melanogaster* species subgroup of *Drosophila*. Chromosoma **67**: 341-363.

10. Benton W.D. and Davis R.W. 1977 Screening λgt recombinant clones by hybridization to single plaques *in situ*. Science **196**: 180-182.

11. Bender J. and Kleckner N. 1986 Genetic evidence that Tn10 transposes by a non-replicative mechanism. Cell **45**: 801-815.

12. Bender W., Akam M., Karch F., Beachy P.A., Peifer M., Spierer P., Lewis E.B. and Hogness D.S. 1983 Molecular genetics of the *Bithorax* complex in *Drosophila melanogaster*. Science **221**: 23-29.

13. Biemont C. 1986 Polymorphism of the Mdg-1 and I mobile elements in *Drosophila melanogaster*. Chromosoma **93**: 393-397.

14. Biessmann H. 1985 Molecular analysis of the *yellow* gene (*y*) region of *Drosophila melanogaster*. Proc. Nat. Acad. Sci. **82**: 7369-7373.

15. Bishop D.H.L., Claybrook J.R. and Spiegelman S. 1967 Electrophoretic separation of viral nucleic acids on polyacrylamide gels. J. Mol. Biol. **26**: 373-387.

16. Boeke J.D., Garfinkel D.J., Styles C.A. and Fink G.R. 1985 Ty elements transpose through an RNA intermediate. Cell **40**: 491-500.

17. Bossy B., Hall L.M.C. and Spierer P. 1984 Genetic activity along 315 kb of the *Drosophila* chromosome. EMBO **3**: 2537-2541.

18. Boyer H.W. 1971 DNA restriction and modification mechanisms in bacteria. Ann. Rev. Micobiology **25**: 153-176.

19. Bregliano J.C., Picard G., Bucheton A., Pelison A., Lavige J.M. and L'Heritier P. 1980 Hybrid dysgenesis in *Drosophila melanogaster*. Science **207**: 606-611.

20. Bridges C. 1935 Supplement to J. Heredity **26**.

21. Britten R.J. and Davidson E.H. 1969 Gene regulation for higher cells: a theory. Science **165**: 349-357.

22. Britten R.J. and Kohne D.E. 1968 Repeated sequences in DNA. Science **161**: 529–540.

23. Brookfield J.F.Y. 1982 Interspersed, repetitive DNA sequences are unlikely to be parasitic. J. Theoret. Biol. **94**: 281–299.

24. Brookfield J.F.Y. 1986 A model for DNA sequence evolution within transposable element families. Genetics **112**: 393–407.

25. Brookfield J.F.Y., Montgomery E. and Langley C.H. 1984 Apparent absence of transposable elements related to the P elements of *Drosophila melanogaster* in other species of *Drosophila*. Nature **310**: 330–331.

26. Brutlag D.L. 1980 Molecular arrangement and evolution of heterochromatic DNA. Ann. Rev. Genet. **14**: 121–144.

27. Bucheton A., Paro R., Sang H.M., Pelisson A. and Finnegan D.J. 1984 The molecular basis of I–R hybrid dysgenisis in *Drosophila melanogaster*. Identification, cloning and properties of the I factor. Cell **38**: 153–163.

28. Bucheton A., Simonelig M., Vaury C. and Crozatier M. 1986 Sequences similar to the I transposable element involved in I–R hybrid dysgenesis in *Drosophila melanogaster* occur in other *Drosophila* species. Nature **322**: 650–652.

29. Bulmer M.G. 1971 The effect of selection on genetic variability. Am. Nat. **105**: 201–211.

30. Cairns J. 1963 The chromosome of *Escherichia coli* Cold Spring Harb. Symp. Quant. Biol. **28**:. 43–46.

31. Cameron J.R., Loh E.Y. and Davis R.W. 1979 Evidence for transposition of dispersed, repetitive DNA families in yeast. Cell **16**: 739–751.

32. Campuzano S., Carramolino L., Cabrera C.V., Ruiz-Gomez M., Villares R., Boronat A. and Modellel J. 1985 Molecular genetics of the *achaete-scute* gene complex of *Drosophila melanogaster*. Cell **40**: 327–338.

33. Campuzano S., Balcells L., Villares R., Carramolino L., Garcia-Alonso L. and Modellel J. 1986 Excess function *hairy-wing* mutants caused by *gypsy* and *copia* insertions within structural genes of the achaete-scute locus of *Drosophila*. Cell **44**: 303–312.

34. Carramolino L., Ruiz-Gomez M., Guerrero M.C., Campuzano S. and Modellel J. 1982 DNA map of mutations at the *scute* locus of *Drosophila melanogaster*. EMBO **1**: 1185–1191.

35. Charlesworth B. and Charlersworth D. 1983 The population dynamics of transposable elements. Genet. Res. **42**: 1-27.

36. Charlesworth B. and Langley C.H. 1986 The evolution of self-regulated transposition of transposable elements. Genetics **112**: 359-383.

37. Charlesworth B., Langley C.H. and Stephan W. 1986 The evolution of restricted recombination and the accumulation of repeated DNA sequences. Genetics **112**: 947-962.

38. Collins M. and Rubin G.M. 1982 Structure of the *Drosophila* mutable allele, *white-crimson* and its *white-ivory* and wild-type derivatives. Cell **30**: 71-79.

39. Collins M. and Rubin G.M. 1983 High frequency precise excision of the *Drosophila* foldback transposable element. Nature **303**: 259-260.

40. Cooper D.W., Johnston P.G., Vandeberg J.L., Maynes G.M. and Chew G.K. 1979 A comparison of genetic variability at X-liked and autosomal loci in kangaroos, man and *Drosophila*. Genet. Res **33**: 243-252.

41. Cooper K.W. 1964 Meiotic conjunctive elements not involving chiasmata. Proc. Nat. Acad. Sci. **52**: 1248-1255.

42. Coyne J.A. 1983 Genetic basis of differences in genital morphology among three sibling species of *Drosophila*. Evolution. **37**: 1101-1118.

43. Coyne J.A. 1984 Genetic basis of male sterility in hybrids between two closely related species of *Drosophila*. Proc. Nat. Acad. Sci.. **81**: 4444-4447.

44. Coyne J.A. 1985 Genetic studies of three sibling species of *Drosophila* with relationship to theories of speciation. Genet. Res. **46**: 169-192.

45. Coyne J.A. and Kreitman M. 1986 EvolutioNucl. Acid Res.y genetics of two sibling species, *Drosophila simulans* and *Drosophila sechellia*. Evolution. **40**: 673-691.

46. Craigie R. and Mizuuchi K. 1985 Mechanism of transposition of bacteriophage mu: structure of a transposition intermediate. Cell **41**: 867-876.

47. Craigie R. and Mizuuchi K. 1986 Role of DNA topology in Mu transposition: mechanism of sensing the relative orientation of two DNA segments. Cell **45**: 793-800.

48. Crow J.F. and Kimura M. 1970 An introduction to population genetics theory. New York Harper and Row, publishers.

49. Crozier R.H. 1976 Counter-intuitive property of effective population size. Nature **262**: 384.

50. Davidson E.H. and Hough B.R. 1971 Genetic information in oocyte RNA. J. Mol. Biol. **56**: 491-506.

51. Davidson E.H., Hough B.R., Amenson C.S. and Britten R.J. 1973 General interspersion of repetitive with non-repetitive sequence elements in the DNA of *Xenopus*. J. Mol. Biol. **77**: 1-23.

52. Dawid I.B., Long E.O. and DiNocera P.P. 1981 Ribosomal insertion-like elements in *Drosophila melanogaster* are interspersed with mobile sequences. Cell **25**: 399-408.

53. Demerec M. and Hoover M.E. 1939 *Hairy-wing* - a duplication in *Drosophila melanogaster*. Genetics **24**: 271-277.

54. Di Nocera P.P. and Dawid I.B. 1983 Interdigitated arrangement of two oligo-(A) terminated DNA sequences in *Drosophila* Nucl. Acid Res. **11**: 5475-5482.

55. Di Nocera P.P., Graziani F. and Lavorgna G. 1986 Genomic and structural organization of *Drosophila melanogaster* G elements. Nucl. Acid Res. **14**: 675- 691.

56. Di Pasquale 1951 Report of istituto di genetica, University of Milan. Dros. Inf. Serv. **25**: 70.

57. Dobzhansky Th. 1963 Genetics of natural populations. XXXIII. A progress report on genetic changes in populations of *Drosophila pseudoobscura* and *Drosophila persimilis* in a locality in California. Evolution. **17**: 333-339.

58. Dobzhansky Th. and Spassky B. 1954 Genetics of natural populations. XXII. A comparison of the concealed variability in *Drosophila prosaltans* with that in other species. Genetics **39**: 472-487.

59. Doolittle F.W. and Sapienza C. 1980 Selfish genes, the phenotype paradigm and genome evolution. Nature **284**: 601-603.

60. Dubinin N.P. 1933 Step-allelomorphism in *Drosophila melanogaster*. J. Genet. **27**: 443-464.

61. Dubinin N.P., Sokolov N.N. and Tiniakov G.G. 1937 Crossing over between the genes *yellow, achaete* and *scute*. Dros. Inf. Serv. **8**: 76.

62. Eanes W.F., Hey J. and Houle D. 1985 Homozygous and hemizygous viability variation on the X chromosome of *Drosophila melanogaster*. Genetics **111**: 831-844.

63. Economou-Pachnis A., Lohse M.A., Furano A.V. and Tsichlis P.N. 1985 Insertion of long interspersed repeated elements in the *Igh* (immunoglobulin heavy chain) and *Mlvi-2* (Molonyleukaemia virus integration 2) loci of rats. Proc. Nat. Acad. Sci. **82**: 2857-2861.

64. Engels W.R. 1979 Extrachromosomal control of mutability in *Drosophila melanogaster*. Proc. Nat. Acad. Sci. **76**: 4011-4015.

65. Engels W.R. 1983 The P family of transposable elements in *Drosophila*. Ann. Rev. Genet. 1983 **17**: 315-344.

66. Engels W.R. 1984 A *trans* acting product needed for P factor transposition in *Drosophila*. Science **226**: 1194-1196.

67. Engels W.R. and Preston C.R. 1979 Hybrid dysgenesis in *Drosophila melanogaster*: the biology of female and male sterility. Genetics **92**: 161-174.

68. Engels W.R. and Preston C.R. 1981 Identifying P factors in *Drosophila* by means of chromosome breakage hotspots. Cell **26**: 421-428.

69. Ewens W.J., Spielman R.S. and Harris H. 1981 Estimation of genetic variation at the DNA level from restriction endonuclease data. Proc. Nat. Acad. Sci. **78**: 3748-3750.

70. Falk R. 1963 A search for a gene control system in *Drosophila*. Am. Nat. **97**: 129-132.

71. Fawcett D.H., Lister C.K., Kellett E. and Finnegan D.J. 1986 Transposable elements controlling I-R hybrid dysgenesis in *Drosophila melanogaster* are similar to mammalian LINE's. Cell **47**: 1007-1015.

72. Finnegan D.J., Rubin G.M., Young M.W. and Hogness D.S. 1978 Cold Spring Harb. Symp. Quant. Biol. **42**: 1053-1063.

73. Fitzpatrick B.J. and Sved J.A. 1986 High levels of fitness modifiers induced by hybrid dysgenesis in *Drosophila melanogaster*. Genet. Res. **48**: 89-94.

74. Feinberg A.P. and Vogelstein B. 1984 A technique for radiolabelling DNA restriction endonuclease fragments to a high specific activity. Anal. Biochem. **132**: 6-13.

75. Flavell A.J. and Ish-Horowicz D. 1983 The origin of extrachromosomal circular *copia* elements. Cell **34**: 415-419.

76. Flavell A.J. 1984 Role of reverse transcription in the generation of extrachromosomal *copia* mobile genetic

elements. Nature **310**: 514-516.

77. Foster T.J., Davis M.A., Roberts D.E., Takeshita K. and Kleckner N. 1981 Genetic organization of the transposon Tn10. Cell **23**: 201-213.

78. Freund R. and Meselson M 1984 Long terminal repeat nucleotide sequence and specific insertion of the *gypsy* transposon. Proc. Nat. Acad. Sci. **81**: 4462-4464.

79. Frischauf A-M., Lehrach H., Poustka A. and Murray N. 1983 Lambda replacement vectors carrying polylinker sequences. J. Mol. Biol. **170**: 842-872.

80. Fry K. and Salser W. 1977 Nucleotide sequences in HS-alpha satellite DNA from kangaroo rat (*Dipodomys ordii*) and characterization of similar sequences in other rodents. Cell **12**: 1069-1084.

81. Gall J.G. and Atherton D.D. 1974 Satellite DNA sequences in *Drosophila virilis*. J. Mol. Biol. **85**: 633-664.

82. Garcia-Bellido A. 1979 Genetic analysis of the achaete-scute system of *Drosophila melanogaster* Genetics **91**: 491-520.

83. Gerasimova T.I., Mizrokhi L.J. and Georgiev G.P. 1984 Transposition bursts in genetically unstable *Drosophila melanogaster*. Nature **309**: 714-716.

84. Gerhenson S. 1933 Studies in the genetically inert region of the X chromosome of *Drosophila* J. Genet. **28**: 297-312.

85. Goldberg M.L., Paro R. and Gehring W.J. 1982 Molecular cloning of the *white* locus region of *Drosophila melanogaster* using a large transposable element. EMBO **1**: 93-98.

86. Gosden J.R., Mitchell A.R., Seuanez H.N. and Gosden C.M. 1977 The distribution of sequences complementary to human satellite DNAs I, II and IV in the chromosomes of chimpanzee (*Pan troglodytes*), gorilla (*Gorilla gorilla*) and orangutan (*Pongo Pygmaeus*). Chromosoma. **63**: 253-271.

87. Gough E.J. and Gough N.M. 1984 Direct calculation of the sizes of DNA fragments separated by gel electrophoresis using a programme written for a pocket calculator. Nucl. Acid Res. **12**: 845-853.

88. Green M.M. 1967 The genetics of a mutable gene at the *white* locus of *Drosophila melanogaster*. Genetics **56**: 467-482.

89. Grindley N.D.F. and Reed R.R 1985 Transpositional

recombination in prokaryotes. Ann. Rev. Biochem. **54**: 863-869.

90. Gronenborn B. and Messing J. 1978 Methylation of a single stranded DNA *in vitro* introduces new restriction endonuclease cleavage sites. Nature **272**: 375-377.

91. Hanahan D. 1983 Studies on transformation of *Escherichia coli* with plasmids. J. Mol. Biol. **166**: 557-580.

92. Harris H. 1966 Enzyme polymorphisms in man. Proc. Roy. Soc. **164**: 298-310.

93. Haynes S.R., Toomey T.P., Leinwand L. and Jelink W.R. 1981 The chinese hamster *Alu* equivalent sequence: a conserved, highly repetitious, interspersed deoxyribonucleic acid sequence in mammals has a structure suggestive of a transposable element. Mol. Cell Biol. **1**: 573-583.

94. Hedgepeth J., Goodman H.M. and Boyer H.W. 1972 DNA nucleotide sequence restricted by the RI endonuclease. Proc. Nat. Acad. Sci. **69**: 3448-3452.

95. Helling R.B., Goodman H.M. and Boyer H.W. 1974 Analysis of *Eco* RI restriction fragments of DNA from lambdoid bacteriophage and other viruses by agarose gel electrophoresis. J. Virol. **14**: 1235-1241.

96. Hennig W., Hennig I. and Stein H. 1970 Repeated sequences in the DNA of *Drosophila* and their location in giant chromosomes. Chromosoma. **32**: 31-63.

97. Hill W.G. 1974 Estimation of linkage disequilibrium in randomly mating populations. Heredity **33**: 229-239.

98. Hill W.G. and Robertson A. 1968 Linkage disequilibrium in finite populations. TAG **38**: 226-231.

99. Hobbs H.H., Lehrman M.A., Yamamoto T. and Russel D.W. 1985 Polymorphism and evolution of *Alu* sequences in the human low density lipoprotein receptor gene. Proc. Nat. Acad. Sci. **82**: 7651-7655.

100. Holmes D.S. and Quigley M. 1981 A rapid boiling method for the preparation of bacterial plasmids. Anal. Biochem. **114**: 193-198.

101. Hudson R.R. 1982 Estimating genetic variation with restriction endonucleases. Genetics **100**: 711-719.

102. Ish-Horowicz D. and Pinchin S.M. 1980 Genomic organization of the 87A7 and 87C1 heat-inducible loci of *Drosophila melanogaster*. J. Mol. Biol. **142**: 231-245.

103. Ilyin Y.V., Tchurikov N.A., Ananiev E., Ryskov A.P., Yenikopolov G.N., Limbourska S.A., Maleeva N.E., Gvozdev V.A. and Georgiev G.P. 1978 Studies on the DNA fragments of mammals and *Drosophila* containing structural genes and adjacent sequences. Cold Spring Harb. Symp. Quant. Biol. Biol. **42**: 959-969.

104. Ish-Horowicz D., Pinchin S.M., Gausz J., Gyurkovics H., Bencze G., Goldshmidt-Clermont M. and Holden J. 1979 Deletion mapping of two *Drosophila melanogaster* loci that code for the 70,000 d heat inducible protein. Cell **17**: 565-571.

105. Jagadeeswaran P., Forget B.G. and Weissman S.M. 1981 Short interspersed repetitive DNA elements in Eukaryotes: transposable DNA elements generated by reverse transcription of RNA pol III transcripts? Cell **26**: 141-142.

106. Jeffreys A.J. 1979 DNA sequence variants in the $^{G}\gamma$-, $^{A}\gamma$-, δ- and β-globin genes of man. Cell **18**: 1-10.

107. Johnson D.A., Gautsch J.W., Sportsman R.J. and Elder J.H. 1984 Improved technique utilising non-fat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose. Gene. Anal. Tech. **1**: 3-8.

108. Kacser H. and Burns J.A. 1981 The molecular basis of dominance. Genetics **97**: 639-666.

109. Kaplan N. and Brookfield J.F.Y. 1983 Transposable elements in mendelian populations III. Statistical results. Genetics **104**: 485-495.

110. Kaplan N., Darden T. and Langley C.H. 1985 Evolution and extinction of transposable elements in mendelian populations. Genetics **109**: 459-480.

111. Karn J., Brenner S., Barnett L. and Cesareni G. 1980 Novel bacteriophage λ cloning vector. Proc. Nat. Acad. Sci. **77**: 5172-5176.

112. Keightly P.D. and Hill W.G. 1987 Manuscript in preparation.

113. Kelly T.J. and Smith H.O. 1970 A restriction enzyme from *Haemophilus influenzae* II. Base sequence of the recognition site. J. Mol. Biol. **51**: 393- 409.

114. Keyl H.G. 1965 Duplikationen von untereinheiten der chromosomalen DNS wahrend der evolution von *Chironomus thumini*. Chromosoma **17**: 139-180.

115. Kidwell M.G. 1979 Hybrid dysgenesis in *Drosophila melanogaster*: the relationship between the P-M and I-R interaction systems. Genet. Res. **33**: 205-217.

116. Kidwell M.G. 1982 Intraspecific hybrid sterility in the genetics and biology of *Drosophila*. Ed. Ashburner M., Carson H.L. and Thompson J.N. Vol. 3C. Academic press, London and New York.

117. Kidwell M.G. 1983 Evolution of hybrid dysgenesis determinants in *Drosophila melanogaster*. Proc. Nat. Acad. Sci. **80**: 1655-1659.

118. Kidwell M.G., Kidwell J.F. and Sved J.A. 1977 Hybrid dysgenesis in *Drosophila melanogaster*: a syndrome of aberrant traits including mutation, sterility and male recombination. Genetics **86**: 813-833.

119. Kidwell M.G., Novy J.B. and Feely S.M., 1981 Rapid unidirectional change of hybrid dysgenesis potential in *Drosophila*. Journal of Heredity **72**: 32-38.

120. Kimura M. 1968a EvolutioNucl. Acid Res.y rate at the molecular level. Nature **217**: 624- 626.

121. Kimura M. 1968b Genetic variability maintained in a finite population due to mutational production of neutral and nearly neutral isoalleles. Genet. Res. **11**: 247-269.

122. Kimura M. 1979 Model of effectively neutral mutations in which selective constraint is incorporated. Proc. Nat. Acad. Sci. **76**: 3440-3444.

123. Kimura M. 1983 The neutral theory of molecular evolution. Cambridge University Press.

124. Kimura M. and Ohta T. 1971 Protein polymorphism as a phase of molecular evolution. Nature **229**: 467-469.

125. King J.L. and Jukes T.H. 1969 Non-darwinian evolution. Science **164**: 788-800.

126. Kleckner N. 1981 Transposable elements in prokaryotes. Ann. Rev. Genet. **15**: 341-404.

127. Kleckner N., Steele D.A., Reichard K. and Botstein D. 1979 Specificity of insertion of the translocatable tetracycline-resistance element Tn10. Genetics **92**: 1023-1040.

128. Kreitman M. 1983 Nucleotide polymorphism at the alcohol dehydrogenase locus of *Drosophila melanogaster* Nature **304**: 412-417.

129. Lande R. 1976 The maintenance of genetic variability by mutation in a polygenic character with linked loci. Genet. Res. **26**: 221-235.

130. Langley C.H., Brookfield J.F.Y. and Kaplan N. 1983 Transposable elements in mendelian populations I: A theory. Genetics **104**: 457-471.

131. Langley C.H., Montgomery E. and Quattlebaum W.E. 1982 Restriction map variation in the *Adh* region of *Drosophila* Proc. Nat. Acad. Sci. **79**: 5631-5635.

132. Langley C.H., Voelker R.A., Leigh Brown A.J., Ohnishi S. Dickson B. and Montgomery E. 1981 Null allele frequencies at allozyme loci in natural populations of *Drosophila melanogaster*. Genetics **99**: 151-156.

133. Lansman R.A., Stacey S.N., Grigliatti T.A. and Brock H.H. 1985 Sequences homologous to the P mobile element of *Drosophila melanogaster* are widely distributed in the subgenus *sophophora*. Nature **318**: 561-563.

134. Larsen A. and Weintraub H. 1982 An altered DNA conformation detected by S1 nuclease occurs at specific regions in active chick globin chromatin. Cell **29**: 609-622.

135. Laski F.A., Rio D.C. and Rubin G.M. 1986 Tissue specificity of *Drosophila* P element transposition is regulated at the level of mRNA splicing. Cell **44**: 7-19.

136. Leigh Brown A.J. 1983 Variation at the 87A heat shock locus in *Drosophila melanogaster* Proc. Nat. Acad. Sci. **80**: 5350-5354.

137. Leigh Brown A.J. 1984 On the origin of the *alu* family of repeated sequences. Nature **312**: 106.

138. Leigh Brown A.J. and Ish-Horowicz D. 1981 Evolution of the 87A and 87C heat shock loci in *Drosophila*. Nature **290**: 677-682.

139. Leigh Brown A.J. and Moss J.E. 1987 Transposition if the I element and *copia* in a natural population of *Drosophila melanogaster* Genet. Res. In press.

140. Levis R., Collins M. and Rubin G.M. 1982 FB elements are the common basis for the instability of the w^{DZL} and w^{c} *Drosophila* mutations. Cell **30**: 551-565.

141. Levis R., O'Hare K. and Rubin G.M. 1984 Effects of transposable element insertions on RNA encoded by the *white* gene of *Drosophila*. Cell **38**: 471-481.

142. Lewontin R.C. 1974 The genetic basis of evolutioNuc'. Acid Res.y change. Columbia University press.

143. Lewontin R.C. and Hubby J.L. 1966 A molecular approach to the study of genetic heterozygosity in natural populations

II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*. Genetics **54**: 595-609.

144. Lifton R.P., Goldberg M.L., Karp R.W. and Hogness D.S. 1978 The organization of the histone genes in *Drosophila melanogaster*: Functional and evolutioNucl. Acid Res.y implications. Cold Spring Harb. Symp. Quant. Biol. **42**: 1047-1051.

145. Lindsley D.L. and Grell E.H. 1967 Genetic variations of *Drosophila melanogaster*. Carnegie institution of Washington publication 627.

146. Mackay T.F.C. 1985 Transposable element induced response to artificial selection in *Drosophila melanogaster*. Genetics **111**: 351-374.

147. Mackay T.F.C. 1986a Transposable element induced fitness mutations in *Drosophila melanogaster*. Genet. Res. **48**: 77-87.

148. Mackay T.F.C. 1986b A quantitative genetic analysis of fitness and its components in *Drosophila melanogaster*. Genet. Res. **47**: 59-70.

149. Maeda N., Bliska J.B. and Smithies O. 1983 Recombination and balanced chromosome polymorphism suggested by sequences 5' to the human δ-globin gene. Proc. Nat. Acad. Sci. **80**: 5012-5016.

150. Maniatis T., Fritsch E.F. and Sambrook J. 1982 Molecular cloning: A laboratory manual. Cold Spring Harbour Laboratory. Cold Spring Harbour. New York.

151. Manning J.E., Schmidt C.W. and Davidson N. 1975 Interspersion of repetitive and non-repetitive DNA sequences in the *Drosophila melanogaster* genome. Cell **4**: 141-155.

152. Mason P.S., Torok I., Kiss I., Karch F. and Udvardy A. 1982 A complex pattern of DNA sequence homology extending far upstream of the hsp 70 genes at loci 87A7 and 87C1 in *Drosophila melanogaster*: evolutioNucl. Acid Res.y implications. J. Mol. Biol. **156**: 21-35.

153. McDonnel M.W., Simon M.N. and Studier F.W. 1977 Analysis of restriction fragments of T7 DNA and determination of molecular weight by electrophoresis in neutral and alkaline gels. J. Mol. Biol. **110**: 119-146.

154. Messing J., Crea R. and Seeburg P.H. 1981 A system for shotgun DNA sequencing. Nucl. Acid Res. **9**: 309-321.

155. Miklos G.L.G. and Nankivell R.N. 1976 Telomeric satellite DNA functions in regulating recombination. Chromosoma. **56:** 143-167.

156. Modellel J., Bender W. and Meselson M. 1983 *Drosophila melanogaster* mutations suppressible by the suppressor of *hairy-wing* are insertions of a 7.3-kilobase mobile element. Proc. Nat. Acad. Sci. **80:** 1678-1682.

157. Montgomery E., Charlesworth B. and Langley C.H. 1987 Manuscript in press.

158. Montgomery E. and Langley C.H. 1983 Transposable elements in mendelian populations II: Distribution of three *copia*-like elements in a natural population of *Drosophila melanogaster*. Genetics **104:** 473- 483.

159. Morgan 1910 Sex limited inheritance of *Drosophila* Science **32:** 120-122.

160. Morton N.E. 1971 population genetics and disease control. Social Biology **18:** 243-251.

161. Moscoso del Prado J. and Garcia-Bellido A. 1984a Genetic regulation of the *achaete-scute* complex of *Drosophila melanogaster*. Roux's Arch. Dev. Biol. **193:** 242-245.

162. Moscoso del Prado J. and Garcia-Bellido A. 1984b Cell interactions in the generation of chaetae pattern in *Drosophila*. Roux's Arch. Dev. Biol. **193:** 242-245.

163. Mount S.M. and Rubin G.M. 1985 Complete nucleotide sequence of the *Drosophila* transposable element *copia*: homology between *copia* and retroviral proteins. Mol. Cell Biol. **5:** 1630-1638.

164. Mukai T. and Yamaguchi O. 1974 The genetic structure of natural populations of *Drosophila melanogaster* XI: genetic variability in a local population. Genetics **76:** 339-366.

165. Muller H.J. 1950 Our load of mutations Am. J. Hum. Genet. **2:** 111-176.

166. Musich P.R. and Dykes R.J. 1986 A long interspersed (LINE) DNA exhibiting polymorphic patterns in human genomes. Proc. Nat. Acad. Sci. **83:** 4854-4858.

167. O'Hare K. and Rubin G.M. 1983 Stuctures of p transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. Cell **34:** 25-35.

168. Ohta T. 1982 Allelic and non-allelic homology of a supergene family. Proc. Nat. Acad. Sci. **79:** 3251-3254.

169. Ohta T. 1983 Theoretical study on the accumulation of selfish DNA. Genet. Res. **41**: 1-15.

170. Ohta T. 1986 Population genetics of an expanding family of mobile genetic elements. Genetics **113**: 145-159.

171. Orgel L.E. and Crick F.H.C. 1980 Selfish DNA: the ultimate parasite. Nature **284**: 604-607.

172. Pardue M.L. and Gall J.G. 1970 Chromosomal localization of mouse satellite DNA. Science **168**: 1356-1358.

173. Parkhurst S.M. and Corces V.G. 1986 Interactions among the *Gypsy* transposable element and the *yellow* and suppressor of *hairy-wing* loci in *Drosophila melanogaster*. Mol. Cell Biol. **6**: 47-53.

174. Peacock A.C. and Dingman C.W. 1968 Molecular weight estimation and separation of RNA by electrophoresis in agarose-acrylamide composite gels. Biochem **7**: 668-674.

175. Peacock W.J., Brutlag D., Goldring E., Appels R., Hinton C.W. and Lindsley D.L. 1973 The organization of highly repeated DNA sequences in *Drosophila melanogaster* chromosomes. Cold Spring Harb. Symp. Quant. Biol. **38**: 405-416.

176. Peacock W.J., Lohe A.R., Dunsmuir P., Dennis E.S. and Appels 1977 Fine structure and evolution of DNA in heterochromatin. Cold Spring Harb. Symp. Quant. Biol. **48**: 1121-1135.

177. Pelison A. 1981 The I-R system of hybrid dysgenesis in *Drosophila melanogaster*. are I factor insertions responsible for the mutator effect of the I-R interaction. Mol. Gen. Genet. **183**: 123-129.

178. Pelison A. and Bregliano J.C. 1987 Mol. Gen. Genet. In press.

179. Picard G. 1976 Non-mendelian sterility in *Drosophila melanogaster*. Hereditary transmission of I factor. Genetics **83**: 107-123.

180. Pierce D.A. and Lucchesi J.C. 1981 Analysis of a dispersed repetitive DNA sequence in isogenic lines of *Drosophila*. Chromosoma **82**: 471-492.

181. Pirrotta V. and Brockl C. 1984 Transcription of the *Drosophila white* locus and some of its mutants. EMBO **3**: 563-568.

182. Potter S.S., Brorein W.J., Dunsmuir P. and Rubin G.M. 1979 Transposition of elements of 412, *copia* and 297 dispersed,

repeated gene families in *Drosophila*. Cell **17**: 415-427.

183. Potter S.S., Truett M., Phillips M. and Mather A. 1980 Eukaryotic tranposable genetic elements with inverted terminal repeats. Cell **20**: 639-647.

184. Rasch E.M., Barr H.J. and Rasch R.W. 1971 The DNA content of sperm of *Drosophila melanogaster*. Chromosoma **33**: 1-18.

185. Rigby P.J., Breckmann M., Rhodes C. and Berg P. 1977 Labelling deoxyribonucleic acid *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. **113**: 237-251.

186. Ritossa F.M. and Spiegelman S. 1965 Localization of DNA complementary to ribosomal RNA in the nucleolus organiser region of *Drosophila melanogaster*. Proc. Nat. Acad. Sci. **53**: 737-745.

187. Rio D.C., Laski F.A. and Rubin G.M. 1986 Identification and immunochemical analysis of biologically active *Drosophila* P element transposase. Cell **44**: 21-32.

188. Robinson M.K., Bennett P.M. and Richmond M.H. 1977 Inhibition of TnA translocation by TnA. J. Bact. **129**: 407-414.

189. Roeder S.G. and Fink G.R. 1982 Movement of yeast transposable elements by gene conversion. Proc. Nat. Acad. Sci. **79**: 5621-5625.

190. Roeder S.G. and Fink G.R. 1983 Transposable elements in yeast. In mobile genetic elements ed. Shapiro 1983. London and New York Academic press.

191. Ronsseray S. and Anxolabehere D. 1986 Chromosomal distribution of P and I transposable elements in a natural population of *Drosophila melanogaster*. Chromosoma **94**: 433-440.

192. Rothfels K., Sexsmith E., Heimburger M. and Krause M.O. 1966 Chromosome size and DNA content of species of Anemone L. and related genera (ranunculaceae). Chromosoma **20**: 54-74.

193. Rowe T.C., Wang J.C. and Liu L.F. 1986 *in vivo* localization of DNA topoisomerase II cleavage sites in *Drosophila* heat shock chromatin. Mol. Cell Biol. **6**: 985-992.

194. Rubenstein I., Thomas C.A. and Hershey A.D. 1961 The molecular weights of T2 bacteriophage DNA and its first and second breakage products. Proc. Nat. Acad. Sci. **47**: 1113-1122.

195. Rubin G.M., Brorein W.J., Dunsmuir P., Flavell A.J., Levis R., Strobel E., Toole J.J. and Young E. 1980 *copia* like transposable elements in the *Drosophila melanogaster* genome. Cold Spring Harb. Symp. Quant. Biol. **45**: 619-628.

196. Rubin G.M., Kidwell M.G. and Bingham P.M. 1982 The molecular basis of P-M hybrid dysgenesis: the nature of induced mutations. Cell **29**: 987-994.

197. Rubin G.M. and Spradling A.C. 1982 Genetic transformation of *Drosophila* with transposable elements vectors. Science **218**: 348-353.

198. Ru Hwu H., Roberts J.W., Davidson E.H. and Britten R.J. 1986 Insertion and/or deletion of many repeated DNA sequences in human and higher ape evolution. Proc. Nat. Acad. Sci. **83**: 3875-3879.

199. Schafit-Zagardo B., Maco J.J. and Brown F.L. 1982 *Kpn* I families of long, interspersed repetitive DNAs in human and other primate genomes. Nucl. Acid Res. **10**: 3175-3193.

200. Schmid C.W. and Jelink W.R. 1982 The *Alu* family of dispersed repetitive sequences. Science **216**: 1065-1070.

201. Shapiro J.A. 1979 Molecular model for the transposition and replication of bacteriophage Mu and other transposable elements. Proc. Nat. Acad. Sci. **76**: 1933-1937.

202. Shapiro J.A. (ed) 1983 Mobile genetic elements. London and New York Academic press.

203. Sharp P.A. 1983 Conversion of RNA to DNA in mammals: *Alu*-like elements and pseudogenes. Nature **301**: 471-472.

204. Shiba T. and Saigo K. 1983 Retrovirus like particles containing RNA homologous to the transposable elements *copia* in *Drosophila melanogaster*. Nature **302**: 119-134.

205. Shrimpton A.E. and Robertson A. 1987a The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. I Allocation of third chromosome sternopleural bristle effects to chromosome sections. Genetics submitted.

206. Shrimpton A.E. and Robertson A. 1987b The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. II Distribution of third chromosome bristle effects within chromosome sections. Genetics submitted..

207. Simmons M.J. and Lim J.K. 1980 Site specificity of mutations arising in dysgenic hybrids of *Drosophila melanogaster*. Proc. Nat. Acad. Sci. **77**: 6042-6046.

208. Singer M.F. 1982 SINE's and LINE's: highly repeated short and long interspersed sequences in mammalian genomes. Cell **28:** 433-434.

209. Slatkin M. 1985 Genetic differentiation of transposable elements under mutation and unbiased gene conversion. Genetics **110:** 145-158.

210. Sokal R.R. and Rohlf J.F. 1969 Biometry. The princples and practice of statistics in biological research. W.H. Freeman. San Francisco.

211. Sokal R.R. and Rohlf J.F. 1969 Statistical tables. W.H. Freeman. San Francisco.

212. Southern E.M. 1970 Base sequence and evolution of guinea pig α-satellite. Nature **227**: 794-798.

213. Southern E.M. 1975 Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98:** 503-517.

214. Southern E.M. 1979 Measurement of DNA length by gel electrophoresis. Anal. Biochem **100:** 319-323.

215. Spierer P. and Spierer A. 1983 Molecular mapping of genetic and chromomeric units in *Drosophila melanogaster.* J. Mol. Biol. **168:** 35-50.

216. Spradling A.C. and Rubin G.M. 1981 *Drosophila* genome organization: conserved and dynamic aspects. Ann. Rev. Genet. **15:** 219-264.

217. Stalder J., Larsen A., Engel J.D., Dolan M., Groudine M. and Weintraub H. 1980 Tissue specific DNA cleavages in the globin chromatin domain introduced by DNase I. Cell **20:** 452-460.

218. Stephan W. 1986 Recombination and the evolution of satellite DNA. Genet. Res. **47**: 167-174.

219. Strobel E., Dunsmuir P. and Rubin G.M. 1979 Polymorphisms in the chromosomal locations of the 412, *copia* and 297 dispersed, repeated gene families in *Drosophila.* Cell **17:** 429-439.

220. Thirion J.P. and Hofnung M. 1972 On some genetic aspects of phage λ resistance in *E.coli* K12 Genetics **71:** 207-216.

221. Tracey M.L. and Ayala F.J. 1974 Genetic load in natural populations: is it compatible with the hypothesis that many polymorphisms are maintained by natural selection? Genetics **77**: 569-589.

222. Udvardy A., Schedl P., Saunder M. and Hsieh T-S. 1985 Novel partitioning of DNA cleavage sites for *Drosophila* topoisomerase II. Cell **40:** 933- 941.

223. Ullmann A., Jacob F. and Monod J. 1967 Characterization by *in vitro* complimentation of a peptide corresponding to an operator-proximal segment of the β-galactosidase structural gene of *Escherichia coli.* J. Mol. Biol. **24:** 339-343.

224. Ullu E. and Tschudi C. 1984 *Alu* sequences are processed 7SLRNA genes. Nature. **312:** 171-172.

225. Vieira J. and Messing J. 1982 The pUC plasmids, an M13mp7 - derived system for insertion mutagenesis and sequencing with universal primers. Gene **19:** 259-268.

226. Voelker R.A., Greenleaf A.L., Gyurkovics H., Wisley G.B., Huang S-M. and Searles L.L. 1984 Frequent imprecise excision among reversions of a P element caused lethal mutation in *Drosophila.* Genetics **107:** 279- 294.

227. Voelker R.A., Langley C.H., Leigh Brown A.J., Ohnishi S., Dickson B., Montgomery E. and Smith S.C. 1980 Enzyme null alleles in a natural population of *Drosophila melanogaster.* frequencies in a North Carolina population. Proc. Nat. Acad. Sci. **77:** 1091-1095.

228. Wahl G.M., Stern M. and Stark G.R. 1979 Efficient transfer of large DNA fragments from agarose gels to diazobenzyloxymethyl paper and rapid hybridization using dextran sulphate. Proc. Nat. Acad. Sci. **76:** 3683-3687.

229. Walter D. and Blobel G. 1982 Signal recognition particle contains a 7SRNA essential for protein translocation across the endoplasmic reticulum. Nature **299:** 691-698.

230. Waring M. and Britten R.J. 1966 Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. Science **154:** 791-794.

231. Weintraub H. and Groudine M. 1976 Chromosomal subunits in active genes have an altered conformation. Science **193:** 848-856.

232. Wensink P.C., Finnegan D.J., Donelson J.E. and Hogness D.S. 1974 A system for mapping DNA sequences in the chromosomes of *Drosophila melanogaster.* Cell **3:** 315-325.

233. Wetmuir J.G. and Davidson N. 1968 Kinetics of renaturation of DNA. J. Mol. Biol. **31:** 349-370.

234. White J.D. 1978 Modes of speciation. W.H. Freeman and Co.

235. Will B.M., Bayev A.A. and Finnegan D.J. 1981 Nucleotide sequence of terminal repeats of 412 transposable elements of *Drosophila melanogaster*. J. Mol. Biol. **153**: 897-915.

236. Wu C. 1980 The 5' ends of *Drosophila* heat shock genes in chromatin are hypersensitive to DNAse I. Nature **286**: 854-860.

237. Young M.W. 1979 Middle repetitive DNA: a fluid component of the *Drosophila* genome. Proc. Nat. Acad. Sci. **76**: 6274-6278.

238. Yukuhiro K., Harada K. and Nukai T. 1985 Viability mutations induced by the p elements in *Drosophila melanogaster*. Jpn. J. Genet. **60**: 531-537.

239. Yunis J.J. and Yasmineh W.G. 1970 Satellite DNA in constitutive heterochromatin of the guinea pig. Science **168**: 263-265.

240. Zhimulev I.F., Pokholkova G.V., Bgatov A.V., Semeshin V.F., Belyaeva E.S. 1981 Fine cytogenetical analysis of the band 10A1-2 and the adjoining regions in the *Drosophila melanogaster* chromosome II. Genetic analysis. Chromosoma. **82**: 25-40.

241. Zissler J., Signer E. and Schaefer F. 1971 The role of recombination in growth of bacteriophage Lambda II: Inhibition of growth by prophage P2. The bacteriophage Lambda. Cold Spring Harbour monograph series no.2. Cold Spring Harbour, New York.