Molecular analysis of transposable element insertions at the alcohol dehydrogenase locus in *Drosophila melanogaster*

A thesis submitted for the degree

of

Doctor of Philosophy

of

The Australian National University

by

Yan-Hong Wu

Molecular and Population Genetics Group

Research School of Biological Sciences

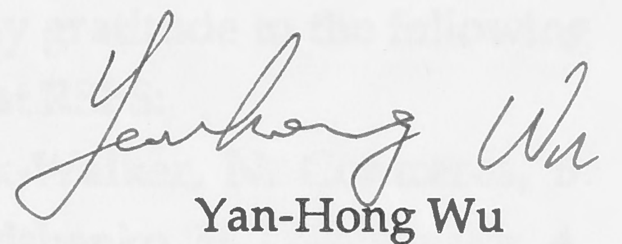
The Australian National University

Canberra

December, 1995

DECLARATION

In accordance with the regulations of the Australian National University, I wish to state that the work described and the results presented in this thesis were carried out by myself under the supervision of Prof. John B. Gibson. Where the work of others is quoted, appropriate references are given.



Yan-Hong Wu

ACKNOWLEDGMENTS

First of all, I would like to express my sincere thanks to my supervisor, Professor John B. Gibson, for his precious guidance and enthusiastically sharing his profound knowledge through the past four years. Without his continuous support and patient help, I were not able to smoothly complete this course of study.

I am also grateful to my advisor, Dr. Xin-Jie Chen, for his helpful discussion and assistance throughout my Ph.D study, and making valuable comments on preparation of this thesis. I thank Dr. Anna Ladavishenko, another advisor, for sharing her knowledge so freely and provided a few stocks used in this study.

I am especially grateful to Ann V. Wilks for her excellent technical instruction and assistance, as well as gifts of several probe DNA needed for this research. I am indebted to Dr. Darryl Reed, Dr. Tom Wilanowski and Slawek Bartoszewski, for their helpful discussion, sharing wonderful ideas and gifts of probe DNA and primers.

I'd like to take this opportunity to express my gratitude to the following colleagues and friends who have enriched my stay at RSBS:
S. Bartoszewski, H-C. Chen, X-J. Chen, D. Clark-Walker, N. Contreras, B. Creaser, C-M. Dong, J. Gibson, M-X. Guan, A. Ladvishenko, H. Liszczynsky, A. Marchant, L-J. Ouyan, D. Reed, D. Shaw, C. Stewart-Moore, X-W. Wang, T. Wilanowski, P. Wilkinson, A. Wilks, E. Wimmer, L-P. Wu, G-J. Xu, J-W. Yu, X-D. Zhang, J-J. Zhou.

The help and advice from the following people has been invaluable during my study and in preparation of this thesis: Jeff Wilson and Maureen Whittaker from the RSBS photography unit; Tara Goodsell and James Whitehead from the RSBS illustration unit; David Sandilands and Philip Drury from RSBS computer unit.

I would like to acknowledge the financial support of the Graduate School and the Research School of Biological Sciences at the Australian National University for the scholarships.

Finally, I'd like to thank my parents and brother for their never-ending support, encouragement and love over the years. Those friends who I can share all the best and worst allowed me to enjoy staying in Australia. A huge thanks must go to my fiancé, Dr. Ying Chen, thanks for his loving, understanding and supporting, as well as criticisms. Without him, I don't think I were able to complete this thesis.

ABSTRACT

Molecular and biochemical analyses were done on three low activity ADH variants, *AHA51, AD369* and *AAS44,* which were all isolated from natural populations. These variants have a similar abnormal developmental pattern of ADH activity, with low levels of enzyme in adult flies but near normal levels in 3rd instar larvae.

ADH activity assays showed that the allele *AHA51* had about 16%, *AD369* 11% and *AAS44* about 34% of the normal activity in adults. The activities in 3rd instar larvae of *AHA51, AD369* and *AAS44* were each slightly lower than the controls *AC8* and *AC5.* Quantitative Northern analyses of the three alleles showed that in adults *AHA51* had 16% of the *AC8* transcript level, *AD369* had 19% and *AAS44* had 55% of the *AC5* transcript level, which are in close agreement with the activity levels. The transcript level in 3rd instar larvae of the three alleles parallels the ADH activity level, both in early and late third instar larvae.

Analysis of the molecular landscape surrounding the *Adh* locus detected insertions in each allele. A 1.15 kb insertion, which has an extra *Pvu*II site, was located in the distal promoter region of the *Adh* gene in *AD369.* A 2.4 kb insertion with extra *Eco*RI and *Pvu*II sites was present at the same region in *AHA51.* In the third allele *AAS44,* there was a 5.7 kb insertion located in the first intron of the *Adh* gene. Small insertions of 50 bp in *AD369* and 200 bp in *AAS44* were detected as well in a 5' upstream region. Except for the insertions, DNA sequence data on the *Adh* gene in three alleles did not reveal any single nucleotide changes that would result in amino acid substitutions that might be responsible for the low ADH activity. PCR amplification of the insertion region produced a band corresponding to the size of the KP element insertion in *AD369.* However, two bands were produced in *AHA51.* A large band had the expected size of the fragment with the *hobo* insertion. The small band had the

size corresponding to this region in the normal *Adh* allele. Sequence data showed that the insertion in *AD369* was a defective P element — 1.15 kb KP element located between the distal TATA box and the transcription start site. In *AHA51* the insertion was a defective *hobo* element located at exactly the same site as the KP element in *AD369*. The KP element in *AD369* and the *hobo* element in *AHA51* had the same target site (GTCCAAGT), which is present in the normal *Adh* sequence at nucleotide -13 to -6. This target site differs from the consensus site for KP element insertion (GNCCAGAC) at 3 out of the 8 nucleotides, while it matches the weak consensus site for *hobo* element insertion (NNNNNNAC). By using inverse PCR, the large insertion in *AAS44* has been defined as *Tirant* , a transposable element first found at the *extramacrochaetae* locus in *Drosophila*. *Tirant* was inserted in the *Adh* gene 409 bp 3' to the distal transcription start site, but 266 bp 5' to the proximal promoter.

Deletion of the KP element insertion in *AD369* by P-M hybrid dysgenesis resulted in normal transcript levels in adults. However, the ADH activity only returned to 50% of the normal. Sequence data confirmed that the KP element had been completely excised and the excision site was perfectly repaired without any of the KP nucleotides remaining. It is likely that the effect of the insertions on transcription arises from the increased distance between the distal promoter and the transcription start site.

Further investigation of the small band amplified in *AHA51* indicates that the defective *hobo* insertion in *AHA51* is possibly unstable both in germline and somatic cells. *Hobo* element movement gave rise to a variety of deletion products which have been cloned and sequenced. These data revealed that there were three different mechanisms involved in *hobo* element excision and post-excision repairing. Possible mechanisms of how the transposable element insertions located at different site in the promoter region affect the ADH activity in adults via transcript levels are discussed.

CONTENTS

Thesis title i

Declaration ii

Acknowledgements iii

Abstract iv

Contents vi

Chapter 1 General Introduction 1

Chapter 2 Materials and Methods

2.1 *Drosophila melanogaster* stocks 19

2.1.1 ADH low activity alleles 19

2.1.2 Other *Drosophila melanogaster* stocks 19

2.1.3 Breeding programs 21

2.2 Food preparation 21

2.3 Detection of ADH protein

2.3.1 ADH protein electrophoresis 22

2.3.2 ADH activity assay (spectrophotometric)

2.3.2.1 Preparation of flies for assay 22

2.3.2.2 Preparation of assay extracts 23

2.3.2.3 Measurement of assay reaction 23

2.4 Cloned *Adh* gene used for detection of *Adh* sequence 23

2.5 Genomic DNA extraction

2.5.1 Total DNA extraction 24

2.5.2 Single fly DNA extraction 25

2.6 Gel electrophoresis

2.6.1 Restriction enzyme digestion of DNA and separation of DNA fragments on agarose gels 26

2.6.2 Isolation of DNA fragments using the DNA purification kit 27

2.7 Hybridisation of genomic DNA

2.7.1 Southern blotting and hybridisation 27

2.7.2 Labelling of nick translated DNA probe 28

2.7.3 Restriction fragment mapping using Southern blots 28

2.8 The polymerase chain reaction

2.8.1 Preparation of oligonucleotide primers 29

2.8.2 Amplification of DNA fragments 29

2.8.3 Inverse PCR 30

2.9 Sub-cloning a PCR fragment into pBluescript

2.9.1 Digestion of insert and vector DNA 31

2.9.2 Ligation

2.9.2.1 'sticky end' ligation with double enzyme digestion 31

2.9.2.2 'blunt end' ligation 31

2.9.3 Making competent bacteria 32

2.9.4 Transformation 32

2.10 Plasmid DNA preparation (boiling mini plasmid preparation) 32

2.11 DNA sequencing

2.11.1 Sequencing of double-stranded plasmid DNA 33

2.11.2 Direct sequencing of PCR amplified DNA 34

2.11.3 Running and reading the sequencing gel 34

2.12 RNA analysis

2.12.1 Total RNA extraction 35

2.12.2 Northern blotting 36

2.12.3 PhosphorImage quantification 36

2.12.4 Removal of probes from nylon membranes 37

2.13 Deleting the KP element

2.13.1 Breeding program 37

2.13.2 Selecting flies with normal ADH levels by exposure to ethanol vapour 37

Chapter 3 The biochemical and molecular phenotypes of the three *Adh* variants

3.1 Introduction 38

3.2 Results

3.2.1 Cellulose acetate electrophoresis of adult and 3rd instar larval extracts 40

3.2.2 ADH activity in adults and 3rd instar larvae

3.2.2.1 *AHA51* 41

3.2.2.2 *AD369* 42

3.2.2.3 *AAS44* 43

3.2.3 ADH transcripts level in adults and larvae

3.2.3.1 Northern blotting for adult flies 43

3.2.3.2 Quantitative Northern analyses

3.2.3.2.1 Comparison of *AHA51/nLA248* with *AC8/nLA248* 45

3.2.3.2.2 Comparison of *AD369/nLA248* with *AC5/nLA248* 45

3.2.3.2.3 Comparison of *AAS44/nLA248* with *AC5/nLA248* 46

3.2.3.3 Quantitative Northern analyses with probe *BSSBN4* 46

3.3 Discussion 47

Chapter 4 Molecular Structure of the three *Adh* low activity alleles

4.1 Introduction 52

4.2 Results

4.2.1 Restriction maps of the *Adh* region in the three variants 53

4.2.2. DNA sequence analyses of the three low activity alleles

4.2.2.1 PCR amplification of the *Adh* gene 55

4.2.2.2 Sequence results of the *Adh* gene region

4.2.2.2.1 *AD369* 57

4.2.2.2.2 *AHA51* 58

4.2.2.2.3 *AAS44* 59

4.2.2.3 The *Adh* adult enhancer (AAE) region in the three variants 60

4.3 Discussion 61

Chapter 5 Molecular characterisation of insertions at the *Adh* locus in the three low activity variants

5.1 Introduction 66

5.2 Results

5.2.1 Identification of the insertion in *AD369* 69

5.2.2 Identification of the insertion in *AHA51* 70

5.2.3 Identification of the insertion in *AAS44* 71

5.3 Discussion 73

Chapter 6 Stability of the transposable elements at the *Adh* locus

6.1 Introduction 80

6.2 Results

6.2.1 Breeding program used to delete KP element from *AD369* 81

6.2.1.1 Ethanol vapour exposure to select flies with higher ADH activity 82

6.2.1.2 Molecular and biochemical analysis of *AD369**

6.2.1.2.1 Single fly PCR of *AD369*/nLA248* 82

6.2.1.2.2 ADH activity assay for *AD369*/nLA248* adults 83

6.2.1.2.3 Sequence analysis of *AD369*/Df(2L)64j* 83

6.2.1.2.4 Northern analysis of *AD369*/nLA248* adults 84

6.2.2 The hobo element in *AHA51* is unstable

6.2.2.1 PCR amplification of *hobo* region in *AHA51* 84

6.2.2.2 Sequence analysis of the 0.7 kb small fragment 85

6.2.2.3 Additional evidence of *hobo* instability 86

6.3 Discussion 88

Chapter 7 General Discussion 95

Bibliography 110

Appendices

Appendix I Comparison of the DNA sequence of the *Adh* gene in the three low activity alleles with the consensus *Adh* gene sequence 132

Appendix II Oligonucleotide primers used in *Adh* sequence analyses 137

Appendix III Oligonucleotide primers used in KP and hobo element sequence analyses 139

Appendix IV Comparison of the *Tirant* element sequences in insertions at the *Adh* locus in *AAS44* and at *emc* locus in the *Ach* mutant (Garrell and Modolell, 1990) 140

Appendix V Northern analysis of *Adh* transcripts in adults and larvae of *AD369/nLA248* and *AC5/nLA248* heterozygotes 141

Chapter 1 General Introduction

The regulation of gene expression in eukaryotes is a complex process which has been extensively explored. Gene expression can be regulated in many different steps, for instance, at the level of transcription, RNA processing, RNA transport, RNA degradation, translation and protein post-translational modifications (Alberts, 1994). However, transcriptional control is the key step that occurs in many genes. Genes can be regulated by modulating the pre-initiation complex (PIC) assembly, PIC activation (DNA melting), transcription initiation, promoter clearance, transcript elongation and termination (Roeder, 1996). Among these, initiation of transcription is the most important step in the regulation of gene expression.

It has been known for a long time that there are three nuclear DNA-dependant RNA polymerases in eukaryotes. Each of them is responsible for transcribing a specific subset of genes. RNA polymerase I transcribes genes coding ribosomal RNA; RNA polymerase III transcribes genes coding for small untranslated RNAs; most eukaryotic genes are transcribed by RNA Polymerase II (Sentenac *et al*, 1992; Willis, 1993).

Regulation of genes transcribed by RNA Pol II requires the interaction of cis- and trans- acting elements. The cis-acting elements include the core-promoter and gene-specific regulatory sequences. The core-promoter normally consists of a TATA box (consensus TATAa/tAa/t), located at position -30 to -25, and a pyrimidine-rich initiator (Inr, consensus YYANt/aYY) located near the transcription start site (Fig 1.0). Gene-specific regulatory sequences are located up to few thousand basepairs upstream of the transcription initiation sites (Verrijzer and Tjian, 1996).

The trans-action elements include general transcription initiation factors and gene-specific transcription activators. General transcription initiation factors of RNA Pol II have been identified from various organisms such as

human, rat, *Drosophila* and yeast. The components of these factors are remarkably conserved from yeast to human (Roeder, 1996). One of the general transcription factors (GTFs), TFIID, has DNA binding activity and is the only GTFs directing sequence-specific binding to eukaryotic promoter DNA. TFIID is a multi-protein complex composed of TATA-binding protein (TBP), which directly recognises the TATA element, and a set of TBP-associated factors (TAFs) required for transcriptional activation (Verrijzer and Tjian, 1996).

In the course of transcription, TFIID is the first component assembled on the core promoter of a TATA-containing Pol II-transcribed gene. It initiates the recruitment of RNA Pol II and other GTFs including TFIIA, TFIIB, TFIIE, TFIIF and TFIIH (Fig 1.0). TFIIA binding stabilises the TBP/DNA complex (Buratowski, 1994). TFIIB functions to recruit RNA Pol II (Buratowski *et al*, 1989). In calf thymus or Hela cell-derived extracts, TFIIF consists of two subunits, RAP30 and RAP74. The smaller, RAP30, is known to stimulate recruitment of RNA Pol II to the DNA template, the larger, RAP74, is required for stabilisation of the complex and to stimulate elongation (Zawel and Reinberg, 1993). TFIIH and TFIIE may be involved in promoter clearance by the polymerase. The kinase activity of TFIIH perhaps stimulate the phosphorylation of the C-terminal domain (CTD) of the Pol II large subunit (Goodrich and Tjian, 1994). The phosphorylation of CTD is an important step in polymerase release from the PIC (Buratowski, 1994).

Gene-specific activator proteins are thought to form the interface to input regulatory information to the transcription apparatus (Kornberg, 1996). Most efforts have been put into elucidating how transcription activators work at the level of pre-initiation complex (PIC) assembly and stabilisation. Most activator proteins have two essential domains: the DNA-binding motif and the activation domain that accelerates the rate of transcription initiation. The activation domain could be an acidic, or a glutamine-rich, or a proline-rich sequence. Various transcriptional activators can interact with GTFs. For

example, the glutamine-rich activation domain of the promoter specific factor Sp1 interacts specifically with $TAF_{II}110$ in *Drosophila* (Gill *et al,* 1994); a 39 amino acid activation domain of the activator virion protein (VP16) binds directly to *Drosophila* $TAF_{II}40$ (Goodrich *et al,* 1993); the acidic activation domain of viral VP16 also interacts with recombinant TBP *in vitro* (Nikolov and Burley, 1994). Interactions between the activator and GTFs affect PIC assembly at different levels including recruitment of GTFs to the PIC, stabilisation of GTFs within the PIC and conformational rearrangements of GTFs (Heard, 1995).

Another level of control of gene expression is the modulation of chromatin structure. It has been shown that the chromatin structure around promoters is in a dynamic state (Kingston *et al,* 1996). As general transcriptional factors (GTFs) can not directly assemble on to a promoter packaged in nucleosomal structure, transcriptional activators have to bind onto upstream enhancer sequences to displace the nucleosome at the promoter region and to make it accessible to GTFs and RNA polymerase (Alberts *et al,* 1994). For example, *in vivo,* the long terminal repeat (LTR) of mouse mammary tumour virus (MMTV) is incorporated into six positioned nucleosomes. When glucocorticoids bind to the glucocorticoid receptors to activate transcription, two nucleosomes adjacent to the TATA box are disrupted and start to assemble a transcription pre-initiation complex over this region (Archer *et al,* 1992; Perlmann and Wrange, 1991).

Chromatin structure can be changed by either translocating the nucleosome core (Georgel *et al,* 1993) or by outcompeting histone H1 — a much more powerful inhibitor (Croston *et al,* 1991). Histone assemblies can also have selective interactions with DNA. The selective wrapping of specific DNA sequences around the histones within positioned nucleosomes have an important role in controlling both the access of transcription factors to chromatin and the transcription process itself (Wolffe, 1994). The gene

products that are directly involved in altering chromatin structure during transcriptional activation have been identified. The products of SWI/SNF genes in yeast *S. cerivisiae* form a complex actively disrupting chromatin structure (Kingston *et al,* 1996). NURF (nucleosome remodelling factor) from *Drosophila* embryo extracts is able to facilitate perturbation of chromatin structure by the GAGA factor and increase restriction enzyme access in an ATP-dependent manner (Tsukiyama and Wu, 1995).

Numerous eukaryotic genes have been used to study different aspect of gene regulation. One of them is the alcohol dehydrogenase gene (*Adh*) in *Drosophila melanogaster*. The *Adh* gene of *Drosophila melanogaster* has become an excellent model system for the study of developmentally regulated gene/enzyme expression in higher eukaryotes since it was identified by Grell *et al* (1965; for review see Sofer and Martin, 1987). This has been largely due to the fact that the *Adh* gene/enzyme system can be easily manipulated and analysed at all levels from DNA to organism (Chambers, 1991).

The alcohol dehydrogenase enzyme (ADH: E.C.1.1.1.1) is a homodimer, with a monomeric subunit molecular weight of 27,500 daltons (Thatcher, 1980; Schwarts *et al,* 1979). *Drosophila* ADH is not a metalloenzyme but is able to oxidise a wide variety of alcohol substrates using NAD^+ as the cofactor. There is a single NAD^+/NADH binding site in each monomer, that gives rise to three isozymes by binding with either none, or one, or two moles of negatively charged NAD^+ (Winberg *et al,* 1982). As substrates, *Drosophila* ADH prefers short chain secondary aliphatic alcohols like propanol-2-ol or butan-2-ol (Sofer and Ursprung, 1968; Thatcher and Camfield, 1977; Winberg *et al,* 1982, Chambers *et al,* 1981). The ADH enzyme has a role in the metabolism of higher alcohols independent of alcohol detoxification (Winberg *et al,* 1982; Heinstra *et al,* 1986); it catalyses the inter-conversion of alcohols and their oxidation products (aldehydes and ketones), and also catalyses the oxidation of acetaldehyde to acetate (Heinstra *et al,* 1986; Eisses, 1989).

Flies lacking ADH die after exposure to high concentrations of ethanol (>6%) (Grell *et al*, 1968; Vigue and Sofer, 1976; David *et al*, 1976), but they are not affected by low concentrations of unsaturated secondary alcohols (1-pentyn-3-ol, 1-penten-3-ol), which rapidly kill wild-type flies. This allows the chemical selection of *Adh* null mutants (Sofer and Hatkoff, 1972; O'Donnell *et al*, 1975).

The ADH activity varies at different developmental stages and between specific tissues (Ursprung *et al*, 1970; Dunn *et al*, 1969; Maroni and Stamey, 1983). In general, ADH activity increases during the larval stages and reaches a peak in late third instar larvae. ADH activity drops just before pupariation and gradually declines further during pupation. After eclosion, ADH activity increases again for 4-6 days to reach a maximum and then declines very slowly during the life of the insect.

Most of the ADH activity is in the larval fat body, midgut, Malpighian tubules and in the adult fat body. Some activity is found in adult brain, gut, Malpighian tubules and part of the male reproductive tract, but not in the larval salivary glands, brain, or imaginal disks (Ursprung *et al*, 1970; Maroni *et al*, 1982). The level of ADH expression is modulated between tissues. For example, ADH is present at a high level in the larval fat body, and at lower levels in larval midgut and Malpighian tubules.

World-wide natural populations carry predominantly two allozymes of ADH: ADH-F and ADH-S (Johnson and Schaffer, 1973; Vigue and Johnson, 1973; Wilks *et al*, 1980; Anderson, 1981; Oakeshott *et al*, 1982), which are derived from Adh^F and Adh^S alleles respectively. The two allozymes differ in amino acid sequence by one residue with Lys-192 in ADH-S and Thr-192 in ADH-F. Both Adh^F and Adh^S strains have similar profiles of ADH activity during development (Maroni *et al*, 1982), but the level of activity differs between Adh^F and Adh^S lines. Generally Adh^F strains express higher ADH enzyme activities than Adh^S strains (reviewed by Laurie-Ahlberg, 1985).

Alcohol dehydrogenase is encoded by a single gene (*Adh*) mapped genetically at position 50.1 on chromosome II (Grell *et al*, 1965), and cytogenetically to 35B2-3 by *in situ* hybridisation (O'Donnell *et al*, 1977; Woodruff and Ashburner, 1979). The *Adh* gene in *Drosophila melanogaster* was first cloned by Goldberg (1980). Its orientation of transcription from 5' to 3' is positioned from left to right in the restriction map (Fig 1.1). The nucleotide sequence of the Adh^S gene was determined by Benyajati *et al* (1981), (1982) and Haymerle (1983). In different *Adh* alleles there are many silent polymorphisms in exons and introns, except the A to C substitution at position +1490 which is responsible for the replacement of the amino acid residue 192 from Lys to Thr in Adh^F alleles (Kreitman, 1983).

The *Adh* gene in *Drosophila melanogaster* extends about 1.8 kb and contains two separate promoters and three exons (Benyajati *et al*, 1981; 1983; Henikoff, 1983; Goldberg, 1980). A region about 1300 bp downstream of the *Adh* gene contains the presumptive coding sequence of *Adh-dup*, a functional gene that is tandemly duplicated from *Adh* (Schaeffer and Aquadro, 1987). The function of *Adh-dup* is not known. In some other species, like *D. mulleri*, there are two closely linked genes to encode two different ADH proteins, ADH-1 (larval) and ADH-2 (adult). *Adh-1* and *Adh-2* are expressed in a similar temporal pattern to that from the *D. melanogaster* proximal and distal promoters (Oakeshott *et al*, 1982; Batterham *et al*, 1983; 1984).

The coding region of the *Adh* gene in *D. melanogaster*, which encodes a 255-amino acid polypeptide (Thatcher, 1980), is split into three exons of lengths 96 bp, 405 bp and 264 bp by two small introns (Benyajati *et al*, 1981; Goldberg, 1980). The first intron is 65 bases and interrupts the protein sequence at amino acid residue 32, whereas the second intron (70 bases) interrupts the protein sequence at amino acid residue 167. The introns in the *Drosophila Adh* gene are not functional elements (Benyajati *et al*, 1981). This has been proved by

removing the introns, either one or both, and showing that the tissue-specific ADH activity was unaffected (Shen *et al*, 1989).

Although the ADH proteins in the larval stages and the adult are identical, their mRNAs are slightly different in size due to a difference at the 5' ends (Fig 1.1; Benyajati *et al*, 1983; Henikoff, 1983). There are two promoters for the *Adh* gene: the distal TATA box is 25 bp 5' to the initiation site of the adult mRNA and the proximal TATA box is 24 bp 5' to the 5' terminus of the larval mRNA. The adult ADH mRNA is transcribed from +1 to +87 and joined to the region from +742; a 654-bp intervening sequence exists from +88 to +741. The 5' end of the larval ADH transcript is initiated at nucleotide +708 and does not contain an intervening sequence. The 3' ends of both mRNAs are identical at +1858 (Benyajati *et al*, 1980; 1983; Kreitman, 1983; Appendix I). Therefore, the adult mRNA is 53 bp longer than the larval mRNA.

The two promoters of the *Adh* gene produce two transcripts at different developmental stages (Fig 1.1; Benyajati *et al*, 1983). The most abundant RNA in larvae is transcribed from the proximal promoter, while the most abundant RNA in adults is transcribed from the distal promoter.

A more detailed study of the temporal and spatial expression of the *Adh* gene shows a rather complex pattern of alternative utilisation of the two promoters (Fig 1.2; Savakis and Ashburner, 1985; Savakis *et al*, 1986; Lockett and Ashburner, 1989). Both promoters are active in the ovaries of adult females — they contribute maternal *Adh* RNA to the eggs. These maternally inherited *Adh* transcripts are distributed randomly in the embryo and decay very rapidly within an hour of fertilisation. The first zygotic expression of the gene starts about 10.5 hr later, and is confined to the fat body with transcripts both from distal and proximal promoters. However, transcription from the distal promoter is transient and can no longer be detected by 15 hr; meanwhile expression from the proximal promoter is increased and is seen in the gut. The steady-state concentration of *Adh* RNA during early larval development is

transcribed only from the proximal promoter and confined to the fat body and midgut. At the end of the larval stages *Adh* RNA level falls dramatically due to the cessation of transcription from the proximal promoter. Transcription from the distal promoter accumulates transitorily again in the larval fat body. Both transcripts disappear just before pupariation and then accumulate again a few hours prior to eclosion. The transcripts from the distal promoter increase quickly and remain high in adults for a few days, while the transcripts from the proximal promoter remain at a relatively low level in adults.

Adh mRNA has the same distribution patterns as ADH protein in most tissues like fat body, gastric caeca, and adult cardiac valve (Anderson *et al*, 1991). However, some tissues such as oocytes, nurse cells, imaginal disks and brain show *Adh* mRNA at a similar level, but with little or no ADH protein. The lack of concordance between *Adh* mRNA and ADH protein expression in oocytes and nurse cells may reflect the packaging of maternal *Adh* transcripts for use in early development. The other non-concordances in protein and mRNA expression suggest that the regulatory mechanisms may act at a post-transcriptional level.

All the *cis*-acting regulatory elements necessary for *Adh* expression are included in an 11.8 kb *Sac*I fragment which contains the *Adh* gene and its 5.5 kb 5'- and 4.5 kb 3'- flanking sequences (Goldberg *et al*, 1983). There are three regions determined to be necessary for correct *Adh* expression both in larvae and adults (Fig 1.3; Posakony *et al*, 1985). Sequences between nucleotide -660 to -69 upstream of the distal transcription start site are required for proper *Adh* expression from the distal promoter in adults. A region located 386 bp upstream of the proximal transcription start site is responsible for the temporal and tissue specificities of larval *Adh* expression. A third region that lies over 2 kb upstream of the proximal promoter controls normal levels of *Adh* expression in larvae (Posakony *et al*, 1985).

Corbin and Maniatis (1990) defined the *cis*-regulatory elements close to the proximal promoter by analysing transcription in deletion mutants (Fig 1.3). Two regions, from -386 to -242 and -110 to -80, are important for larval *Adh* expression as the transcript levels were decreased when these regions were absent. Deletion of the region from -386 to -242 reduces the ADH activity level in fat body and midgut, and also eliminates the ADH activity in Malpighian tubules (Corbin and Maniatis, 1990). The regions from -270 to -230 and from -98 to -76 have been shown by Heberlein *et al* (1985) to be protected from DNAase digestion.

To determine the importance of each element in the region up to -386 bp from the proximal promoter, a series of 50-bp deletions were made by Shen *et al* (1989). Their data showed that a region spanning the proximal TATA box was essential for *Adh* expression in larvae. Deletion of the region from -115 to -66 affected tissue specificity. Other deletions in this region affected the level of *Adh* transcription from the proximal promoter. In particular, a 53-bp sequence from -341 to -289 behaves like an enhancer, as it can stimulate *Adh* expression when positioned either side of the *Adh* gene in transformed constructs (Rothberg *et al*, 1991). A 12-bp sequence from -305 to -293 is important. When this 12-bp sequence was removed, the resultant construct did not show any ADH activity (Shen *et al*, 1989; 1991). These results show that each sequence element in this region is important for wild-type *Adh* expression in larvae (Shen *et al*, 1991; Corbin and Maniatis, 1990).

The *Adh* larval enhancer (ALE), which is necessary and sufficient for the normal level of *Adh* expression in all larval tissues, lies further upstream of the proximal promoter (Posakony *et al*, 1985; Corbin and Maniatis, 1989b). The ALE includes three elements with the 5' region lying between -5000 to -3550 bp and the 3' region lying between -1845 to -660 bp from the distal transcription start site (Fig 1.3). The first element, located between -660 and -1845, slightly boosts the larval enhancer activity; the second element, located between -1845

to -2395, is necessary for wild-type *Adh* expression in larvae; the third element, located between -3550 and -5000, is essential for activating the larval enhancer. The individual larval enhancer elements cannot stimulate *Adh* expression independently, but act synergistically to achieve maximal *Adh* expression in larvae.

Sequences immediately upstream of the distal promoter control *Adh* expression in adults. When the region of the distal promoter including the TATA box and transcription start site was deleted, *Adh* transcripts from the proximal promoter increased (Posakony *et al*, 1985). This result implies that the distal promoter is a negative regulatory element to the proximal promoter and the upstream sequence from the distal promoter functions as a positive regulatory element to the proximal transcription. A 142-bp *Adh* adult enhancer (AAE) is located in this region from -615 to -473 (Fig 1.3; Falb and Maniatis, 1992a; Appendix I). Further studies have shown that the AAE, which contains both positive and negative elements, is necessary for the *Adh* expression from the distal promoter (Falb and Maniatis, 1992a).

The *Adh* adult enhancer (AAE) and larval enhancer (ALE) differ from each other in function. The AAE is stage-specific and stimulates *Adh* expression only in late third instar larvae and in adults; it can stimulate transcription both from distal and proximal promoters. The ALE is active in all developmental stages, but it can only interact with the proximal promoter and not directly with the distal promoter (Corbin and Maniatis, 1989b).

Adh expression during the development is switched from the proximal promoter to the distal promoter by stage-specific activation of the distal promoter and subsequent repression of the proximal promoter (Fig 1.4; Corbin and Maniatis, 1989a). The factors bound to ALE can interact with the proximal promoter directly to initiate *Adh* expression from the proximal promoter in larvae. Until late third instar larvae, the factors bound to ALE and AAE interact with each other to activate the AAE and then to stimulate the

transcription from the distal promoter. Distal transcription interferes with the interaction between the ALE and the proximal promoter. When distal transcription increases in adults, proximal transcription is repressed by read-through transcription from the distal promoter (Corbin and Maniatis, 1989b). The switch from proximal to distal transcription occurs both in Malpighian tubules and in larval fat body in a single cell (Savakis *et al*, 1986).

By comprehensive analyses of the results of deletion mutagenesis experiments and DNA binding experiments, numerous trans-acting factors have been found to bind to the *Adh* adult enhancer (AAE), the distal promoter and the proximal promoter region (Heberlein *et al*, 1985).

At least three regulatory elements within the AAE are required for adult *Adh* gene expression. Two of them are bound with positive factors: Box B-binding factor (BBF-2) and CCAAT/enhancer-binding protein (C/EBP). The other one is bound with a negative factor — adult enhancer factor-1 (AEF-1) (Fig 1.5; Abel *et al*, 1992; Falb and Maniatis, 1992b). BBF-2 is a member of the CREB/ATF family of transcriptional regulatory proteins and binds to all fat body-specific enhancers of the *Adh* and the *york* gene (Abel *et al*, 1992). BBF-2 protects a 19-bp region in AAE from -576 to -557, which is overlapped with a weak C/EBP binding site (Abel *et al*, 1992). C/EBP is a transcriptional activator that strongly protects a region in AAE from -518 to -510 upstream of the distal transcription start site. The C/EBP can increase the level of ADH activity about 50~100 fold, and can be repressed by the negative factor AEF-1. AEF-1 binds adjacently to the C/EBP site from -534 to -518 in AAE. Both factors competitively bind to their site with the AEF-1 factor playing a dominant role, as the AEF-1 can displace pre-bound C/EBP *in vitro* (Falb and Maniatis, 1992a).

Using a culture cell system and transient-transfection analyses of deletion mutations, further positive and negative protein binding sites have been found in AAE (Fig 1.5; Ayer and Benyajati, 1990). A negative site bound to the distal enhancer protein 4 (DEP4) is found between -531 to -524, and this

overlaps with the AEF-1 site in AAE (Fig 1.5). Another region from -496 to -481 is found to be bound with two distal enhancer proteins DEP1 and DEP2. One of them, DEP1, is recognised by FTZ-F1, a *Drosophila* member of the steroid hormone receptor superfamily (Ayer and Benyajati, 1992). A second steroid hormone receptor, the *Drosophila* Hormone Receptor 39 (DHR39), also binds to this site (Ayer *et al*, 1993). DHR39 represses distal *Adh* expression, while FTZ-F1 activates distal *Adh* expression. As both FTZ-F1 and DHR39 share the same site in AAE, it implies that they act competitively to modulate the level of *Adh* transcription. The third distal enhancer protein (DEP3) binds to a region from -454 to -427 and is not correlated to the level of ADH expression (Ayer and Benyajati, 1992).

Two more factors have been shown to bind to the upstream sequences of the *Adh* distal promoter (Fig 1.5). The *Adh* distal factor-2 (Adf-2), which binds to a 10-bp direct repeat motif, functions as a repressor of distal transcription in cells which contain inactive distal promoter (Benyajati *et al*, 1992). There are two binding sites for Adf-2, which are located at +8 and -202, flanking the *Adh* distal promoter factor-1 (Adf-1) binding site and the distal TATA box (Fig 1.5). Adf-1 is a transcriptional activator binding to a region from -85 to -47 upstream of the distal transcription start site. Adf-1 also binds to a region upstream of the proximal promoter from -150 to -104 (P1 site) (Fig 1.5). However, Adf-1 stimulates transcription only from the distal promoter, not from the proximal promoter (Heberlein *et al*, 1985; England *et al*, 1990; 1992).

Transcription from the proximal promoter is regulated by two other proteins that bind to the upstream region of the proximal promoter. The P2 site lies between -270 to -230 and the P0 site lies between -92 to -76 from the proximal transcription start site. The P2 and P1 binding sites are overlapped with *cis*-regulatory elements for *Adh* expression from the proximal promoter (Fig 1.3 and 1.5; Heberlein *et al*, 1985). The P0 protein is a Box A-binding factor (ABF) which is found in *D. mulleri* and at the proximal promoter of the *Adh*

gene in *D. melanogaster*. ABF belongs to the GATA family of transcription factors and functions as an activator of proximal transcription (Abel *et al*, 1993).

Another level of transcriptional control *in vivo* is the chromatin structure. The chromatin structure of *Adh* distal promoter is altered from the stages not transcribing *Adh* to the stage transcribing *Adh* (Ewel *et al*, 1990). In cells actively transcribing *Adh*, the distal promoter was occupied by Adf-1 or TATA binding factors. The DNA between the TATA box and Adf-1 are possibly bent or kinked to facilitate Adf-1 interacting with general transcription machinery. It forms a multi-protein complex involved in specific transcription initiation of the distal promoter. A longer linker between -140 to +30 is bound in the multi-protein initiation complex. These DNA-protein interactions account for an open chromatin structure at the distal promoter. These regions are nucleosome free and accessible to non-histone proteins. DNase I hypersensitivity exists around -200 to +50 of the distal RNA start site.

In cells not transcribing the distal *ADH* RNA, although both Adf-1 and the TATA binding factor are present and functionally active (Benyajati *et al*, 1987; Heberlein *et al*, 1985), the binding sites for Adf-1 and TATA binding factors at the distal promoter are occupied by a specific positioned nucleosome and factor Adf-2. The two Adf-2 binding sites (see Fig 7.2), which are nearly 200 bp apart, might be in closer proximity as the DNA in between them is apparently bound to a positioned nucleosome core. A short linker (from -40 to +30 for the distal RNA initiation site) exists between the positioned nucleosomes, which interacts with a specific DNA-binding protein. The chromatin structure of these regions is protected from DNase I (Ewel *et al*, 1990). The *Adh* larval enhancer (ALE) is large and complex and its chromatin structure has not been precisely defined (Jackson and Benyajati, 1993).

Taken together, we can see that there are potentially many different ways in which part or all of normal transcription could be interrupted by mutations. Mutations artificially induced or naturally occurring have been

extensively used in elucidating the structure and function of the *Adh* gene. The classical genetic approach to analyse enzyme function has been to artificially induce mutations which affect the production of enzyme. Some chemicals have been used as mutagens to create ADH negative mutants that can be isolated by using chemical selection procedures (Sofer and Hatkoff, 1972; O'Donnell *et al*, 1975). For example, formaldehyde treatment has been used to produce small deletions between 6 to 69 base pairs in the *Adh* gene (Benyajati *et al*, 1981; Le *et al*, 1990). The mutation frequency at the *Adh* locus after formaldehyde treatment is about 30-50 times higher than the spontaneous mutation rate (Benyajati *et al*, 1981; 1982; Hipeau-Jacquotte *et al*, 1989; Le *et al*, 1990; O'Donnell *et al*, 1975; Place *et al*, 1987).

Other chemical mutagens such as ethyl methane sulphonate (EMS) and N-nitroso urea (ENU) have been used to induce mutants at the *Adh* locus. One of the EMS-induced mutants, Adh^{n4}, produced mRNA at only 10% of the wild type level because of changes in the sequences flanking the *Pvu*II site within the second exon, despite the normal processing of the transcript (Chia *et al*, 1987). Flies treated with ENU have point mutations and the frequency of the ENU-induced recessive lethals is approximately 300 times higher than the spontaneous frequency (Fossett *et al*, 1990).

Another mutagen, X-rays, have been commonly used to produce mutations at the *Adh* locus. The frequency of mutation at the *Adh* locus following X-ray irradiation is estimated as 7×10^{-8} per rad per locus (Aaron, 1979). The intragenic mutations produced by X-ray irradiation are single-site or small deletions. Aaron (1979) recovered 31 mutations induced by X-ray, twenty-two were classified as deletions; three were null alleles but they had normal mRNA; in the remaining four, two showed insertions and two showed deletions by Southern analysis (Kelley *et al*, 1985). One of the alleles, Adh^{nLA248}, has been shown to produce a 250-bp longer transcript but at a similar level to the wild-type gene (Chia *et al*, 1985b). This allele has been used

as an *in vivo* control in studying levels of *Adh* transcript (Laurie and Stam, 1988; Savakis *et al*, 1986).

The laboratory-induced mutations were all found to be structural changes in the *Adh* gene that altered ADH activity, but no control mutations were obtained — none were in putative regulatory regions. Therefore, one putative control mutant, Adh^{SL}, isolated by Thompson *et al* (1977) from a natural population is of particular interest. This mutant showed lower ADH activity in adults than in larvae. Genetic analyses of this allele showed that the low activity phenotype was separable by recombination from the structural gene. Later, molecular structure analyses revealed that Adh^{SL} carried a 4.5-kb insert approximately 3 kb 5' to the distal promoter (Schott *et al*, 1988).

In natural populations of *Drosophila melanogaster*, high levels of DNA polymorphism within the structural and regulatory regions of *Adh* genes have been found and reviewed by Aquadro *et al* (1990). The polymorphisms include basepair substitutions, unique sequence insertions/deletions or transposable elements insertions. Whether certain types of polymorphism have greater effects on the levels of *Adh* expression is not clear; individual variants need to be isolated and analysed for their phenotypic effects (reviewed by Laurie-Ahlberg, 1985).

In a survey of variation in restriction fragment length polymorphism (RFLP) in the *Adh* gene region, 80% of the chromosomes had an insertion or deletion compared to a consensus restriction map (Langley *et al*, 1982; Aquadro *et al*, 1986). Most of the length changes are unique insertion/deletions up to ~200 bp, but there are also several transposable element insertions ranging in size from 0.34 to 10.2 kb. Investigation of restriction endonuclease variation in Australian populations revealed ten insertions (Jiang and Gibson, 1992a), two of them shared homology with known mobile elements. The majority of the insertions and deletions had only a slight effect on ADH activity, with only a 1.5 kb insertion 2kb from the 5' end of the *Adh* coding region and a 5 kb

insertion within the first intron of the *Adh* being associated with low activity (Jiang and Gibson, 1992b). Another well-studied low ADH activity allele, Adh^{RI42}, contained a *copia* insertion located approximately 240-bp upstream of the distal promoter (Strand and McDonald, 1989). The *copia* insertion mainly accounted for the unusual low ADH activity in Adh^{RI42} adults (Dunn and Laurie, 1995).

Most spontaneous mutations used in *Drosophila* genetics are caused by the insertions of transposable elements (Finnegan and Fawcett, 1986; Green, 1980; Rubin, 1983). The range of transposable elements sizes are 1-9 kb. The *Drosophila* genome contains about 3000 copies of transposable elements, which consists of 30 to 50 different families and occupies about ten percent of the fly DNA (Charlesworth and Langley, 1989; Biemont, 1992). *Drosophila* transposable elements are distributed throughout the genome (Charlesworth and Langley, 1989), and are integrated into almost every band of the giant chromosomes in the salivary glands (Ananiev *et al*, 1984). Many families of transposable elements tend to accumulate in the euchromatic polytene regions embedded in the chromocenter (Berg and Howe, 1989; Finnegan, 1989; McDonald, 1993). Recent studies have shown that transposable elements are stable components in regions of mitotic heterochromatin in *Drosophila melanogaster* (Pimpinelli *et al*, 1995; Biemont, 1992).

Transposable element insertion mutations have been associated with all aspects of alterations in chromosome structure, recombination, replication and gene regulation (McDonald, 1993). Transposable element insertions at different site have different effect on host genome. TE euchromatic insertions usually reduce the fitness of their host genome and won't be expected to have any positive roles in the evolution of their hosts (Fitzpatrick and Sved, 1986; Eanes *et al*, 1988). The majority of heterochromatic TE insertions do not cause mutations with negative impact on fitness, therefore they are not subjected to

elimination by recombination (Montgomery *et al*, 1987; Charlesworth and Langley, 1989).

Most TE mutations are against natural selection (Fitzpatrick and Sved, 1986; Eanes *et al*, 1988; Charlesworth *et al*, 1992). Only those insertion mutations which do not have negative effects on fitness are selected and preserved to fix the TE insertion mutant. For example, some of transposable elements contain control sequences known to be critical for proper regulatory expression of eukaryotic genes. These short stretches of regulatory sequences are preserved by natural selection (McDonald, 1993).

A high number of insertions are maintained in *Drosophila* genome, but do not produce many mutations. Spontaneous morphological mutations or lethal mutations occur at a fairly low rate — about 10^{-3} to 10^{-4} per element per generation (Young and Schwartz, 1981; Pierce and Lucchesi, 1981). However, transposition rates differ among different transposable elements and in chromatin regions. The rates can be induced to increase in certain circumstances such as the P element in the hybrid dysgenesis (Biemont, 1992). Spontaneous loss of elements from their chromosomal locations is usually very rare or imprecise (Rubin, 1983; Woodruff *et al*, 1987). Usually, the rates of excision are ten times smaller than the rates of transposition. In addition, removal of insertions usually leads to reversion or alteration of a mutant phenotype which is associated with the insertion (Engels, 1989; Finnegan and Fawcett, 1986).

Therefore, transposable elements are maintained in *Drosophila* populations as a result of transpositional increase in copy number, balanced by some opposing forces (Charlesworth and Langley, 1989). Transposable elements could act indirectly as a source of mutational variation for evolutionary change — they have played a major role in population evolution by making genes mutate or producing gross chromosome rearrangements, and help populations to adapt to changing environments (Biemont, 1992).

Naturally occurring variants with null or reduced activity have recently been intensively analysed, as they might provide a different spectrum of mutations from those artificially produced. *Adh* null alleles were found at unexpectedly high frequencies in Australian populations (Freeth and Gibson, 1985). Although they are similar in properties, the causes of the null activity are different. An 8 bp insertion in intron 2 disrupting correct splicing has been found in all the Tasmanian null alleles analysed. It causes an altered transcription pattern (Gibson and Wilks, 1989; Freeth *et al*, 1990). Among the mainland null alleles an insertion of 320 bp in intron 1 and a deletion of 438 bp including most of exon 2 were found (Gibson *et al*, 1992).

In addition to the structural null alleles, which were null both in larvae and adults, a number of alleles were identified in the Australian surveys which had normal levels of ADH activity in 3rd instar larvae but very low activity in adults. These alleles were similar in phenotype to the Adh^{SL} allele . Three of these *Adh* low activity alleles, *AHA51, AD369* and *AAS44*, are the focus of this thesis. Detailed molecular investigations of these *Adh* low activity alleles are worthwhile, because they might provide further clues about the way the *Adh* gene-enzyme system functions and how the expression of the *Adh* gene is differentially regulated.

The origins of the *Adh* variants that I have studied are described in Chapter 2; the biochemical and molecular phenotypes of the three *Adh* variants are presented in Chapter 3. Chapter 4 describes the molecular structure of the three *Adh* low activity alleles and based on the findings in this chapter, detailed characterisation of the insertions at the *Adh* locus in each of the three variants are described in Chapter 5. Finally, the effect of the insertion on the expression of the *Adh* gene is investigated and the stability of the inserts is discussed in Chapter 6.

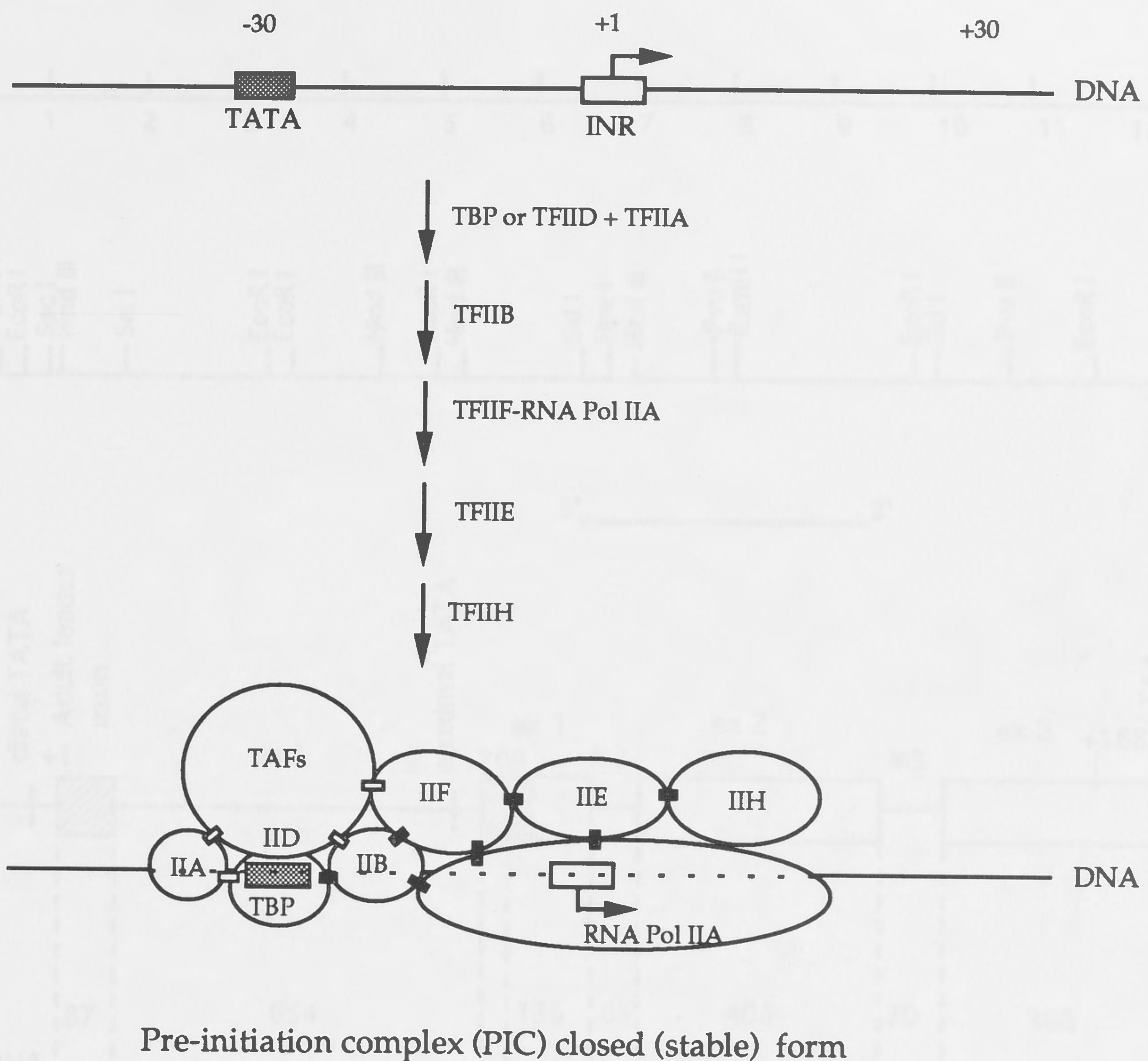
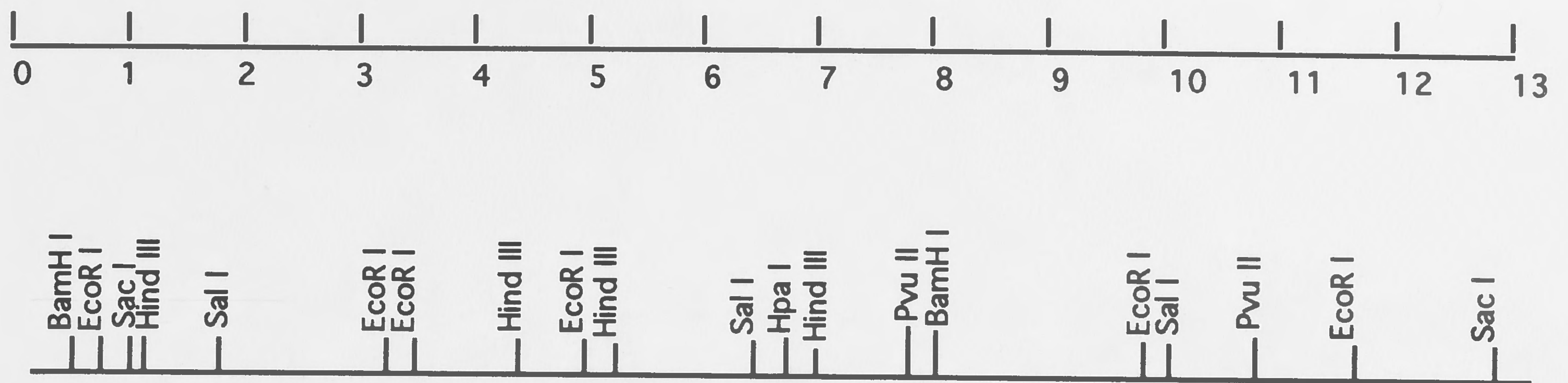


Figure 1.0 Model for pre-initiation complex (PIC) assembly on a TATA-containing core promoter (derived form Roeder, 1996).

The core-promoter elements, TATA (TATA box) and INR (initiator) are indicated. Together with the TATA-binding protein (TBP), a minimal set of general transcription factors (TFIIA, -B, -E, -F, -H and RNA Plo II) are required for the stepwise assembly of the pre-initiation complex (PIC). Short solid bars indicate protein-protein interactions between the factors. Short open bars represent stablilizing interactions between TAFs and other factors.

a).



5' 3'

b).

distal TATA
Adult leader exon
+1
in 1
proximal TATA
+708
ex 1
in2
ex 2
in3
ex 3
TAA
+1681
poly A
+1858
87
654
135
65
405
70
388
distal mRNA
169
proximal mRNA

Figure 1.1 Structural organization of the *Adh* gene of *Drosophila melanogaster*.

a). A restriction map of the *Adh* region. The 5' to 3' orientation of the ADH mRNA is indicated as a thick line from left to right below the map. b). The *Adh* exons (ex) are shown as boxes with the untranslated regions hatched; introns (in) are shown as single lines. The adult and larval transcripts are indicated together with the sizes (bp) of the different regions.

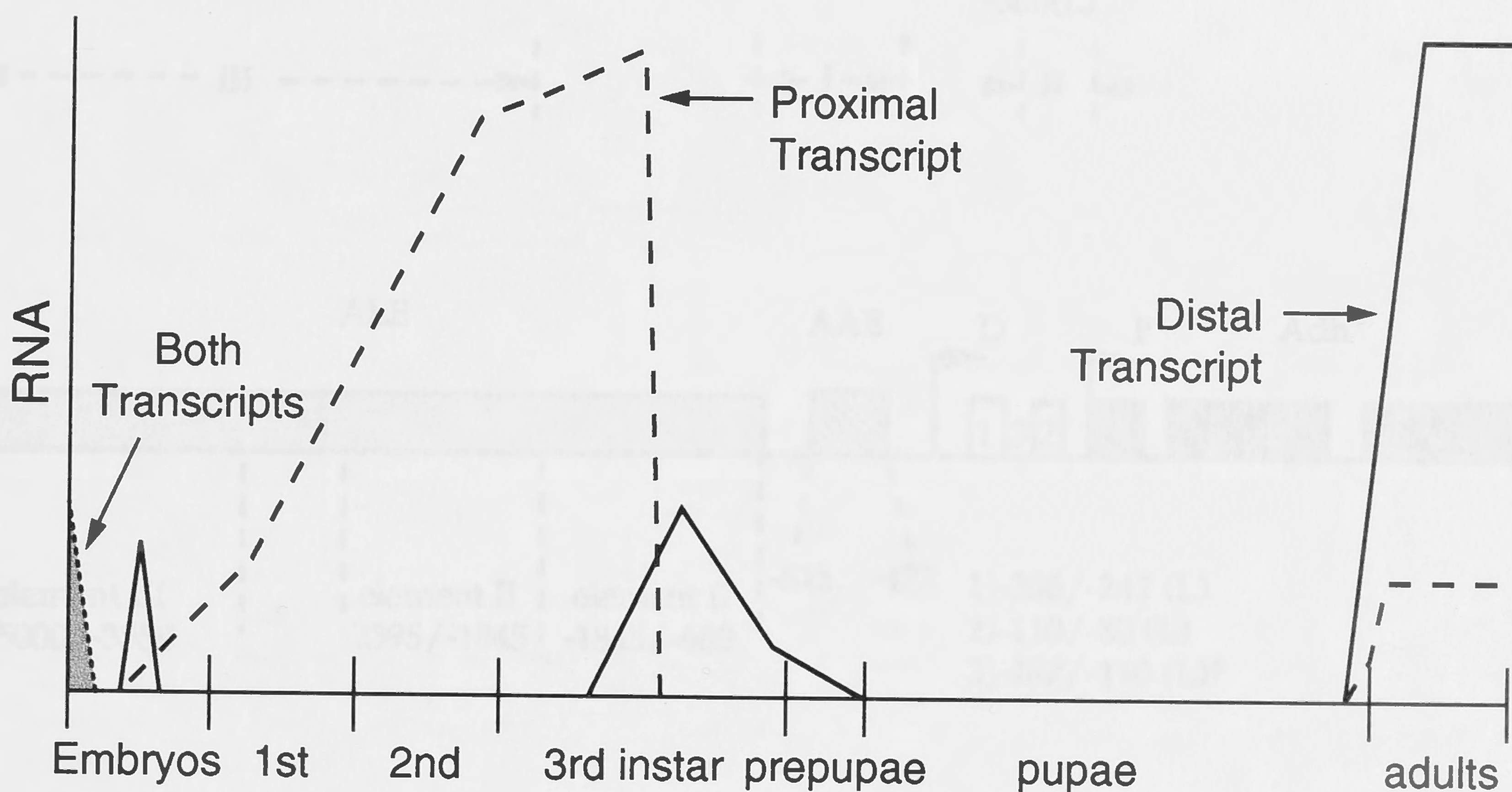


Figure 1.2 ***Adh*** **transcripts during the development of** ***D. melanogaster***
The vertical scale reflects the *Adh* mRNA concentrations and the horizontal scale shows different developmental stages. Transcription from the distal promoter is indicated as solid lines, from the proximal promoter as dashed lines. Maternally inherited transcript transcribed from both promotors is shown as a shaded area.

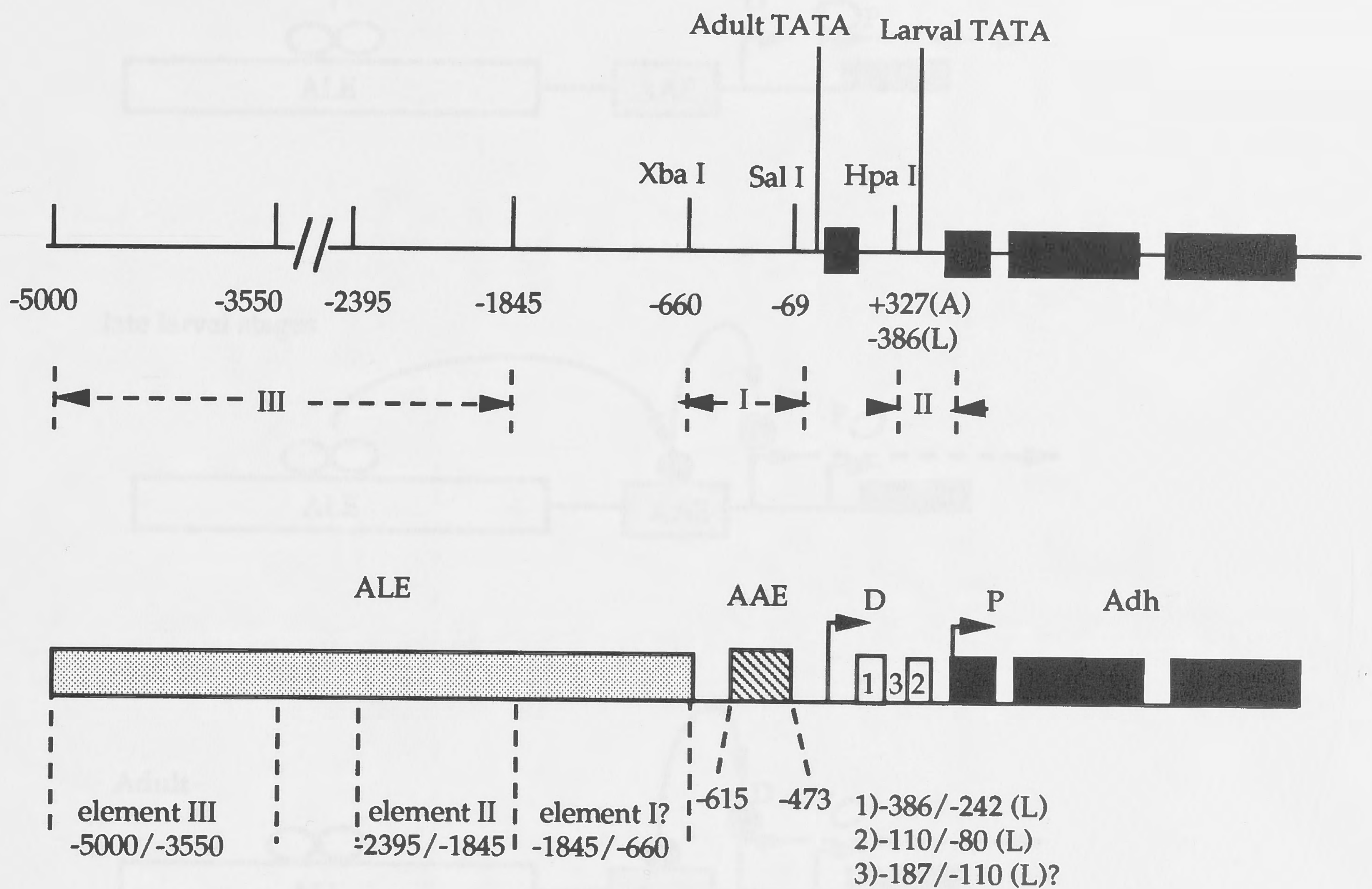


Figure 1.3 Localization of the *cis*-regulatory sequences of the *Adh* gene in *Drosophila melanogaster*

The *Adh* exons are shown as black boxes with the positions of the TATA boxes indicated in the upper figure. In the lower figure, the *Adh* adult enhancer (AAE) is shown as a shaded box and the larval enhancer (ALE) as a dotted box; the limits of the three larval enhancer elements are indicated. The regulatory elements near the proximal promoter are shown as small white boxes marked with numbers. The transcription start sites are indicated by solid arrows.

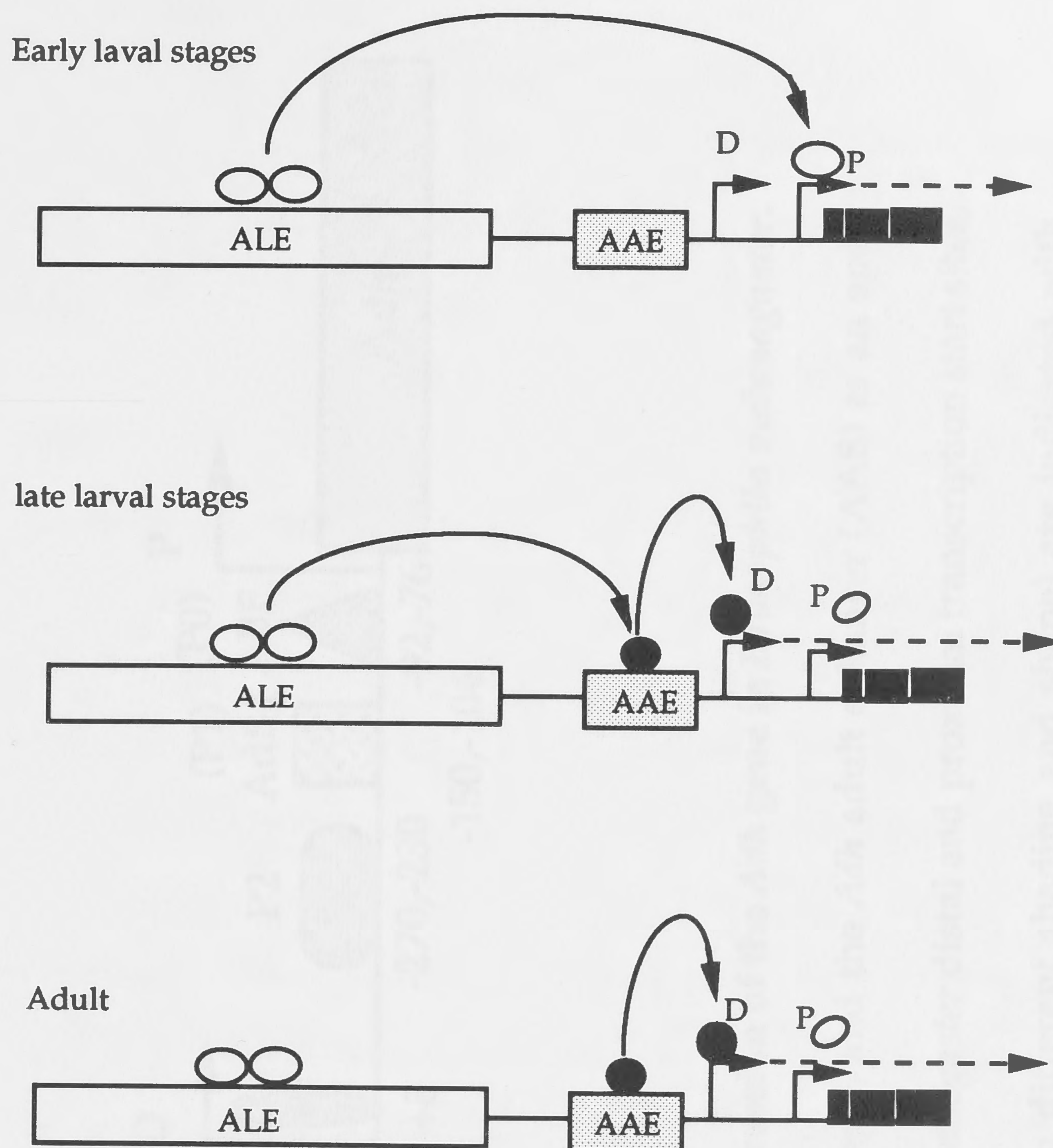


Figure 1.4 Diagram showing enhancer-promoter interactions in different developmental stages (Corbin and Maniatis, 1989)

In early larval stages, the *Adh* larval enhancer (ALE) stimulates transcription directly from the proximal promoter by interaction between ALE-binding factors (open circles) and the proximal promoter.

In late third instar larvae, the distal promoter is stimulated by activators bound both to ALE and AAE (solid circles) to interact with the distal promoter factor. The proximal promoter factor is repressed by read-through transcription from the distal promoter.

In adults, only factors bound to AAE and the distal promoter interact to stimulate transcription from the distal promoter, while the proximal promoter is turned off by transcriptional interference.

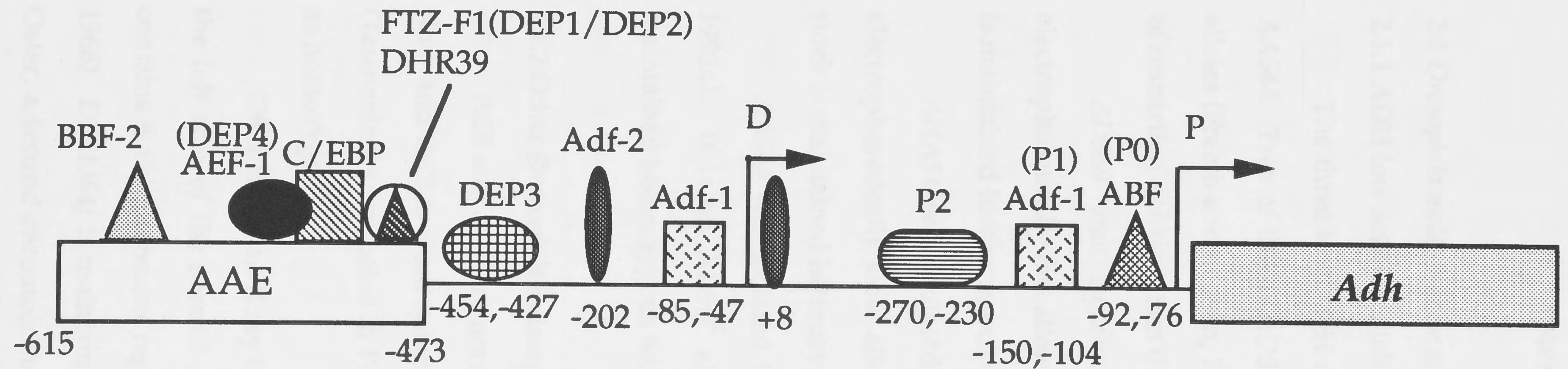


Figure 1.5 Trans-acting factors regulating the expression of the *Adh* gene in *Drosophila melanogaster*. The *Adh* coding region is shown as a dotted rectangle, and the *Adh* adult enhancer (AAE) as an open rectangle. D and P labelled arrows indicate the *D. melaogaster* distal and proximal transcription start sites, respectively. The binding sites for each factor (in different shading and shape) are indicated with numbers which refer to distances from the approriate transcription start site.

Chapter 2 Materials and Methods

2.1 *Drosophila melanogaster* stocks

2.1.1 ADH low activity alleles

The three low ADH activity alleles I studied were *AD369*, *AHA51* and *AAS44*. Two of them (*AD369* and *AHA51*) came from a screen for *Adh* null alleles (Freeth and Gibson, 1985). The third line *AAS44* was found in a survey of restriction site variation (Jiang and Gibson, 1992a).

AD369 was isolated in 1987 from All Saints (VIC). It is an electrophoretically fast allele for ADH. The homozygotes are sterile. The stock is maintained heterozygous with *CyO* or *Df(2L)64j* (see 2.1.2).

AHA51 was isolated in 1984 from Coffs Harbour (NSW). It is an electrophoretically slow allele for ADH. The homozygotes are sterile. The stock is maintained heterozygous with *CyO*.

AAS44 was isolated in 1986 from All Saints (VIC) (Jiang and Gibson, 1992a). It is an ADH^F allele and the homozygote is lethal. *AAS44* is maintained heterozygous with *Df(2L)64j*.

2.1.2 Other *Drosophila melanogaster* stocks

AC8 and *AC5* are normal activity *Adh* alleles. *AC8* is electrophoretically slow and *AC5* is electrophoretically fast. They were isolated from the Cygnet (Tasmanian) population in 1984 (Gibson *et al*, 1991). Both stocks are maintained as homozygotes.

Df(2L)64j is an X ray induced deficiency covering 34D1-2 to 35B9-C1 on the left arm of the second chromosome (Lindsley and Zimm, 1992), which contains the chromosome region encompassing the *Adh* gene (Lindsley & Grell, 1968). *Df(2L)64j* is maintained as a heterozygote with *CyO*. *CyO* is Curly of Oster, a second chromosome that contains multiple inversions, *In(2LR)O, Cy* $dp^{lvI} pr\ cn^2$, the order is 21-22D/33F5-30F/50D1-58A4/42A2-34A1/22D2-

30E/50C10-42A31/58B1-60 (Lindsley and Zimm, 1992). It contains an active *Adh* fast allele (*Adh*^{F}), and some visible markers including the dominant marker *Cy*. It is a balancer for the second chromosome.

Df(2L)GpdhA is an X-ray induced deficiency that includes the *Gpdh* locus. The region from 25D7-E1 to 26A8-9 on the left arm of second chromosome is absent (Lindsley and Zimm, 1992). It is kept heterozygous with *CyO*.

Df(2L)A379 is an X-ray induced ADH null allele. It has a deficiency from 35B1-3 on the left arm of the second chromosome and has In(2LR)35B1-3 (Lindsley and Zimm, 1992).

Adh^{nB} is an opal (UGA) nonsense mutant (Kubli *et al*, 1982) and was isolated after EMS mutagenesis of a stock carrying the *CyO* chromosome. A single base substitution at the TGG tryptophan codon at amino acid 235 to a TGA termination codon accounts for the reduction in size of the alcohol dehydrogenase polypeptide (Martin *et al*, 1985). *CyO Adh*^{nB} is maintained as a balanced stock in heterozygotes with *Df(2L)A379*.

Adh^{nLA248} is an X-ray induced *Adh* null allele (Aaron, 1979). It has a 250 bp insertion formed by unequal crossover between exon 3 (at +708) and exon 2 (at +465) with 7 bp (GTGCAAC) inserted at the junction (Chia *et al*, 1985b). The chromosome bearing *Adh*^{nLA248} also carries the markers *cinnabar* (*cn*, localised at 2-57.5) and *brown* (*bw*, localised at 2-104.5) which when homozygous give white-eyed flies (Lindsley and Grell, 1968) — the *Adh* locus is at 50.1 on chromosome II. This stock was used to generate heterozygotes with *Adh* low activity alleles to provide *in vivo* controls for Northern analyses. The stock is maintained as a homozygote.

w^1*;Sb e P[ry*$^+$*Δ2-3]/TM6,Ubx* was obtained from Dr. G. Miklos. It contains an incomplete P element to provide a source of transposase (Robertson *et al*, 1988).

SM5,CyO/L, was obtained from Dr. A. Ladvishenko. *Lobe* is a dominant gene localised on the second chromosome at 2-72.0. It causes a variable

reduction in eye size of heterozygotes and homozygotes. It is kept heterozygous with *CyO* (Lindsley and Zimm, 1992).

2.1.3 Breeding programs

All the stock flies I used were maintained on single yeast food at 20°C. For experimental purposes, the cultures were made on a high protein food at 25°C. Single crosses were made in 4 inch vials. Mass cultures were made in 200ml bottles with 5-10 pair flies in each. Virgin females were collected less than 12 hours after emergence at 25°C. They were kept at 20°C on single yeast food until used for setting up crosses.

2.2 Food preparation

Single yeast food was made of 10g agar, 6g yeast, 26g sucrose, 50g glucose, 50g maize meal and 22.5g wheat germ, mixed with 1 litre water, boiled and simmered for a few minutes. When the food cooled to 45°C, 12.5ml acid mix (proportions: 541ml H_2O, 418ml propionic acid, 41ml orthophosphoric acid) was added and mixed well before pouring into bottles and vials.

High protein food was made of 10g agar, 50g yeast, 15g sucrose, 50g malt, 30ml karo and 10g hi-pro with 1 litre water, boiled and simmered for at least 20 minutes. 4.5ml propionic acid was added when the food cooled to 45°C. This medium was used for harvesting flies for nuclei acid extraction or for culturing flies for ADH assays.

Riboflavin food is specially used for raising *Drosophila* larvae to distinguish those that are homozygous or hemizygous for the white eye gene. It was made by adding 0.3g riboflavin per litre high protein food (Ashburner, 1989).

Apple food is used for collecting *Drosophila* eggs. It was made as follows: 22.5g agar was mixed with 750ml water, boiled and simmered for a few minutes. 22g sucrose was dissolved in 250ml apple juice, then mixed with

agar and poured into petrie dishes. Just before use, freshly made yeast paste (dry yeast mixed with water to form a paste) was spread onto the surface of this food.

2.3 Detection of ADH protein

2.3.1 ADH protein electrophoresis

ADH protein electrophoresis was carried out on Cellulose Acetate Membranes (Chemetron). Before electrophoresis the membranes were removed from storage buffer (30% methanol) and soaked in the running buffer (35mM Tris, pH8.8; 5mM boric acid; 3.5mM EDTA) for at least 20 minutes. A single fly or larva was ground in 10µl distilled water before applying to the membrane. Electrophoresis was carried out at 4mA (300V) for 30 minutes per gel. Staining for ADH was carried in 50mM orthophosphate buffer (pH7.8, KH_2PO_4 0.722g/l, K_2HPO_4 8.01g/l, EDTA 37.22mg/l), 0.5mM tetrazolium salt (MTT), 1% isopropanol, 0.5mg/ml NAD and 0.01mg/ml phenazine methosulphate. Staining was continued for about 10 minutes in the dark until the intensity was satisfactory and it was then stopped by washing the membranes in water.

2.3.2 ADH activity assay (spectrophotometric)

2.3.2.1 Preparation of flies for assay

Adult flies used for ADH assay were raised from single crosses on high-protein food at 25°C. Cultures of homozygous flies were set up by putting a single fertilised female in each vial. Crosses were set up by pairing a virgin female from one line and a male from the other line. Parents were removed after four days. Male progenies were collected after the flies emerged. They were kept on single yeast food at 20°C for 7-10 days. Flies were then frozen in liquid nitrogen and stored at -90°C for up to 1-7 days. Control flies *AC8* and *AC5* were treated in exactly the same way as the experimental lines.

Larvae were raised from mass cultures. 30 pairs of flies were allowed to lay eggs on apple food for 8 hours at 20°C. Eggs were collected and transferred to the riboflavin high protein food in a vial. About 100 eggs were placed in each vial and incubated at 25°C. Larvae were collected at 72-80 hours and 88-96 hours by washing them out from the food using 20% sucrose and the larvae were then classified under the microscope. Larvae were frozen in liquid nitrogen and kept at -90°C for assay and total nucleic acid extraction.

2.3.2.2 Preparation of assay extracts

10 male adults or larvae were weighed and homogenised with a glass grinder on ice in cold 100mM sodium phosphate buffer (pH 7.5) to a final concentration of 10mg/ml fly weight. The homogenates were centrifuged at 12,000 rpm (Sorvall SM-24) for 30 minutes. The supernatant was kept on ice until assayed.

2.3.2.3 Measurement of assay reaction

ADH activity was measured in a 1ml reaction mixture of 150mM isopropanol and 2mM NAD^+ in 100mM sodium phosphate buffer (pH7.5) with 10-100μl extract depending on the line to be assayed. Two replicates were assayed for each extract. The reaction was monitored at 340nm to record NADH production for 5 minutes on a Gilford 250 spectrophotometer. One unit of activity was defined as an increase in absorbency at 340nm of 0.001/min at 25°C (i.e. 1.61×10^4 nmoles NADH produced /minute). Enzyme activity was expressed as unit per milligram fly weight and was termed activity unit.

2.4 Cloned *Adh* gene used for detection of *Adh* sequence

The *p*SAF-2 plasmid contains an 11.8 kb *Sac*I fragment from Adh^F allele (Fig. 2.1) which was sub-cloned into a *pBR*322-derived *Sac*I vector (*pSV2*)

(Goldberg, 1980). Plasmid DNA used as a probe was provided by Slawek Bartoszewski in the concentration of 100ng/µl.

The *p*SAC-1 plasmid contains a 4.7 kb *Eco*RI fragment which was originally derived from *Canton-S* and sub-cloned into *pBR322* (Fig. 2.1; Goldberg, 1980). Plasmid DNA used for the probe was provided by Ann V. Wilks.

BSBH6 is a plasmid which carries a 600 bp fragment of *Bam*HI - *Hin*dIII (Fig. 2.2) cloned into *M13mp8*. The entire *Eco*RI /*Hin*dIII polylinker region containing the fragment was sub-cloned into pBluescript (Lockett, T. J., personal communication). This probe was used as a common probe both for adult and larval mRNA. The DNA was provided by Ann V. Wilks.

BSSBN4 is a plasmid containing a 386 bp *Sal*I/ *Hpa*I fragment that covers the 5' noncoding exon of the distal transcripts from -63 to +326 (Fig. 2.2, Lockett, T. J., personal communication). It was cloned into *M13mp8* .

The *pGPDH* (-5.3/7.70) plasmid, provided by Slawek Bartoszewski, has the non-duplicated *Gpdh* gene cloned into the *pUC8* plasmid. It was used as a *Gpdh* specific probe in Northern analyses.

The *pπ25.1* plasmid contains a *Bam*HI fragment from *λπ25* which was sub-cloned into *pBR322* (O'Hare and Rubin, 1983). It was used as a P-element specific probe in Southern analyses. Probe DNA was provided by Dr. Anna Ladvishenko.

pDmras is a plasmid which contains *ras* sequences in a 3.5 kb *Pst*I fragment (Mozer *et al*, 1985). It was used as a control probe for Northern analyses. The DNA was provided by Ann V. Wilks.

2.5 Genomic DNA extraction

2.5.1 Total DNA extraction

Genomic DNA was extracted according to Chia's method (Chia *et al*, 1985a). Flies were starved for 2 hours at 20°C to de-yeast them before they

were frozen in liquid nitrogen. About 50 adult flies were ground with a glass grinder in 0.35ml of 10mM Tris-HCl (pH7.5), 60mM NaCl, 10mM EDTA, 0.15mM spermine, 0.15mM spermidine, 5%(w/v) sucrose in a 1.5ml eppendorf tube. An equal volume (0.35ml) of 0.2M Tris-HCl (pH9.0) , 30mM EDTA, 2%(w/v) sodium dodecyl sulphate, 200μg/ml proteinase K, 5%(w/v) sucrose was added. The mixture was incubated at 37°C for 2-3 hours and then was extracted with an equal volume of neutralised phenol and centrifuged at 12K rpm for 5 minute in an eppendorf bench centrifuge. The aqueous phase (600ml) was re-extracted with an equal volume of phenol: chloroform (1:1). Subsequently 500ml aqueous phase was precipitated by two volumes of chilled ethanol with one-tenth volume 3M sodium acetate (pH5.2) at -90°C for 30 minutes or -20°C overnight. The DNA pellet was collected after centrifuging at 15K rpm for 15 minutes at 4°C in an eppendorf centrifuge, then washed in 70% ethanol and dried under vacuum for 10 minute. The DNA was re-suspended in 100μl TE buffer (10mM Tris-HCl, pH8.0, 1mM EDTA) and treated with 1μl RNase (10mg/ml) at 37°C for 30 minutes.

2.5.2 Single fly DNA extraction

Single fly DNA was extracted according to Jowett's methods (Roberts, 1986). One frozen fly was ground in 100μl of 10mM Tris-HCl (pH7.5), 60mM NaCl, 50mM EDTA, 0.15mM spermine, 0.15mM spermidine, an equal volume of 1.25% SDS, 0.3M Tris-HCl (pH9.0), 0.1M EDTA, 5% sucrose, 0.75% diethylpyrocarbonate (DEPC) was added and mixed well before incubating at 60°C for 30-40 minutes. 30μl 8M potassium acetate was added to the mixture and allowed to cool on ice for 45 minutes, then spun in an eppendorf centrifuge for 1 minute. The supernatant was transferred and mixed with 2 volumes chilled ethanol at room temperature for 15 minutes. The DNA pellet was centrifuged and washed with 70% ethanol, dried under vacuum and re-

suspended in 20µl of TE, treated with RNase to a concentration of 100ng/µl at 37°C for 15 minutes.

2.6 Gel electrophoresis

2.6.1 Restriction enzyme digestion of DNA and separation of DNA fragments on agarose gels

0.2-1µg DNA was digested with an enzyme with 10x digestion buffer. Single enzyme digests were done in the 10x digestion buffer which was provided with the enzyme (Boehringer Mannheim, Amersham). Double enzymes digests were usually carried out in either high, medium or low buffer as recommended (Sambrook *et al*, 1989). The reactions were carried out at the recommended temperature for 1-2 hours and stopped by mixing with one-fifth volume of sample loading dye (30% sucrose, 0.01% bromophenol blue, 50mM EDTA) or incubating at 65°C for 10 minutes.

The DNA fragments were separated on the basis of size by electrophoresis on an agarose gel. The concentration of the agarose gel varied between 1%-2%. TBE buffer (89mM Tris-HCl, 89mM boric acid, 25mM EDTA, pH8.3) was used for quick electrophoresis in a mini-gel tank (Bio-Rad) at 80v, 75mA for 1-2 hour. TAE buffer (0.04M Tris-HCl, pH7.8; 5mM sodium acetate; 1mM EDTA) was used for slow electrophoresis at 20mA overnight when the gel was run for Southern blotting or 50mA for 3-4 hour for isolating DNA fragments.

The DNA was visualised by adding ethidium bromide (0.01%) both to the gel and running buffer, because DNA is fluorescent under the ultraviolet light emitted from a transilluminator (305nm). The results were recorded on Polaroid Type 55 film.

A DNA size marker was run on each gel. Two sets of markers were used: the first one is phage *lambda* DNA (Boehringer Mannheim) digested with *Hin*dIII, the second one is *Spp-1* DNA (Bresatec) digested with *Eco*RI.

2.6.2 Isolation of DNA fragments using the DNA purification kit (Progenius)

The required DNA fragment was isolated from a TAE gel and the piece of gel was sliced into small pieces. Three volumes of NaI stock solution was added and incubated at 50°C for few minutes until the gel dissolved. 5µl of silica suspension was added to bind the DNA and mixed with the solution, incubated at room temperature for 5 minutes with frequent gentle mixing. The silica beads to which DNA had bound were sedimented after 30 seconds spinning in a bench eppendorf centrifuge and they were washed twice with 600µl ethanol wash solution. DNA was eluted into 10µl sterile distilled water at 55°C after centrifuging at full speed for 30 seconds.

2.7 Hybridisation of genomic DNA

2.7.1 Southern blotting and hybridisation

DNA was transferred from a 1% agarose gel to nitrocellulose membranes basically according to Sambrook *et al* (1989). The gel was soaked twice in Blot 1 (0.8M NaCl, 0.4M NaOH) for 30 minutes with general shaking, then in Blot 2 solution (1.5M NaCl, 0.5M Tris-HCl, pH7.4) for 30 minutes and transferred to a nitrocellulose membrane by 20x SSC (2.7M sodium chloride, 0.27M sodium citrate) for 14-16 hour at room temperature. When the filters had been blotted dry, they were baked at 80°C in a vacuum oven for 2 hours.

Baked filters were soaked in 2x SSC before being pre-hybridised in 25ml solution of 10x Denhardt's solution (Denhardt, 1966: 3x SSC, 0.02% bovine serum albumin, 0.02% Ficoll 400, 0.02% polyvinylpyrrolidone) with 30µg/ml of heat denatured herring sperm DNA, 0.1% SDS and 0.1% sodium pyrophosphate at 65°C for 1 hour. Denatured probes were added to the prehybridisation solution. Hybridisations were done in the hybridisation oven at 65°C for 14- 16 hour. Soaked filters were rolled with a sheet of mesh and inserted into a glass bottle and unrolled by rotating the bottle in the opposite direction. Hybridised filters were washed in 2x SSC, 0.1% SDS and 0.1%

sodium pyrophosphate solution three times for 35 minutes each at 65°C. After washing, the filters were exposed to the PhosphorImage cassette at room temperature or to a sheet of X-ray film (Kodak XRP-1) in a light proof cassette sandwiched between intensifying screens at -70°C.

2.7.2 Labelling of nick translated DNA probe

Double-stranded DNA probes used to hybridise to the Southern and Northern blots were made by nick translation (Sambrook *et al,* 1989). 10µl (~0.5µg) probe DNA, 5µl nick translation buffer (0.05M Tris-HCl, pH7.2; 0.01M $MgSO_4$; 1mM dithiothreitol (DTT); 0.05mg/ml BSA), 4µl unlabelled dNTPs (dATP, dGTP, dTTP mixed in 20mM each), 1µl DNA polymerase I (10U, Bresatec) and 2µl DNase (0.1µg/ml) were mixed and made up to 47µl with sterile distilled water. Finally 3-4µl labelled α^{32}P-dCTP (30-40µCi) (Amersham, Dupond) was added. The labelling mixture was incubated at 14°C for 2 hours to allow incorporation of the labelled nucleotides, then at 65°C for 10 minutes to inactive the enzymes. To remove any unincorporated nucleotides, the probe DNA was passed through a Sephadex (G-50) (Sigma) column by spinning at 2000 rpm for 5 minute in a bench centrifuge. 10µl sample of labelled probe DNA was checked for the amount of incorporation by a Geiger counter (the reading was usually in the range 400-550 cpm). The probe was boiled for 5 minutes to make the DNA single stranded and left on ice for a further 5 minutes before adding it to the pre-hybridisation mixture.

2.7.3 Restriction fragment mapping using Southern blots

The sizes of the bands observed in Southern blots were determined graphically from a semi-logarithmic plot of standard marker DNA fragment (kilo bases) versus relative mobility (centimetres). By analysing the fragments produced both by single and double digests (simultaneous digestion of the one extract with two enzymes), it was possible to identify the insertion of extra

DNA sequence and changes in the position of restriction endonuclease recognition sites.

2.8 The polymerase chain reaction

2.8.1 Preparation of oligonucleotide primers

The primers used in PCR and for sequencing were prepared by the Molecular Resource Facility, Centre for Molecular Structure and Function, The Australian National University on an Applied Biosystems 380B DNA synthesiser. The primers were obtained in about 3.5ml of a concentrated ammonia solution. The solution was aliquoted to 130µl each in eppendorf tubes, extracted with 1.3ml n-butanol. After vortexing for 15-20 seconds, the primer mix was spun on a bench centrifuge for 1 minute and then the supernatant was tipped off and the pellet was dried under vacuum for 20 minutes. Primers were stored dry at -20°C. One of the aliquots was re-suspended in 100µl TE to use as a working sample. The concentration of the primer was measured at $O.D_{260}$. Primers for sequencing and PCR were used at a concentration of 1 pmol/µl and 20 pmol/µl respectively.

2.8.2 Amplification of DNA fragments

Double-stranded DNA template was added to a mixture of a pair of primers (0.1 nmole each), four deoxynucleotide triphosphates (200µM each dNTPs, Boehringer Mannheim), 10µl 10x reaction buffer (500mM KCl; 100mM Tris-HCl, pH8.3; 15mM $MgCl_2$; 0.1%w/v gelatine), *Taq* polymerase (2.5U) and sterile water to make final volume of 100µl. Mineral oil was over-layered to prevent evaporation . The reaction mixture was heated to 94°C for 30 seconds to denature the template, then chilled to 60°C for 1 minute to allow the primers to anneal to the template, and the temperature was increased to 72°C for 1 minute to allow the polymerase to extend from the primer along the template. An extension step at 72°C for 7 minutes was added at the end of the 35 cycles.

Other parameters (annealing time and extension time) were changed in order to amplify large size fragments (see chapter 4). A DNA Thermal Cycler (Perkin-Elmer Cetus) was used to control the variation in time and temperature. When the PCR was completed, the DNA samples were cooled to room temperature before being removed from under the oil layer using a pulled pasteur pipette. Each sample was electrophoresised on a 1% agarose gel to determine the relative yield and size of DNA.

2.8.3 Inverse PCR

The normal PCR method requires the use of primers complementary to opposite strands of DNA and orientated so that primer extension proceeds 5' to 3' from each primer towards the other primer. When the sequence is known only in one region, but one or both sides of this region need to be amplified, it fails to meet the requirement of primers. Inverse PCR provides a simple solution to this problem: a restriction fragment containing the region of interest is circularised, and then primers are orientated so that primer extension proceeds 'outward' from the region of known sequence to amplify the flanking DNA (Fig 2.3; McPherson, 1991). Genomic DNA was digested with an appropriate restriction enzyme which generated a fragment between 300 bp - 2 kb that would be suitable for circularisation and amplification. Digested DNA was extracted by phenol: chloroform (1:1), precipitated by ethanol and re-suspended in TE buffer. DNA was diluted to 10μg/ml in 20μl ligase buffer, 1 unit T4 DNA ligase was added and incubated at 14°C overnight to allow ligation. The ligation mixture was incubated at 65°C to inactivate ligase and then digested with another appropriate restriction enzyme to make it linear. The restriction enzyme was inactivated at 65°C for 10 minutes. Linearized DNA was used for PCR.

2.9 Sub-cloning a PCR fragment into pBluescript

2.9.1 Digestion of insert and vector DNA

1μg aliquot of pBluescript vector DNA was digested with appropriate enzymes and purified by phenol : chloroform extraction. PCR fragment DNA (~200ng) was digested with a suitable enzyme to give compatible cohesive ends, or was treated with T4 DNA polymerase (1U) in T4 DNA polymerase buffer [(0.33M Tris acetate, pH8.0; 0.66M potassium acetate; 0.1M magnesium acetate; 5mM dithiothreitol; 1mg/ml bovine serum albumin (Fraction V, Sigma)] and 2mM dNTPs mix to give blunt ends suitable for ligation. The reaction was carried out at 37°C for 15 minutes then inactivated at 65°C for 10 minutes. The insert DNA was purified by the DNA purification kit (see 2.5.2).

2.9.2 Ligation

2.9.2.1 'sticky end' ligation with double enzyme digestion

Double enzyme digested vector DNA 1μl (~200ng) was mixed with 5μl (200ng-1μg) insert DNA, 1μl 10mM ATP, 1μl 10x T4 ligation buffer (Boehringer, 500mM Tris-HCl, pH8.0; 70mM $MgCl_2$; 10mM DTT), 0.5μl T4 DNA ligase (1.5 unit) and sterile water to a final volume of 10μl. A control ligation sample was prepared in which the insert DNA was left out to test the efficiency of self-ligation of the digested vector. The ligation mix was incubated at 14°C for 14-16 hours.

2.9.2.2 'blunt end' ligation

1μl *Eco*RV digested vector DNA and 7μl blunt ended PCR amplified fragment DNA were mixed with 1μl 10x T4 ligation buffer and 1μl T4 DNA ligase (3U). The mixture was incubated at room temperature (23°C) for 1-2 hours and then at 18-20°C for 14-16 hours. Then the ligation samples were ready for transformation.

2.9.3 Making competent bacteria

Bacterial stock *DH5α* was kept in a glycerol preparation at -90°C. One drop of such preparation was dilution streaked onto an LB plate (1% Bacto-tryptone, 0.5% Bacto-yeast extract, 1% NaCl, pH7.5, 1.8% agar for solid media) and incubated at 37°C overnight. A single colony was transferred into 3ml liquid LB media and incubated overnight at 37°C in a shaker. Next day the bacterial culture was diluted 1:100 into fresh LB media and incubated in the shaker at 37°C. When the $O.D_{600}$ reading was between 0.3-0.6 (usually less than 2 hours) the cells were in the required exponential growth phase. The bacteria were immediately chilled on ice and centrifuged for 2 minutes. The pellet was re-suspended in 0.5 volume of cold 50mM $CaCl_2$, incubated on ice for 30 minutes and centrifuged again for 2 minutes. Finally the bacteria were suspended in 0.01 volume of 50mM $CaCl_2$ and 15% glycerol and stored at -90°C until used.

2.9.4 Transformation

The competent cells were thawed on ice, mixed with the ligation mixture and kept on ice for 30 minutes. Then the mixture was heat shocked at 42°C for 5 minutes and incubated with 1ml LB media at 37°C for 45 minutes to 1 hour. The transformed bacteria were spread onto LB plates which contained ampicillin (50μg/ml) and X-gal (32μg/ml) and incubated at 37°C overnight. White colonies were picked-out for plasmid DNA preparation.

2.10 Plasmid DNA preparation (boiling mini plasmid preparation)

A single bacterial colony was inoculated into 3ml LB medium with ampicillin (50μg/ml) and incubated overnight in a 37°C shaker. The following morning the bacterial culture was chilled on ice and spun in an eppendorf tube at 4°C for 3 minutes. The pellet was re-suspended in 220μl STETL (8% sucrose, 5% triton X-100, 50mM EDTA, 0.5mg/ml lysozyme), incubated at room

temperature for several minutes and boiled in a water bath for 30 seconds. The cell debris were removed and 220μl isopropanol was added to the supernatant and centrifuged again for 15 minutes. The DNA pellet was washed with 70% ethanol and re-suspended in 10μl sterile water and incubated with 1μl RNase solution (10mg/ml) at 37°C for 20 minutes. The DNA sample was suitable for restriction enzyme digestion at this stage. Two more extractions with phenol: chloroform (1:1) and precipitation with 2.5 volume cold ethanol at -90°C for 30 minutes were carried when the DNA samples were being prepared for sequencing. The DNA pellets were washed with 70% ethanol, dried under vacuum and re-suspended in 20μl sterile water.

2.11 DNA sequencing

2.11.1 Sequencing of double-stranded plasmid DNA

The sequence reactions were performed using the Sequenase DNA Sequencing Kit (United States Biochemical). 2-3μg of plasmid DNA from one mini-prep was diluted to 19μl with sterile water. DNA was denatured with 1μl of freshly made 4M NaOH at 37°C for 10 minutes and precipitated by adding 8μl 5M ammonium acetate (pH4.8) and 100μl cold ethanol, and leaving it at -90°C for 20 minutes and centrifuging at 4°C for 15 minutes. The pellet was washed with 70% ethanol, dried under vacuum and re-suspended in 1μl of appropriate primer (1 pmol/μl), 2μl of 5x sequence reaction buffer and 7μl sterile water. Once the pellet was dissolved, the sample was incubated at 60°C for 15 minutes followed by 15 minutes at 37°C and then transferred to room temperature. 1μl 0.1M DTT, 2μl diluted labelling mix, 1μl label (α-^{35}S dATP, 10μCi/μl) and 2μl dilute Sequenase were added to the annealing mix and incubated at room temperature 2-5 minutes. Then 3.5μl of mixture was transferred to each of four tubes containing 2.5μl of each termination mixture (A, C, T, G) and incubated at 42°C for 5 minutes. The reaction was stopped by adding 4μl loading dye.

2.11.2 Direct sequencing of PCR amplified DNA

Samples of PCR amplified DNA products were electrophoresed on an agarose gel to determine the yield and purity (ie: whether any non specific fragments had been amplified). Then the remainder of the product was extracted by phenol : chloroform (1:1) to remove any of the unused *Taq* polymerase, 0.1 volume of 3M sodium acetate and 2.5 volume chilled ethanol were added to the DNA and the mixture was left to precipitate at -90°C for 20 minutes. The sample was spun at 4°C in an eppendorf centrifuge for 15 minutes to obtain a DNA pellet which was subsequently washed with 70% ethanol, dried under vacuum and finally dissolved in 20µl TE. This DNA was applied to a Sepharose column (Linker-6, Boehringer Mannheim) and spun at 1100g as specified by the manufacturers for 10 minutes. DNA collected from the column was usually between 40-50µl and used for sequencing without any further treatment.

3µl of appropriate primer (1pmol/µl), 2µl of 5x reaction buffer and 5µl purified DNA were made as an annealing mix. The mix was boiled for 3 minutes and then snap cooled on dry ice for at least 5 minutes. The sample was thawed and mixed with 1µl DTT, 2µl dilute labelling mix, 1µl label (α-^{35}S dATP, 10µCi/µl) and 2µl diluted Sequenase at room temperature for 45 seconds. After that 3.5µl of the mix was transferred to each of four tubes containing 2.5µl of each termination mixture (A, C, T, G) and incubated at 37°C for 5 minutes. The reaction was stopped by adding 4µl loading dye.

2.11.3 Running and reading the sequencing gel

The electrophoresis was performed in a Bio-Rad Sequi-Gen apparatus. The gel plates were washed thoroughly with pyroneg, rinsed with distilled water and then wiped thoroughly with ethanol. The back plate was siliconized. The spacer and slot former were washed with distilled water and ethanol. 200ml gel solution [7M urea, 1x TBE (89mM Tris-HCl, 89mM boric acid, 25mM

EDTA, pH8.3), 5% polyacrylamide mix] were filtered through Whatman paper and degassed for 15 minutes. 210μl (250mg/ml) ammonium persulphate and 150μl TEMED (Bio-Rad) were used for catalysis to quickly polymerise 30ml of gel solution which was used to seal between the base of the plates. Another 221μl ammonium persulphate and 170μl TEMED were added to the remaining 170ml of gel solution. They were mixed well and poured between the plates. The slot former was inserted and the gel was allowed to set for 90 minutes. Then the slot former was removed and the slots were washed and saturated with 1xTBE until the gel was used. 1800ml 1xTBE running buffer were added to the back and bottom chambers. The gel was pre-run for 1 hour at 125 watts to reach a temperature of about 50°C. The reaction mixes were heated for 2 minutes at 80°C before 3μl was loading into each slot. The samples were loaded in the order of GATCG to facilitate reading the gel. The gel was run at 100w for 1 hour 45 minutes (short run) or 3.5 hours (long run).

2.12 RNA analysis

2.12.1 Total RNA extraction

For all RNA work, glassware was baked at 240°C overnight and solutions were autoclaved. 0.1% DEPC treated water was used where possible. Equipment unable to be autoclaved was thoroughly wiped with ethanol to prevent contamination by nucleases.

Approximately 50 adult flies (4-8 days old) were starved for 2 hours to remove yeast. Then they were snap frozen in liquid nitrogen and stored at -90°C until used. For larvae, approximately 50mg of properly aged larvae were collected and snap frozen in liquid nitrogen. The material was homogenised in 500μl of grinding buffer (100mM Tris-HCl, pH7.5; 10mM EDTA; 350mM NaCl; 2% SDS; 7M urea). The homogenate was extracted twice with 500μl phenol : chloroform (1:1). The final aqueous phase was precipitated by adding 2 volumes of chilled ethanol and leaving at -90°C for 30 minutes. The RNA was

pelleted by spinning at 4°C in an eppendorf centrifuge, re-suspended in 200µl sterile water and re-precipitated overnight at -20°C. Finally the nucleic acid was re-suspended in 50µl sterile water.

2.12.2 Northern blotting

5µl (about 10 µg) of total RNA was mixed with 4µl di-ionised glyoxal, 10µl de-ionised DMSO and 2µl sodium phosphate buffer (100mM, pH7.0) and incubated at 50°C for 1 hour. The glyoxal and dimethylsulfoxid (DMSO) were di-ionised freshly before use by passing them through a mixed bed resin (Bio-Rad, AG501-X8D). 2µl of loading dye (50% glycerol; 5% 50mM $NaPO_4$, pH7.0; 0.09% bromo-phenol blue) was added to each sample before loading onto a 1% agarose gel (10mM sodium phosphate buffer, pH7.0). 1600ml 10mM phosphate buffer was poured into the tank and the buffer was pumped from the cathode (-) side to the anode (+) side at the flow rate of 400ml per hour. The gel was run overnight at 40mA, 35V and then the RNA was transferred to a nylon membrane by blotting using 20x SSC (Sambrook *et al*, 1989). Next morning the filter was fixed by placing it on the UV transilluminator for 5 minutes and then baking the filter in an 80°C oven for 2 hours.

2.12.3 PhosphorImage quantification

Hybridised filters were exposed to the PhosphorImage cassette at room temperature over one night for quantification. Hybridisation signals stored on the phosphor screen were scanned on the PhosphorImager instrument (Molecular Dynamics). The PhosphorImage signals are linear with the amount of radiation in the bands in a large range. Each band observed was measured separately for its signal intensity. The relative intensity between the normal sized-*Adh* band and the *nLA248* band was calculated for each sample. The average value of relative intensity for 3 to 5 repeat samples is shown in the tables in chapter 3. The relative RNA level between the variant and the control

allele are derived from the value of relative intensity between normal sized-*Adh* band and the *nLA248* band.

2.12.4 Removal of probes from nylon membranes

Filters were hybridised with a nick translated probe (see 2.7.1). When the filter had to be reprobed, it was washed twice at 75°C for one hour in a solution of 1mM Tris-HCl (pH8.0), 1mM EDTA (pH8.0), 0.1x Denhardt's solution, and then rinsed in 0.01x SSC, blotted dry and placed with a PhosphorImage cassette for exposure to check if all the probe was removed.

2.13 Deleting the KP element

2.13.1 Breeding program

sbΔ2-3/Tm6 was used as a strong P strain. For convenience, *Lobe* was used as a dominant mark on the second chromosome. *L/+, sbΔ2-3/+* male flies were crossed with the M strain *AD369/CyO* virgin females (Fig.2.4). In the next generation, *L/AD369, sbΔ2-3/+* male flies were selected and crossed to virgin females homozygous for the *Adh* null allele Adh^{nLA248}. *AD369/nLA248* flies were picked out from the progeny and tested for increased ADH activity.

2.13.2 Selecting flies with normal ADH levels by exposure to ethanol vapour

In a 4 inch vial, a cotton wool disc (~7mm thick) which is made from half of a large size cotton ball (Smith Nephew), was pushed to the bottom and soaked with 1ml 8% ethanol, 3% sucrose. Tubes were made air-tight with Nescofilm to retain moisture (Fig 2.5). 15-20 male flies (5-10 days old) were exposed to ethanol vapour in each vial at 25°C for 24 hours. All surviving flies were rescued onto single yeast food and kept at room temperature for a few days. They were used to set up single crosses with homozygous *nLA248* virgin females to preserve the genotype and were then used in single fly PCR for further molecular analyses.

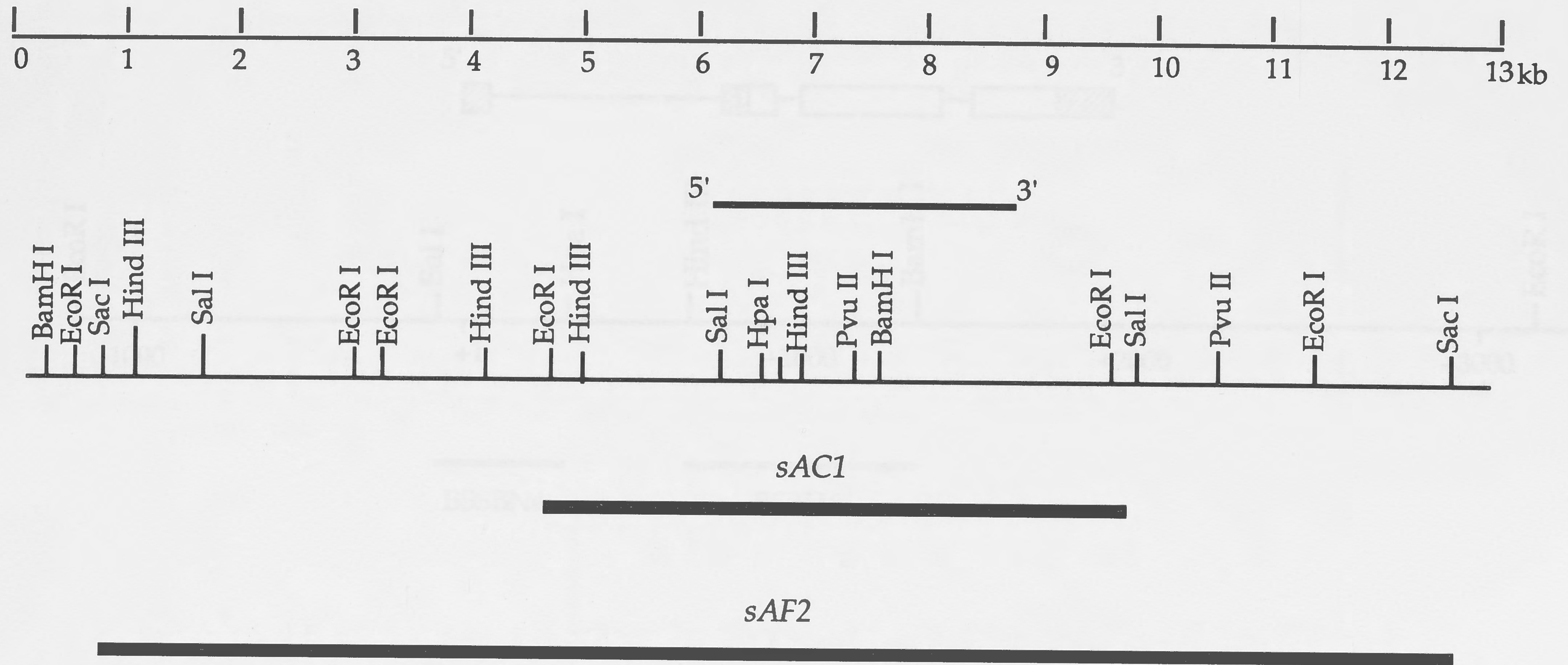


Figure 2.1 Restriction map of the *Adh* gene and surounding region.

The thin bar above the restriction map shows the position of the *Adh* gene and the direction of its transcription.

The thick bars below the map indicate the probes which were used to detect the *Adh* gene. *sAC1* contained a 4.7-kb *Eco*RI fragment. *sAF2* contained an 11.8-kb *Sac*I fragment.

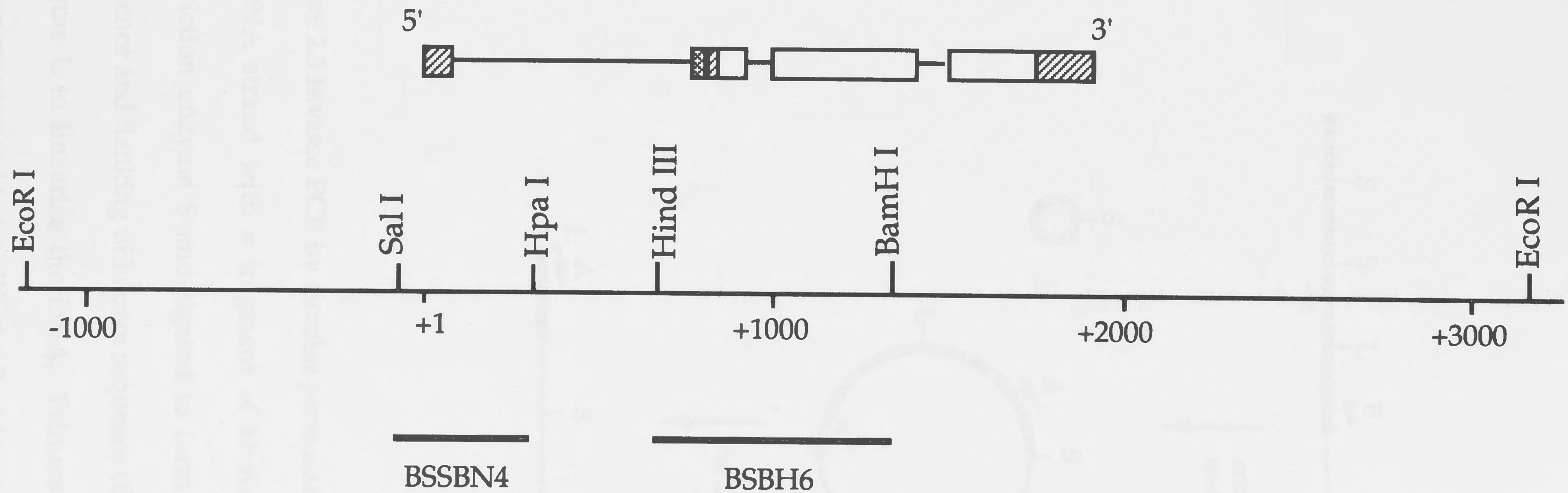


Figure 2.2 Detailed restriction map of the region surounding the *Adh* gene.

The distal transcriptional start site is numbered +1. The structure of the *Adh* gene of *D. melanogaster* is shown above the map. Bars below the restriction map represent the two probes used for Northern analyses: BSSBN4 and BSBH6.

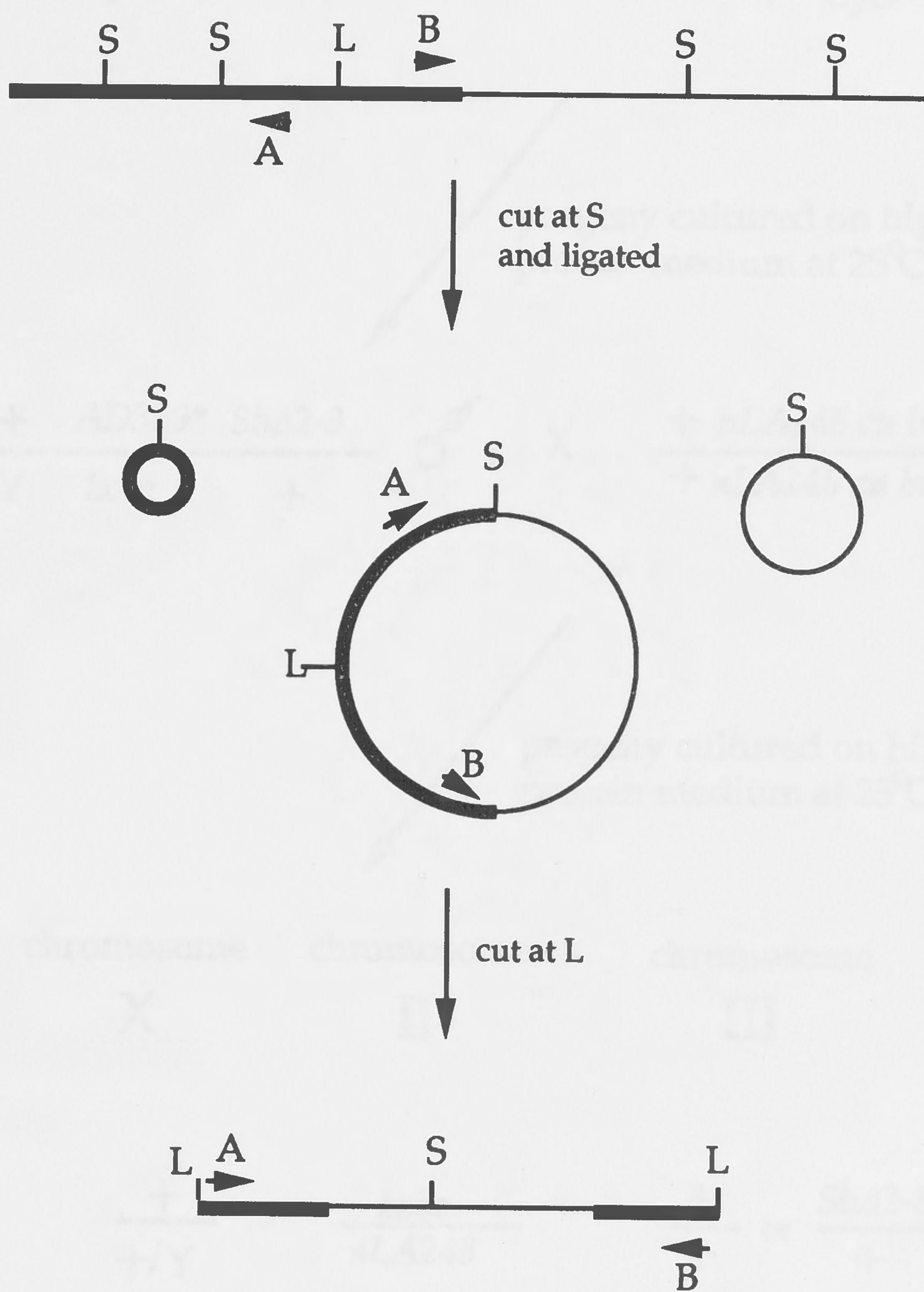


Figure 2.3 Inverse PCR by circular permutation.

A DNA strand with a fragment of known sequence (thick line) is cut by restriction enzyme S and ligated to form a circle. The circle with known sequence and flanking unknown sequence (thin line) is cut by another restriction enzyme L to linearize the DNA. Primers A and B matching to the known sequences are used to amplify the flanking unknown sequence.

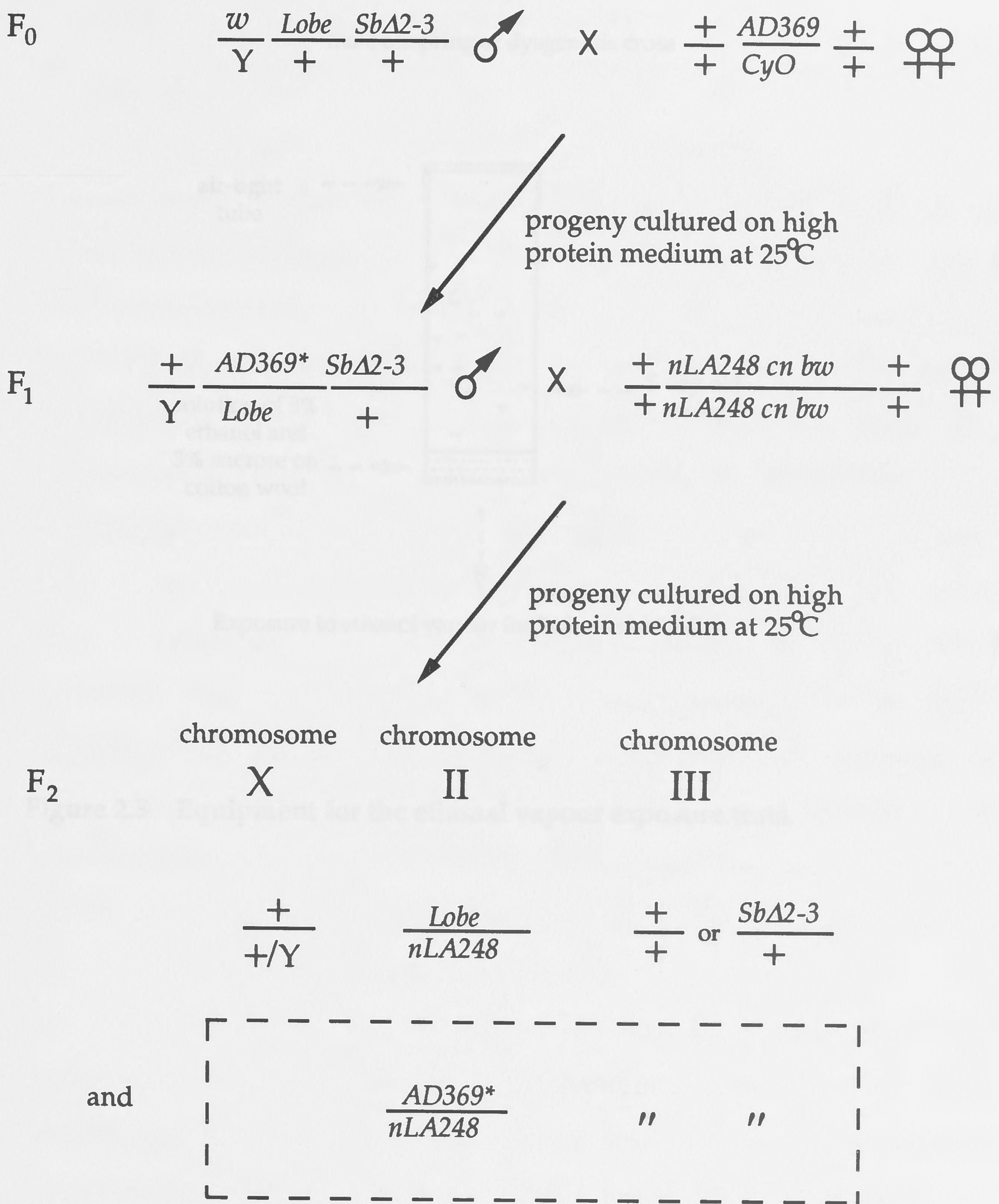


Figure 2.4 Breeding programe for deleting the KP element

Genotypes of flies crossed in each generation are shown. Flies whose genotype is indicated in the dashed box in the F_2 generation were collected for further analyses.

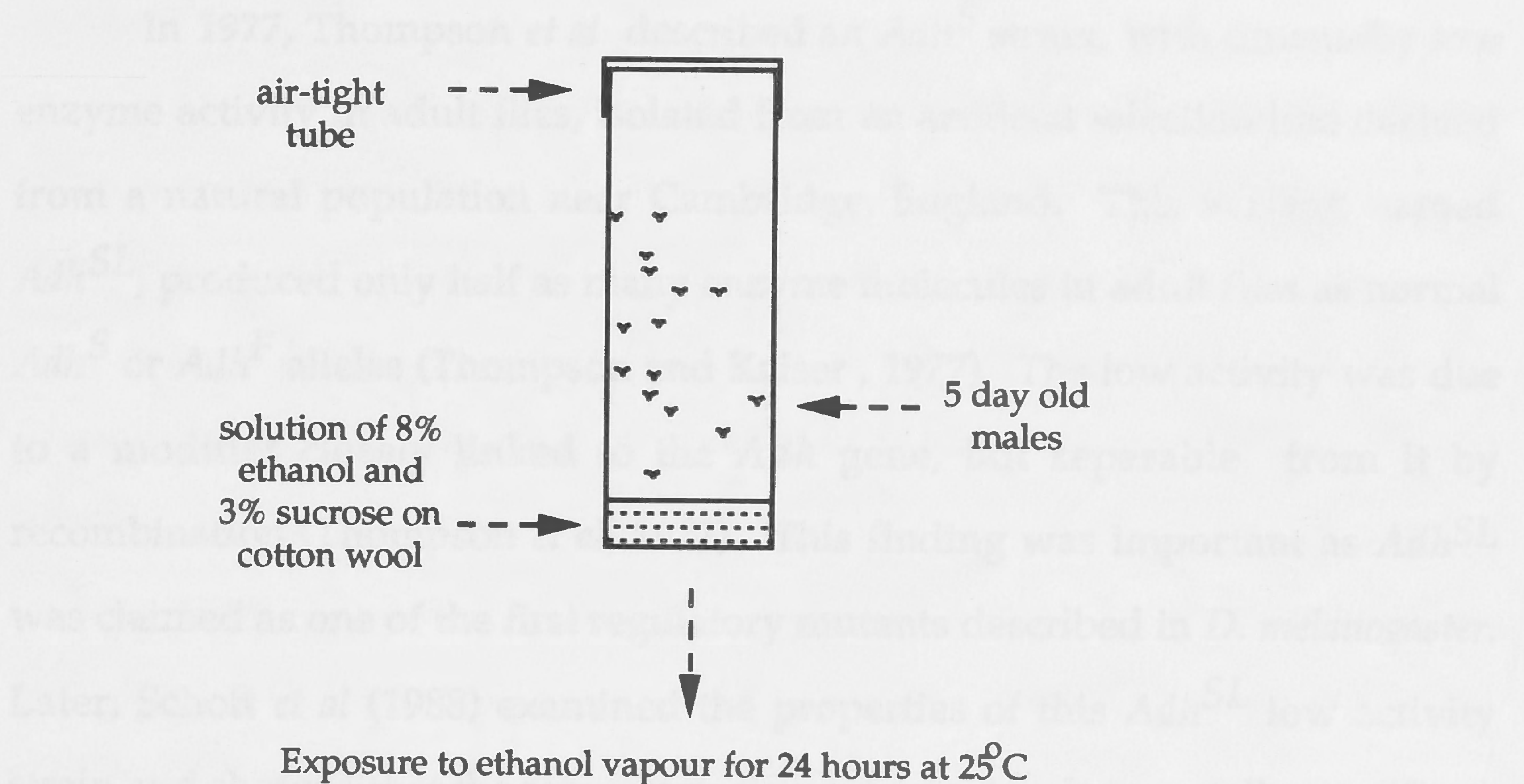


Figure 2.5 Equipment for the ethonal vapour exposure tests.

Chapter 3 The biochemical and molecular phenotypes of the three *Adh* variants

3.1 Introduction

In 1977, Thompson *et al* described an Adh^S strain, with unusually low enzyme activity in adult flies, isolated from an artificial selection line derived from a natural population near Cambridge, England. This variant, named Adh^{SL}, produced only half as many enzyme molecules in adult flies as normal Adh^S or Adh^F alleles (Thompson and Kaiser , 1977). The low activity was due to a modifier closely linked to the *Adh* gene, but separable from it by recombination (Thompson *et al*, 1977). This finding was important as Adh^{SL} was claimed as one of the first regulatory mutants described in *D. melanogaster*. Later, Schott *et al* (1988) examined the properties of this Adh^{SL} low activity strain and showed that the regulatory element was developmentally specific as expression from both the proximal and distal promoters was minimally affected in third instar larvae, but expression from the distal promoter was much more affected in adults. They found that Adh^{SL} carried a 4.5 kb insert approximately 3 kb 5' upstream to the distal promoter. They did not prove that the insertion was responsible for the regulatory phenotype of Adh^{SL} as they did not sequence the gene or the 5' or 3' regions.

In a survey of Adh^{null} alleles in Australian populations of *D. melanogaster*, Freeth and Gibson (1985 and unpublished data) found two alleles, *AHA51* and *AD369*, which initially were both classified as null activity variants. The allele *AHA51* was detected as a low activity Adh^S allele in a survey of the Coffs Harbour (NSW) population in 1984. The allele was identified by the null allele assay technique of Voelker *et al* (1980) as flies heterozygous for Adh^F and that allele produced no heterodimer bands after electrophoresis and staining. The second chromosome bearing this allele was isolated by crossing male flies to $T(2{:}3)ap^{Xa}/CyO;TM6$ (Lindsley and Zimm,

1992) and the allele was maintained as a balanced stock with *CyO*. Preliminary electrophoretic studies indicated that although *AHA51/CyO* adults produced no heterodimer bands of activity, these bands were present in 3rd instar larvae. When first isolated, most female flies in the stock homozygous for *AHA51* were sterile.

The Adh^F allele *AD369* was isolated in 1987 from a population at All Saints winery near Rutherglen (VIC) about 900 km SSW of the Coffs Harbour site. A large number of single female lines were set up and in the progeny of one of these females there were flies with an unusual phenotype — no heterodimer bands in adults flies expected to be heterozygous Adh^F/Adh^S. The second chromosome bearing this unusual Adh^F allele was isolated by crossing male progeny to $T(2{:}3)ap^{Xa}/CyO;TM6$. Flies homozygous for the unusual allele were sterile and crosses between *AHA51 /CyO* and *AD369/CyO* produced sterile wild type progeny. Again, preliminary electrophoretic experiments suggested a difference in phenotype between adults and third instar larvae of *AD369/CyO*.

In addition, Jiang and Gibson (1992a) described the restriction map of an *Adh* allele (*AAS13*), extracted in 1986 from the All Saints (VIC) population, which had a 5 kb insertion in the adult intron. This allele was originally named *AAS44* after the single female line from which it was isolated, and this nomenclature has been retained in my thesis. Jiang and Gibson(1992b) reported that this allele had relatively low ADH activity in adult flies, but they did not assay larvae. Because of the position of the insertion I decided to include it in my sample of *Adh* alleles for detailed analysis. The allele had been maintained as a heterozygote with *Df(2L)64j* and the chromosome had acquired recessive lethal mutations.

The first step in the analysis of these three alleles was to confirm the phenotypes of *AHA51* and *AD369* and to investigate the phenotype of *AAS44* by cellulose acetate electrophoresis. This chapter also describes quantitative

assays of ADH activity and the correlation with RNA levels in the three alleles. Some earlier unpublished data collected by Freeth and Gibson on *AHA51* are also presented.

3.2 Results

In all of my work, the normal activity alleles *AC5* (*Adh^F*) and *AC8* (*Adh^S*) were used as controls. They were isolated from the Cygnet (Tasmania) population in 1984 (Gibson *et al*, 1991).

3.2.1 Cellulose acetate electrophoresis of adult and 3rd instar larval extracts.

Cellulose acetate electrophoresis was first used to examine the adult phenotypes of the three variants. In order to make comparison, adult flies heterozygous for each variant and *Df(2L)64j* were produced for electrophoresis. These data (Fig 3.1) showed that the extract of *AD369* was less stained than that of the control *AC5* (ADH-F). The extract of *AAS44* was stained less strongly than *AC5* but heavier than *AD369*. The *AHA51* extract was stained more lightly than the extract of the normal activity control *AC8* (ADH-S).

To check the phenotypes, I had first to produce crosses in which larvae of the correct genotype could be identified. This is necessary because the *AD369* and *AHA51* homozygous flies are sterile, and the alleles are kept heterozygous with *CyO*. It was not possible to distinguish the homozygous wild-type larvae in these stocks. In order to identify the appropriate larvae, flies were made heterozygous for the variant alleles and the null allele *nLA248*. The second chromosome of the *nLA248* allele carries the markers *black(b) cinnabar(cn) brown(bw)* and is homozygous viable (Kelley *et al*, 1985; see 2.1.2). Males heterozygous for *nLA248* and *AC8* , or *AHA51*, or *AC5*, or *AD369* , or *AAS44* were crossed to *nLA248* homozygous virgin females. In the progeny of these crosses heterozygous larvae could be distinguished from the *nLA248* homozygous larvae because the pigment in the wild type eye of the

heterozygote caused the colour of the Malpighian tubules to be yellow (Brehme and Demerec, 1942). In the homozygous *nLA248* larvae, there is no colour in the Malpighian tubules as they are homozygous for *cn* and *bw*. By feeding the larvae riboflavin, the distinction between heterozygous and homozygous *nLA248* larvae can be enhanced (Ashburner, 1989; Fig 3.2). For comparison, the adult flies produced in these crosses were also subjected to cellulose acetate electrophoresis. These data (Fig 3.3) showed that the phenotypes of the adult flies were as expected from the previous experiments. They also showed that the 3rd instar larval extract of *AD369* was less stained than the control *AC5,* as was the extract of *AAS44*. The *AHA51* larval extract appeared to be stained as strongly as the *AC8* control extract. In both *AD369* and *AHA51* the larval extracts appeared to be more heavily stained than the adults.

3.2.2 ADH activity in adults and 3rd instar larvae

3.2.2.1 *AHA51*

Dr. Freeth and Prof. Gibson had done some preliminary experiments investigating the level of ADH activity and ADH CRM in adults and 3rd instar larvae of *AHA51*. They have allowed me to quote their unpublished results as theirs' are the only data available on ADH CRM levels in *AHA51*.

Freeth and Gibson (unpublished) compared ADH activity and ADH CRM in a variety of genotypes, including *AC8* homozygotes and *AHA51/Df(2L)A379* heterozygotes, *AC8* homozygotes and *AHA51/*Adh^{nB} heterozygotes. They measured ADH CRM as described by Lewis and Gibson (1978). Their data for adult flies are given in Table 3.1. The data from experiment 1 and 2 were similar but significantly different from that in experiment 3. Experiment 1 and 2 had lower ADH activity than experiment 3 for each genotype. The relative ADH activities obtained in each experiment were not significantly different. The results showed that ADH activity in *AHA51/*Adh^{nB} adults was about 15% of the control *AC8/*Adh^{nB}, and in

AHA51/Df(2L)A379 adults was about 19% of the *AC8/Df(2L)A379* control. The differences in ADH CRM paralleled the differences in activity. *AHA51/Adh*^{nB} had about 16% of the control level, whilst in *AHA51/Df(2L)A379*, it was about 20%.

Freeth and Gibson (unpublished) also assayed ADH activity and ADH CRM in 3rd instar larvae produced in crosses between *AHA51* homozygotes (some of which when first isolated were fertile) and *Df(2L)A379/CyO Adh*^{nB} heterozygotes. These larvae were a mixture of the genotypes of *AHA51/Df(2L)A379* and *AHA51/CyO Adh*^{nB}, but in either case only the *AHA51* allele contributed to ADH activity levels. These data (Table 3.2) showed that the *AHA51* larvae had about 71% activity of the normal level, a much higher relative level than in adults. The ADH CRM in larvae was as high as 61% of the normal activity allele, consistent with the enzyme activity level in larvae.

My assay data for adults homozygous for *AHA51* (Table 3.3) also showed that the ADH activity was much lower than that in the control *AC8*. *AHA51* had only about 16% of the normal activity. The activity of heterozygotes *AC8/AHA51* was not significantly different to the expected mid-parent level.

The activity of adult flies heterozygous for *AC8* and *nLA248* showed no significant difference to the expected mid-parent activity (Table 3.4), but in flies heterozygous for *AHA51* and *nLA248* the activity was significantly lower than the mid-parent — about half of the expected value. In these data the relative activity of *AHA51* to *AC8* was reduced to 8%. In 72-78hr larvae, the activity of *AHA51/nLA248* heterozygotes was very similar to that of the *AC8/nLA248* larvae. However in late 3rd instar (90-96hr), the activity of *AHA51/nLA248* was similar to the activity in early larvae but the activity of *AC8/nLA248* was increased. The relative activity was lower (57%) in the late 3rd instar larvae.

3.2.2.2 *AD369*

ADH assay data for *AD369* homozygotes are summarised in Table 3.5. Compared to the control *AC5,* adult flies homozygous for *AD369* had very low activity — 11% of the normal level. In *D369/Df(2L)64j* hemizygotes the activity was not significantly different to half the level in the *AD369* homozygotes. The activity of *AC5/Df(2L)64j* hemizygotes was also not significantly lower than the expected mid-parent level. In the heterozygotes *AC5/AD369,* the activity was not significantly different to the expected mid-parent level.

In heterozygotes with *nLA248*, a similar pattern of ADH activity differences was found (Table 3.6). *AD369/nLA248* had about 11% the activity of *AC5/nLA248*. In early 3rd instar larvae the activity of *AD369/nLA248* was 70% of that of the control *AC5/nLA248*. In late 3rd instar larvae, the activity of *AD369/nLA248* was about 66% of the control.

3.2.2.3 *AAS44*

Because of the recessive lethal gene on the second chromosome bearing *AAS44*, it was impossible to obtain homozygous viable flies. Heterozygotes with *nLA248* were assayed. These data (Table 3.7) showed that the activity of *AC5/nLA248* adults was similar to that obtained in previous assays (see Table 3.5). In *AAS44/nLA248* the activity was about 34% of the activity in control heterozygous adult flies. In early 3rd instar larvae the activity of *AAS44/nLA248* was about 75% of the control but was 89% in later 3rd instar. The activity profile for *AAS44* larvae was different to that of the other two alleles as it increased during the 3rd instar larvae whilst in *AD369* and *AHA51* activity was lower in late 3rd instar larvae.

3.2.3 ADH transcripts level in adults and larvae

3.2.3.1 Northern blotting for adult flies

To fully interpret the information obtained on the relative ADH activities in larvae and adults it was necessary to assay transcript levels in the

three strains. Total RNA extracted from adult flies of *AHA51* was examined on Northern blots and compared with the control *AC8* . The filters were probed with a nick-translated labelled *p*SAF-2 plasmid, which is an *Adh* specific probe (see Chapter 2.4). Compared with the *AC8* control, the *AHA51* adults showed very low amounts of *Adh* transcript, but the size was the same as the control (Fig 3.4 upper part). To check the RNA loading on the gel, the level of the *Gpdh*-specific transcript was assessed. To do this, the radioactivity was washed off the filter and the filter was re-hybridised with the nick-translated labelled *pGPDH*(-5.3/7.7), a *Gpdh* specific probe. However the data (Fig 3.4, middle part) showed that there was a lower *Gpdh* transcript level in *AHA51* than in *AC8*. The GPDH activity level in *AHA51* was then checked after electrophoresis on a cellulose acetate gel. This experiment (data not given) showed that *AHA51* had a low level of GPDH activity as well. Thus a *Gpdh* specific probe was not appropriate to assess RNA loadings. I then used *pDmras* as a probe. The data (Fig. 3.4 lower part) confirmed that *AHA51* and *AC8* had very similar RNA loadings. Thus the difference in the hybridisation signals using the *Adh*-specific probe is likely due to differences in transcript level between the two lines.

The *AD369* homozygotes and *AAS44/Df(2L)64j* heterozygotes were investigated by Northern blots probed with *p*SAF-2 (Fig 3.5). In comparison with the *AC5* control, both *AD369* and *AAS44* had a much lower level of hybridisation signals in adults (Fig 3.5, upper part). To check the RNA loadings the filter was re-probed with *pGPDH*. The data showed that the RNA loadings were greater for *AD369* and *AAS44* than for *AC5* (Fig 3.5, lower part).

The Northern blot experiments did not reveal any novel transcripts that would hybridise to the *Adh*-specific probe in any of the variants. There was only a single band in each lane (Fig 3.4 and Fig 3.5). All the bands in each line were of the same size and about 1150 bases as expected (Benyajati *et al*, 1983).

3.2.3.2 Quantitative Northern analyses

3.2.3.2.1 Comparison of *AHA51/nLA248* with *AC8/nLA248*

Heterozygotes between *nLA248* and each variant and the control allele were produced to provide *in vivo* controls for quantification of *Adh* transcripts level (see 2.1.2; Laurie and Stam, 1988). Total RNA was extracted from *AHA51/nLA248* and *AC8/nLA248* adults and from larvae aged 72-80 hours and 88-96 hours. The total RNA was then electrophoresed on an agarose gel and transferred to a nylon filter. The filters loaded with adult RNA were probed with *p*SAF-2 and the filters loaded with both adult and larval RNA were probed with *BSBH6* (see 2.4). The results are shown in Fig 3.6 and the PhosphorImager quantification of the data is summarised in Table 3.8. These data showed that the hybridisation signal for *nLA248* was lower than that of the normal activity control *AC8* and was about one third of the *AC8* level in any developmental stage. But the *AHA51* hybridisation signal in adults was about half of the *nLA248* signal. Therefore the relative intensity of the signal in *AHA51* is about 16% of that in *AC8* adults.

In larvae, both in early 3rd instar (72-80hr) and late 3rd instar (88-96hr), the hybridisation signal for *AHA51* was twice as intense as that for *nLA248* , and the hybridisation signal for *AC8* was three times stronger than that for *nLA248*. These data suggest that in larvae the relative level of hybridisation signal in *AHA51* is about 76% of that in *AC8* .

3.2.3.2.2 Comparison of *AD369/nLA248* with *AC5/nLA248*

The results of quantitative Northern analyses for *AD369/nLA248* and *AC5/nLA248* are summarised in Fig 3.7 and Table 3.9. The hybridisation signal for *nLA248* was about one third that of *AC5*. But the signal intensity of *nLA248* was 1.5-fold higher than that of *AD369*. Thus the relative level of the *Adh*-specific signal in *AD369* in comparison with the *AC5* control was 19.9%. In larvae, the signal level for *AD369* was higher than that for *nLA248* but still

lower than that of *AC5*. The relative signal level of *AD369* compared to *AC5* was about 55.5% in early 3rd instar and about 60% in late 3rd instar.

3.2.3.2.3 Comparison of *AAS44/nLA248* with *AC5/nLA248*

The quantitative Northern analyses showed that the level of *Adh*-specific signals for *AAS44* and *AC5* were both higher than that of *nLA248*, in adults and larvae (Fig 3.8 and Table 3.10). The *Adh*-specific signal of *AC5* was 2.7 times stronger than that of *nLA248*, whilst the intensity of the *AAS44* signal was 1.5 times higher than that of *nLA248*. So in comparison with the *AC5* control, the *AAS44* hybridisation level was 55% of the normal in adults. In larvae the signal for *AAS44* was 125% of that for *AC5* at 88-96hr and 98.8% of the *AC5* signal at 72-80hr; these data suggest that the level of the larval signal in *AAS44* is very similar to that in the control, *AC5*.

3.2.3.3 Quantitative Northern analyses with probe *BSSBN4*

For measuring the transcript levels from different promoters in larvae, *BSBH6* and *BSSBN4* were used as probes to detect different transcripts from the proximal promoter and the distal promoter. *BSBH6* is a common probe both for transcripts from the proximal and distal promoters, and thus reflects the total transcript level. The results obtained from the *BSBH6* probe have been described in previous sections of this chapter. *BSSBN4* is a specific probe used to detect transcripts from the distal promoter only (See Fig 2.2). There is a transient accumulation of RNA with distal sequence between 80-100hr in the third larval instar (Savakis *et al*, 1986). Larvae at 72-78hr and 88-96hr were collected with the aim to detect if the transient accumulation of transcription from the distal promoter occurred in these three variants. Unfortunately, the probe *BSSBN4* did not give any signal in any of the larvae I tested either from the variants or the control.

3.3 Discussion

Cellulose acetate electrophoresis indicates that *AHA51* and *AD369* have similar levels and development profiles for ADH activity. They both have very low activity in adults but nearly normal levels of activity in 3rd instar larvae. *AAS44* shows a relatively higher activity than *AHA51* and *AD369* in adults, but still lower than the normal activity control (Fig. 3.1).

The ADH activity levels suggested by the staining on the cellulose acetate membranes were confirmed by quantitative ADH activity assays. The assay data revealed that in *AHA51* there was about 16% of normal activity in adults and about 60% of normal activity in 3rd instar larvae. The previous data obtained by Freeth and Gibson and my data are highly consistent in terms of the relative activity in adult *AHA51* compared to that in the control *AC8* (Table 3.1 and 3.3), although the actual levels of activity obtained in the different sets of assays varied slightly. These differences are probably due to the different genetic backgrounds in the two sets of materials. The relative level of CRM between *AHA51* and *AC8* is consistent with the differences in activity, being 18% in adults and 61% in larvae.

ADH activity data for *AD369, AAS44* and *AC5* supported the differences found between these three lines on cellulose acetate gels. *AD369* had very low activity in adults, — only 11% of the normal activity allele — while *AAS44* had a higher activity of about 34% of the normal allele. The relative activities both for hemizygotes and homozygotes of *AD369* compared to *AC5* were highly consistent (Table 3.5 and 3.6). The activities of the heterozygotes *AC5/AD369* and *AC8/AHA51* (Table 3.3 and 3.5) did not differ significantly from the expected mid-parent level. Thus the normal *Adh* allele and the variant *Adh* allele do not affect the contribution of each other to ADH activity.

The larval activities of *AHA51, AD369* and *AAS44* were each lower than the controls *AC8* and *AC5* but not as much as in the adults. In *AHA51* early 3rd instar larvae, the activity was very similar to that of *AC8* . In *AHA51* late 3rd

instar, there was about 57% of the activity of *AC8*. In the other two variants, *AD369* had 70% and *AAS44* had 75% of the activity of *AC5* in early 3rd instar larvae. Then in late 3rd instar larvae, *AD369* had a similar level of 66% and *AAS44* had an increased level of 89% of the activity of *AC5*. *AAS44* showed a different ADH profile from the other two alleles in that it had relatively higher activity in adults and nearly normal activity in the larval stage.

Adh^F alleles generally have 2-3 times higher ADH enzyme activity per fly than *Adh*^{S} lines (Winberg *et al,* 1985). A large part of the activity difference is due to the difference in the concentration of ADH protein, as well as the difference in the catalytic efficiency of the enzyme (Laurie and Stam, 1988; Laurie *et al,* 1991). *Adh*^{F} alleles produce more ADH molecules than *Adh*^{S} alleles and each of those molecules has twice the catalytic efficiency of the ADH-S isozyme (Laurie *et al,* 1990; Winberg *et al,* 1985). This difference in catalytic activity per molecule explains much of the characteristically higher activity of *Adh*^{F} strains. Also, the Thr/Lys amino acid replacement at residue 192 causes the difference in the catalytic efficiency, but does not affect the level of ADH protein (Choudhary and Laurie, 1991).

The activity difference between *Adh*^{F} and *Adh*^{S} is not related to the mRNA level. Flies do not have higher ADH mRNA levels in Fast lines than Slow lines (Laurie and Stam, 1988; Collet, 1988 and my data). This result suggests either there is a difference in the translation rates of the two RNAs or there is a difference in the degradation rates of the two proteins *in vivo*. Other studies have eliminated differences in protein stability as a possible cause, therefore it appears that the ADH protein level difference is caused by a difference in translational efficiency (Choudhary and Laurie, 1991).

My results showed that the *Adh*^{F} allele *AC5* and the *Adh*^{S} allele *AC8* gave similar levels of hybridisation signal in adults, although the ADH activity of *AC5* was 3-4 times higher than *AC8*. This result supports the conclusion of Laurie and Stam (1988) who found that *Adh*^{F} strains do not have

characteristically higher levels of ADH mRNA than *Adh*^{S} strains. Contrary to this result, Anderson and McDonald (1983) reported that their F strain had a >2-fold higher ADH RNA level than the S strain, which paralleled the differences in ADH activity.

Quantitative Northern analyses for heterozygotes *AC5 /nLA248* and *AC8/nLA248* showed that the hybridisation signals for *AC5* and *AC8* alleles were higher than that for *nLA248* (Table 3.8, 3.9 and 3.10; Fig. 3.6, 3.7 and 3.8). Chia *et al* (1985b) first reported that in *nLA248* the level of the ADH transcripts and the developmental pattern of their expression were very similar to that in the wild-type *Canton-S Adh* gene. Laurie and Stam (1988) also used *nLA248* as an *in vivo* control in heterozygotes. From the data they provided, the intensity of the *nLA248* signal was about one third less intense than the normal sized *Adh* signal. Our PhosphorImager data showed that the signal intensity of *nLA248* was about one third that for the normal alleles *AC5* and *AC8* I used as controls.

Using *nLA248* as an *in vivo* control, the Northern analyses showed that in adults *AHA51* had 16% of the *AC8* hybridisation signal; *AD369* had 19% and *AAS44* had 55% of the *AC5* hybridisation signal. The intensity of the hybridisation signal in *AHA51, AD369* or *AAS44* adults was in close agreement with the observed activity levels (Fig 3.9). *AHA51* and *AD369* have similar low levels of activity and hybridisation signals while *AAS44* had higher levels than those two variants. *AHA51* had 76% of the *AC8* hybridisation signal both in early third instar larvae and in the late 3rd instar larvae. *AD369* had 60% of the *AC5* hybridisation signal in the early 3rd instar larvae and 55% in the late 3rd instar larvae. *AAS44* had a similar hybridisation signal to *AC5* in third instar larvae (Table 3.8, 3.9 and 3.10). The hybridisation signal level in larvae of the three lines paralleled the ADH activity level in larvae (Fig 3.10).

In summary the results obtained from the quantitative assays of ADH activity and the Northern analyses paralleled each other in the three variants (Fig 3.9 and 3.10). It is likely, but not proved, that the hybridisation signal

represents the amount of RNA transcript produced by the *Adh* gene, but it is possible that the differences also reflect differences in the stability of the transcripts produced (Brawerman, 1987).

The two variants *AHA51* and *AD369* have very similar phenotypes; both have very low levels of activity (8% in *AHA51* and 11% in *AD369*) and transcript (16% in *AHA51* and 20% in *AD369*) in adults compared to the controls. They also have reduced activity (57% in *AHA51* and 66% in *AD369*) and hybridisation signals (76% in *AHA51* and 60% in *AD369* l) in 3rd instar larvae, but these levels are much higher than in adults. It is possible that *AD369* and *AHA51* represent two copies of the same allele as they have very similar ADH profiles in adults and larvae.

AAS44 is clearly different as it has a higher ADH level (34%) and transcript level (55%) in adults and nearly normal activity (89%) and transcript level (98%) in 3rd instar larvae (Table 3.7 and 3.10). The activity profile in larvae is different to that of the other two alleles as it increases from early to late 3rd instar larvae (Table 3.7) while in *AD369* and *AHA51* activity is lower in late than in early 3rd instar larvae.

The ADH activity and *Adh* transcript profile in 3rd instar larvae differed in *AAS44* from *AHA51* and *AD369*, in addition the activities in adults of *AHA51* and *AD369* were much reduced compared to that in *AAS44* . These results suggest that the *Adh* expression in *AHA51* and *AD369* from the distal promoter is much more affected than that from the proximal promoter. Using the probe *BSSBN4* specific for transcripts from the distal promoter, I expected to be able to demonstrate that in the late 3rd instar larvae of *AD369* and *AHA51*, the transient accumulation of the distal transcript did not occur; while transcription from the distal promoter in control larvae was normal (see Appendix V). These results would have indicated that the ADH activity profile in *AD369* and *AHA51* is mainly due to the reduction of *Adh* expression from the distal promoter. Unfortunately, the distal promoter specific probe did not give any

signal in any of the larvae either from the variants or the control, albeit it was expected to give signal at least in larvae from the control.

These data also show that the modifier(s) present in the three variants act in *cis-* but not in *trans-*, that is the effect is only on the *Adh* allele on the same chromosome as the postulated modifier(s). Assay data from heterozygote of *AC8/AHA51* and *AC5/AD369* were not significantly different to the expected mid-parent activities (Table 3.3 and 3.5), which means the modifier(s) present in the *AHA51* or *AD369* lines do not affect the ADH activity of the normal allele on the other chromosome. Quantitative Northern analysis of *nLA248* heterozygotes confirm that only *cis*-acting effects are observed. *Trans*-acting effects have previously been demonstrated for some GPDH low activity variants (Reed, 1993).

In comparison with my data, the Adh^{SL} allele shows a similar ADH profile (Schott *et al*, 1988). The Adh^{SL} allele had about half the total ADH activity of their standard Adh^{S} strain and a reduction of 44% in the amount of ADH mRNA in adults. The Adh^{SL} allele consistently showed reduced ADH activity (73% of the normal) and mRNA level (83% of the normal) in larvae, but the reduction was much smaller than that seen in adults.

My results suggest that each variant *AHA51, AD369* and *AAS44* may have a factor(s) that affects the amount or stability of the *Adh* transcripts but does not affect the size of the transcripts. Also the factor(s) is likely to be closely linked to the *Adh* locus as each variant behaves in breeding programs like an *Adh* allele. Thus in future work it will be important to analyse the molecular structure of the three variants and to try to identify the factor(s) responsible for the abnormal ADH profile.

AC8/Df (slow)　*AHA51/Df*　*AAS44/Df*　*A D369/Df*　*AC5/Df* (fast)

Figure 3.1　Electrophoretic phenotypes of adults

Stained ADH proteins after electrophoresis on cellulese acetate membrane of extracts from adult flies of the low activity variants *AHA51/Df(2L)64j*, *AAS44/Df(2L)64j*, *AD369/Df(2L)64j* and two control lines *AC8/Df(2L)64j* (electrophoretically "slow") and *AC5/Df(2L)64j* (electrophoretically "fast").

Figur 3.2 Distinction between heterozygous and homozygous *nLA248* larvae.

Heterozygous and homozygous *nLA248* larvae are distinguishable after riboflavin feeding (see Chap. 2.2). Heterozygous *nLA248/+* larvae (number 1 and 2) are darker yellow, while homozygous *nLA248/nLA248* larvae (number 3) are light yellow.

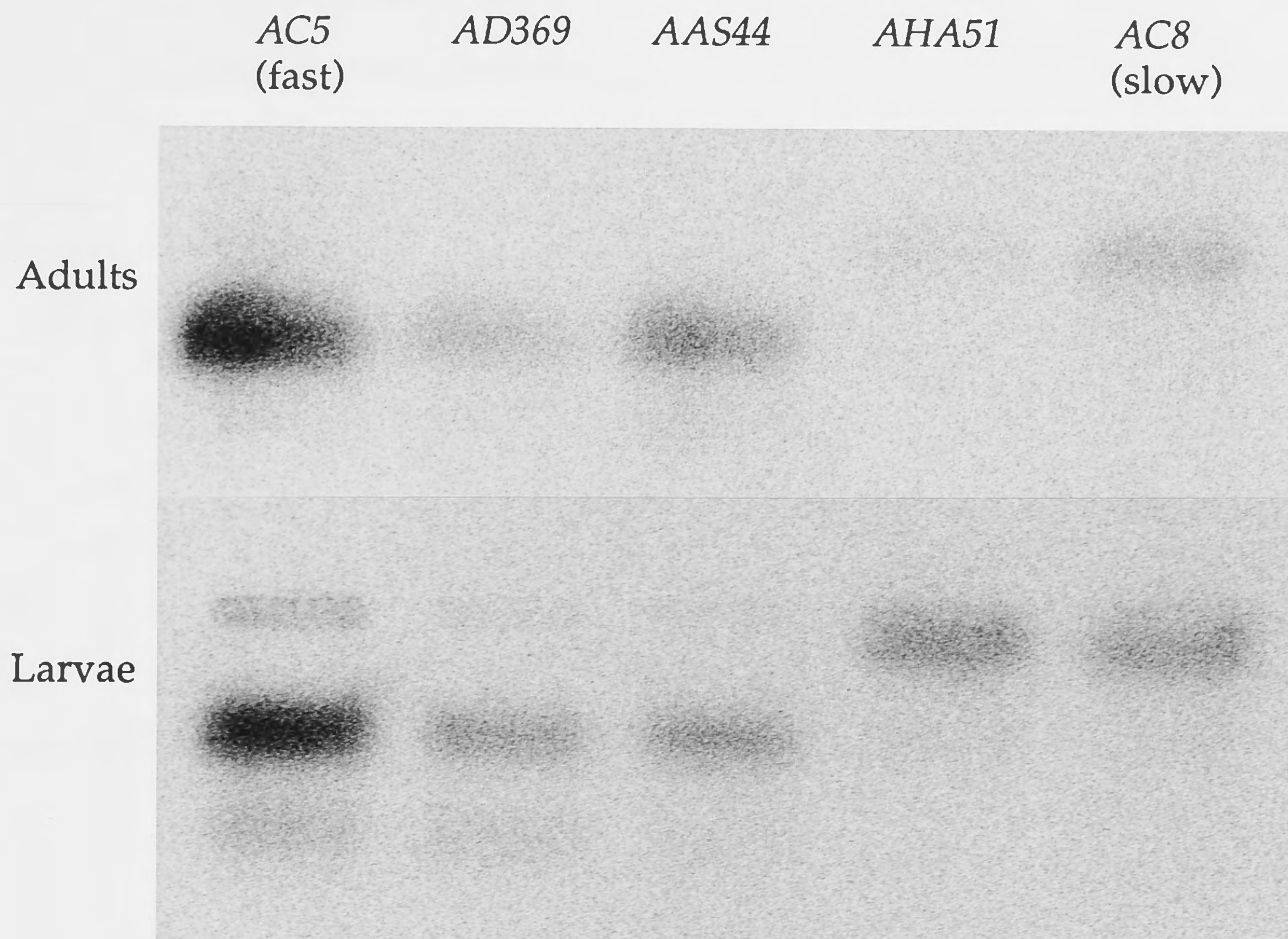


Figure 3.3 Electrophoretic phenotypes of variants heterozygous with *nLA248* in adults and larvae

Stained ADH proteins after electrophoresis on cellulese acetate membrane of extracts from the low activity variants heterozygous for *nLA248*. They are *AD369/nLA248*, *AAS44/nLA248*, *AHA51/nLA248* and two control lines *AC5/nLA248* (electrophoretically "fast") and *AC8/nLA248* (electrophoretically "slow").

Table 3.1 ADH activity and the amount of CRM relative to the control, in flies heterozygous for either *Df(2L)A379* or Adh^{nB}. Each value is the mean of assays of extracts from progeny in four separate cultures each initiated by crosses between 20 pairs (Experiments 1 and 2) and 10 pairs (Experiment 3) of flies.

Genotype	Experiment	ADH activity (±SE)	Relative activity	ADH CRM (±SE)	Relative CRM level
$AC8/Adh^{nB}$ (control)	1	30.9(±1.4)	1	53.4(±2.9)	1
	2	33.0(±1.2)	1	59.3(±1.3)	1
	3	51.9(±2.0)	1	79.6(±2.8)	1
$AHA51/Adh^{nB}$	1	4.3(±0.1)	0.139	9.2(±0.3)	0.172
	2	5.5(±0.4)	0.167	11.2(±0.4)	0.189
	3	7.4(±1.2)	0.143	10.4(±0.4)	0.131
AC8/Df(2L)A379 (control)	1	30.8(±1.9)	1	47.9(±4.3)	1
	2	28.9(±1.9)	1	48.5(±4.3)	1
	3	46.3(±1.9)	1	62.7(±2.3)	1
AHA51/Df(2L)A379	1	5.5(±0.2)	0.179	10.0(±0.1)	0.209
	2	6.6(±0.5)	0.228	11.2(±0.4)	0.231
	3	7.4(±0.6)	0.160	10.2(±0.3)	0.163

Table 3.2 ADH activity and the amount of CRM relative to the heterozygote control, in larvae heterozygous for either *Df(2L)A379* or Adh^{nB}. Each value is the mean of assays of extracts from progeny in four separate cultures each initiated by crosses between 20 pairs (Experiments 1 and 2) and 10 pairs (Experiment 3) of flies.

Genotype	Experiment	ADH activity (±SE)	Relative activity	ADH CRM (±SE)	Relative CRM
AC8/Adh^{nB} or *Df(2L)A379*	1	19.9(±0.7)	1	65.0(±6.3)	1
	2	18.5(±1.9)	1	58.5(±9.1)	1
	3	19.0(±0.6)	1	70.7(±3.3)	1
AHA51/Adh^{nB} or *Df(2L)A379*	1	13.5(±0.7)	0.678	40.9(±0.4)	0.629
	2	13.6(±1.0)	0.735	37.8(±3.0)	0.646
	3	13.6(±0.6)	0.716	40.1(±3.1)	0.567

Table 3.3 ADH activity for *AC8* and *AHA51* homozygous and heterozygous adults. Each value is the mean of assays of extracts from progeny in single pair crosses. Standard errors are shown in parenthesis.

Genotype	Number of assays	ADH activity	Expected mid-parent activity	Relative activity
AC8/AC8	5	111.5(±11.3)		1
AHA51/AHA51	4	18.3(±1.8)		0.16
AC8/AHA51	5	61.4(±5.8)*	64.9	

* not significantly different to the mid-parent activity

Table 3.4 ADH activity of *AC8/nLA248* and *AHA51/nLA248* heterozygotes.

Each value is the mean of assays of extracts from the progeny of single pair crosses for adults or a mass cross for larvae. Standard errors in parenthesis.

Genotype	Developmental stage	Numbers of Assay	ADH activity	Expected mid-parent activity	Relative activity
AC8/nLA248	adults	5	58.18(±3.09)	55.8	1
AHA51/nLA248	adults	5	4.76(±0.33)*	9.2	0.08
AC8/nLA248	72-78hr larvae	4	22.85(±1.98)	34.8 (±5.4)	1
AHA51/nLA248	72-78hr larvae	4	24.98(±4.72)		1.09
AC8/nLA248	90-96hr larvae	5	37.66(±4.79)	36.98(±5.2)	1
AHA51/nLA248	90-96hr larvae	5	21.64(±2.89)		0.57

* significantly lower than the expected mid-parent level

Table 3.5 ADH activity in adult flies of *AC5* and *AD369* homozygotes and heterozygotes, and hemizygotes with *Df(2L)64j*. Each value is the mean of the assays of progeny from single pair crosses. The standard errors are in parenthesis.

Genotype	Number of assays	ADH activity	Expected mid-parent activity	Relative activity
AC5/AC5	4	409.1(±11.4)		1
AC5/Df(2L)64j	5	188.6(±7.5)*	204.6	
AD369/AD369	5	46.7(±3.99)		0.11
AD369/Df(2L)64j	5	24.2(±2.9)*	23.4	
AC5/AD369	5	203.1(±14.0)*	227.9	

* not significantly different to the mid-parent level

Table 3.6 ADH activity of *AC5/nLA248* and *AD369/nLA248* heterozygotes. Each value is the mean of assays of the progeny from single pair crosses for adults or from a mass cross for larvae. Standard errors are in parenthesis.

Genotype	Developmental stage	Numbers of Assay	ADH activity	Expected mid-parent activity	Relative activity
AC5/nLA248	adults	5	229.8(±11.8)*	204.6	1
AD369/nLA248	adults	5	24.3(±2.6)*	23.4	0.11
AC5/nLA248	72-78hr larvae	4	67.6(±4.3)		1
AD369/nLA248	72-78hr larvae	4	47.6(±2.4)		0.70
AC5/nLA248	90-96hr larvae	2	80.4(±0.6)		1
AD369/nLA248	90-96hr larvae	2	53.4(±3.1)		0.66

*: not significantly different to mid-parent level

Table 3.7 ADH activity of *AC5/nLA248* and *AAS44/nLA248* heterozygotes. Each value is the mean of assays of the progeny from single pair crosses for adults or a mass cross for larvae. Standard errors in parenthesis.

Genotype	Developmental stage	Number of Assays	ADH activity	Relative activity
AC5/nLA248	adults	5	229.6(±9.7)	1
AAS44/nLA248	adults	5	78.5(±7.6)	0.34
AC5/nLA248	72-78hr larvae	4	71.7(±17.5)	1
AAS44/nLA248	72-78hr larvae	5	53.7(±3.5)	0.75
AC5/nLA248	90-96hr larvae	2	67.5(±20.1)	1
AAS44/nLA248	90-96hr larvae	2	60.3(±9.1)	0.89

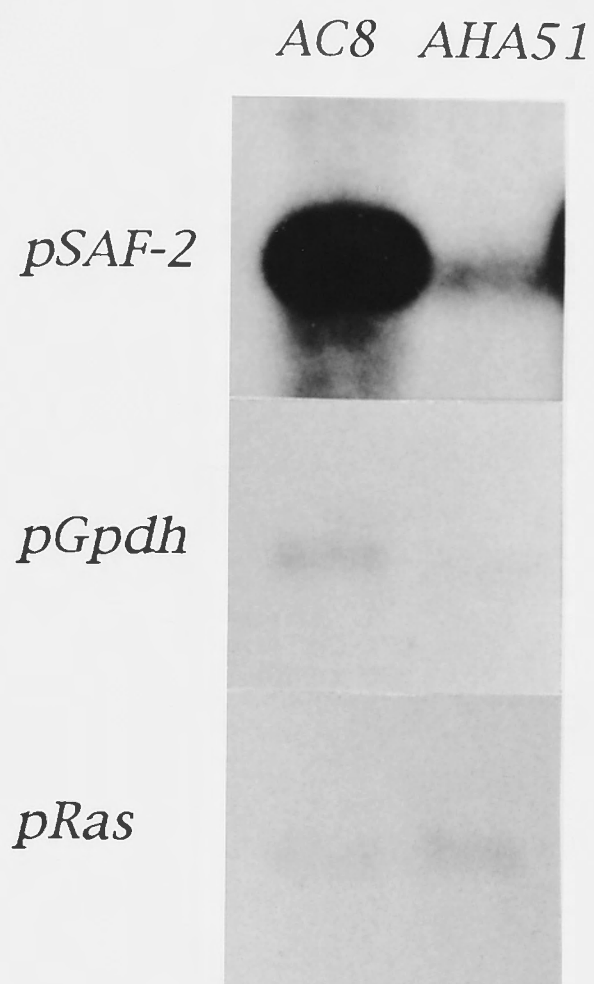


Figure 3.4 Northern blots of total RNA from *AC8* and *AHA51* homozygous adults

Total RNA extracted from *AHA51* homozygous adults was electrophoresed and a Northern blot was prepared. Total RNA from the normal activity allele *AC8* was loaded as a control. The filter was probed with *pSAF-2* and revealed a signal of normal size but low amount in *AHA51*. The middle part and lower part of the figure shows the hybridisation pattern obtained from the same filter after it was washed and re-hybridised to the *pGpdh* and *pRas* probes, which were used as controls to check that equal amounts of RNA were loaded on the gel.

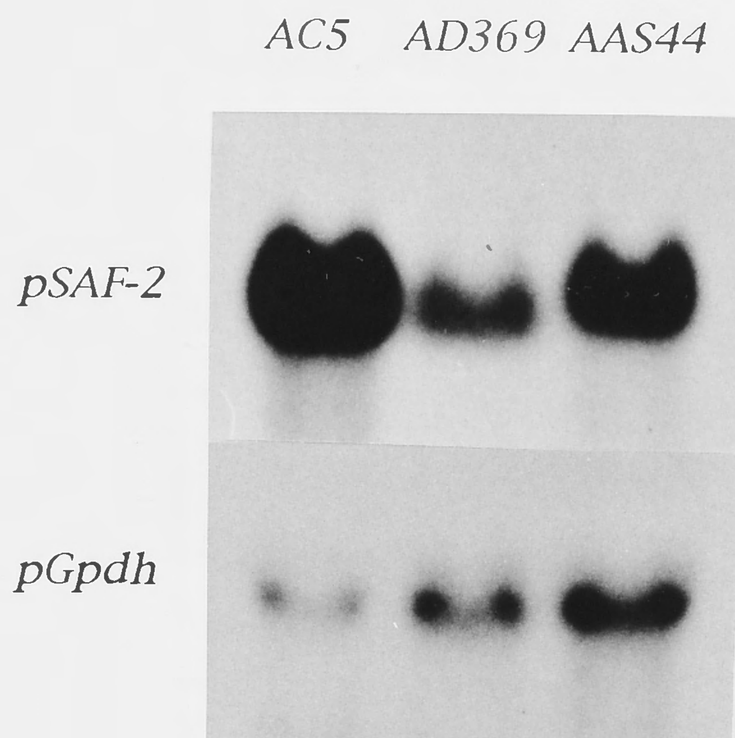


Figure 3.5 Northern blots of total RNA from *AC5* and *AD369* homozygotes and *AAS44/Df(2L)64j* heterozygous adults

Total RNA extracted from *AD369* homozygous adults and *AAS44 /Df(2L)64j* heterozygous adults were electrophoresed and Northern blots were prepared. Extract from the normal activity allele *AC5* was loaded as a control. The filter was probed with *pSAF-2* and gave a normal sized signal but at a lower level in *AD369* and *AAS44* than in the control *AC5*. The lower part of the figure shows the hybridisation pattern for the *pGpdh* probe which was used as a control to check that equal amounts of RNA were loaded on the gel.

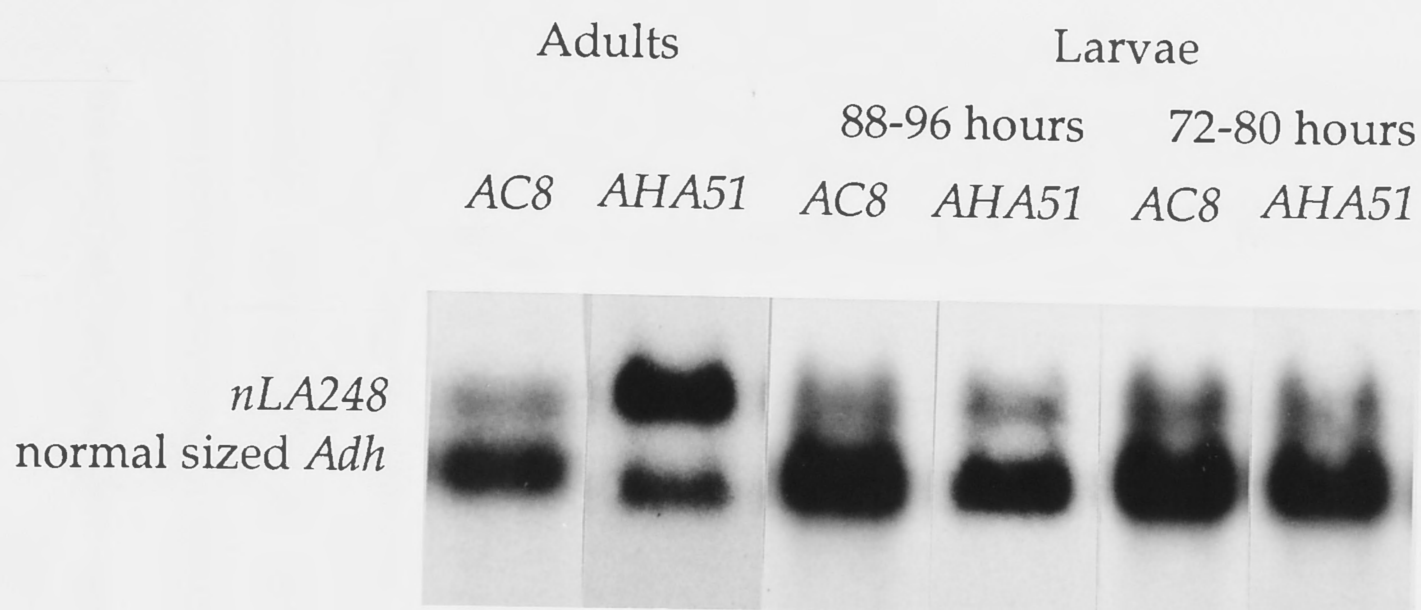


Figure 3.6 Quantitative Northern analysis for adults and larvae of *AHA51/nLA248* and *AC8/nLA248* heterozygotes

From left to right, samples were loaded as: *AC8* adults, *AHA51* adults, *AC8* 88-96 hours larvae, *AHA51* 88-96 hours larvae, *AC8* 72-80 hours larvae and *AHA51* 72-80 hours larvae. The adult filters were probed with the nick-translated labelled *pSAF-2* and the larval filters were probed with the nick-translated labelled *BSBH6*. The upper band represents the *nLA248*-specific transcript. The lower band represents the normal sized *Adh*- specific transcript.

Table 3.8 Quantitative Northern Analyses for *AHA51/nLA248* and *AC8/nLA248* heterozygotes

The quantitative results of the Northern analysis shown in Figure 3.6 are summarised in this table. The intensity of each band was measured using the PhosphorImager instrument (Molecular Dynamics). Each value is the mean intensity of three measured samples. The relative *Adh/nLA248* specific intensity was calculated separately for each sample, to control for differences in RNA loadings. Only the average value of the relative intensity between the normal sized-*Adh* band and the *nLA248* band is shown in the table. Standard errors are shown in parenthesis.

Developmental stage	Adult		88-96 hours larvae		72-80 hours larvae	
Genotype	*AC8*	*AHA51*	*AC8*	*AHA51*	*AC8*	*AHA51*
nLA248 1330-bases band intensity	13172 (±2155)	18890 (±2980)	102646 (±11856)	57813 (±8700)	155223 (±20960)	102417 (±7585)
normal 1130-bases band intensity	43641 (±2988)	10259 (±1644)	286429 (±9448)	122947 (±2197)	441808 (±47938)	222918 (±15678)
normal band:*nLA248* band	3.313	0.543	2.79	2.13	2.85	2.18
AHA51 : AC8	0.164		0.763		0.765	

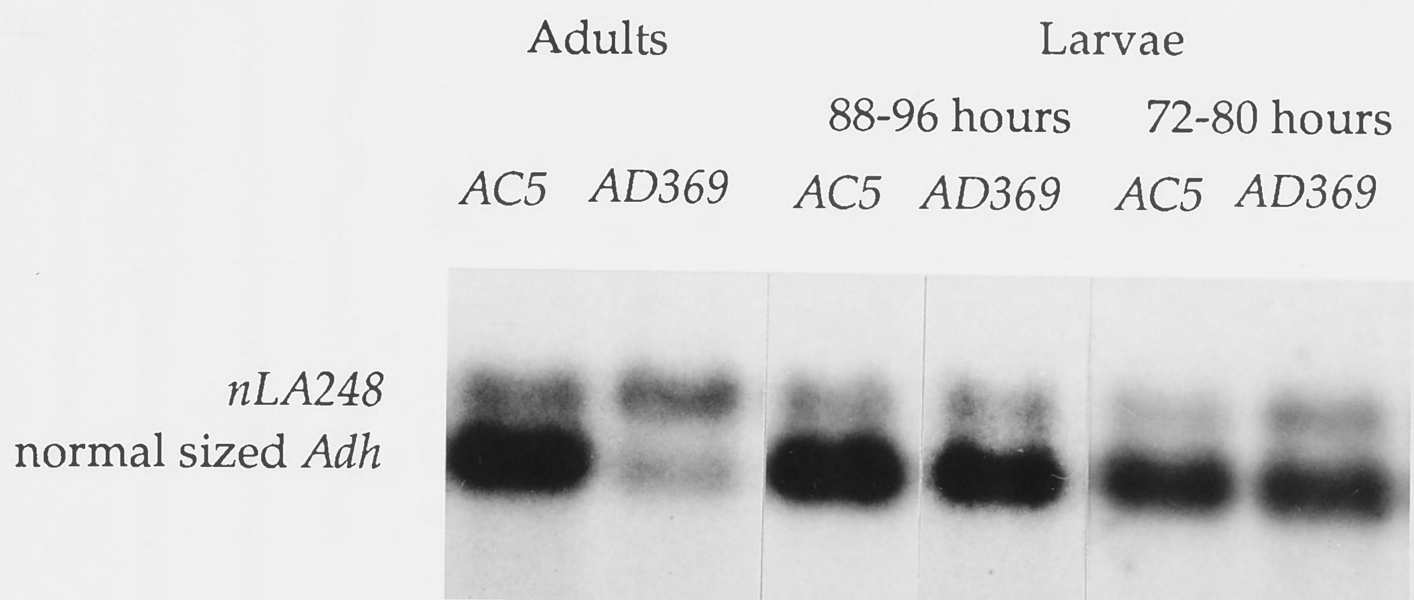


Figure 3.7 Quantitative Northern analysis for adults and larvae of *AD369/nLA248* and *AC5/nLA248* heterozygotes

From left to right, samples were loaded as: *AC5* adults, *AD369* adults, *AC5* 88-96 hours larvae, *AD369* 88-96 hours larvae, *AC5* 72-80 hours larvae and *AD369* 72-80 hours larvae. The adult filters were probed with the nick-translated labelled *pSAF-2* and the larval filters were probed with the nick-translated labelled *BSBH6*. The upper band represents the *nLA248* specific signal. The lower band represents the normal *Adh* specific signal.

Table 3.9 Quantitative Northern Analyses for *AD369/nLA248* and *AC5/nLA248* heterozygotes

The quantitative results of the Northern analysis shown in Figure 3.7 are summarised in this table. The intensity of each band was measured using the PhosphorImager instrument (Molecular Dynamics). Each value is the mean intensity of three measured samples. The relative *Adh/nLA248* specific intensity was calculated separately for each sample, to control for differences in RNA loadings. Only the average value of the relative intensity between the normal sized-*Adh* band and the *nLA248* band is shown in the table. Standard errors are shown in parenthesis.

Developmental stage	Adult		88-96 hours larvae		72-80 hours larvae	
Genotype	*AC5*	*AD369*	*AC5*	*AD369*	*AC5*	*AD369*
nLA248 1330-bases band intensity	17246 (±2154)	37925 (±7749)	17677 (±4350)	12047 (±4661)	2428(±617)	2918(±608)
normal 1130-bases band intensity	54753 (±14433)	23864 (±5257)	69925 (±13461)	28842 (±6284)	13000 (±1686)	8659(±1175)
normal band: *nLA248* band	3.17	0.63	3.96	2.39	5.35	2.97
AD369 : AC5	0.199		0.601		0.555	

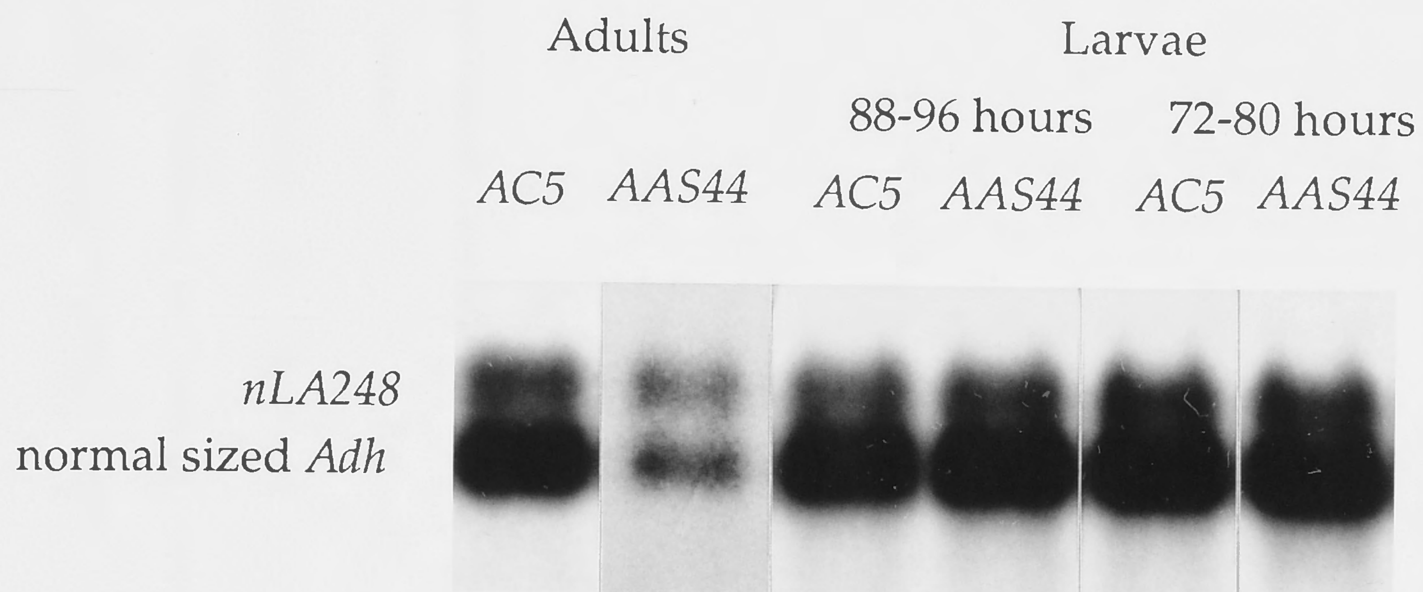


Figure 3.8 Quantitative Northern analysis for adults and larvae of *AAS44/nLA248* and *AC5/nLA248* heterozygotes

From left to right, samples were loaded as: *AC5* adults, *AAS44* adults, *AC5* 88-96 hours larvae, *AAS44* 88-96 hours larvae, *AC5* 72-80 hours larvae and *AAS44* 72-80 hours larvae. The adult filters were probed with the nick-translated labelled *pSAF-2* and the larval filters were probed with the nick-translated labelled *BSBH6*. The upper band represents the *nLA248*- specific level. The lower band represents the normal sized*Adh*-specific level.

Table 3.10 Quantitative Northern Analyses for *AAS44/nLA248* and *AC5/nLA248* heterozygotes

The quantitative results of the Northern analysis shown in Figure 3.8 are summarised in this table. The intensity of each band was measured using the PhosphorImager instrument (Molecular Dynamics). Each value is the mean intensity of three measured samples. The relative *Adh/nLA248* specific intensity was calculated separately for each sample, to control for differences in RNA loadings. Only the average value of the relative intensity between the normal sized*Adh* band andthe *nLA248* band is shown in the table. Standard errors are shown in parenthesis.

Developmental stage	Adult		88-96 hours larvae		72-80 hours larvae	
Genotype	*AC5*	*AAS44*	*AC5*	*AAS44*	*AC5*	*AAS44*
nLA248 1330-bases band intensity	37350 (±2958)	9983(±843)	27826 (±3253)	39891 (±2060)	46188 (±372)	51784 (±3327)
normal 1130-bases band intensity	102014 (±9358)	15002 (±4118)	104661 (±28013)	188207 (±4989)	217705 (±10247)	241070 (±10615)
normal band: *nLA248* band	2.73	1.50	3.76	4.72	4.71	4.66
AAS44 : AC5	0.550		1.25		0.988	

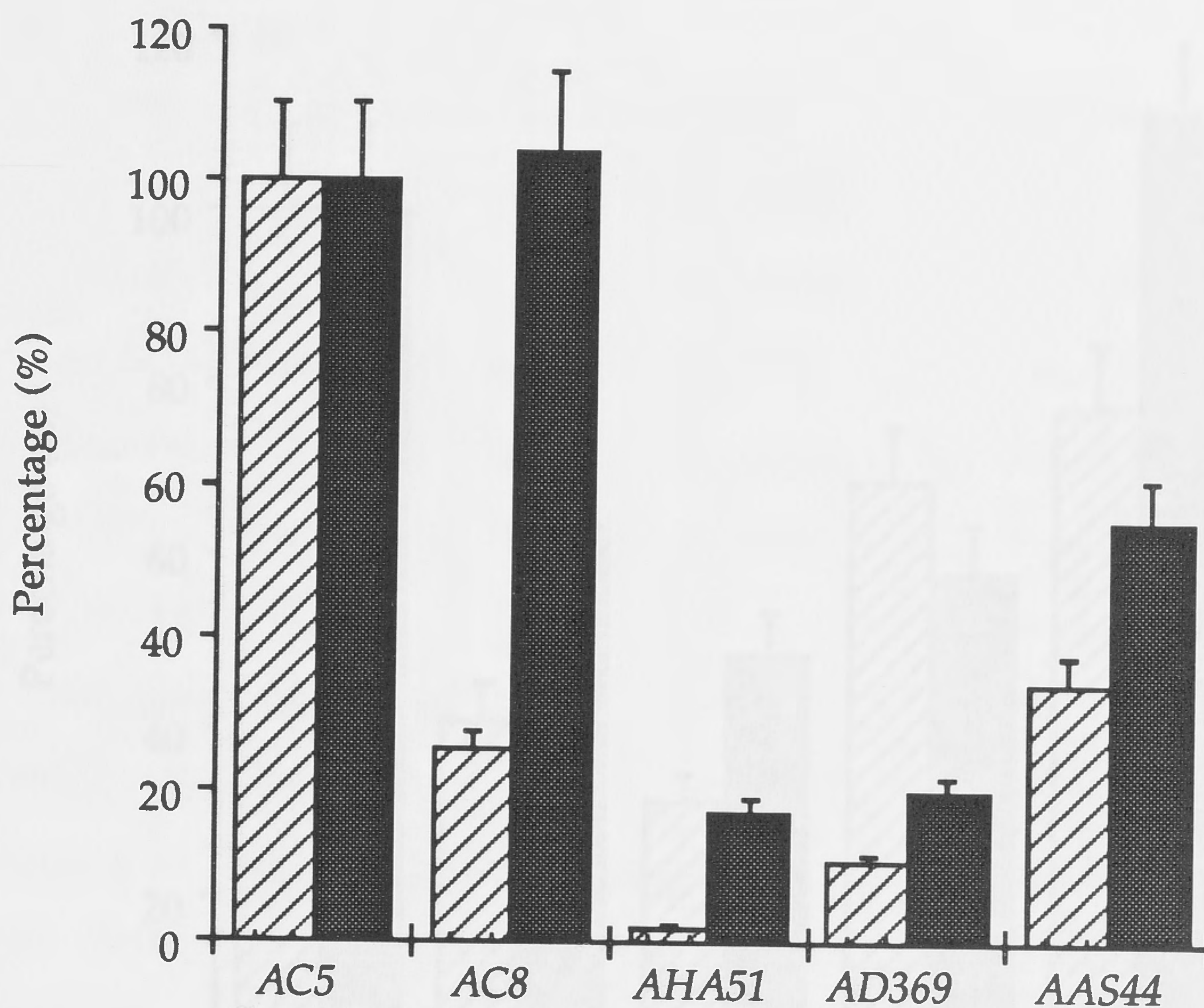


Figure 3.9 Comparison of activity and RNA hybridisation signal level in adults

This diagram shows the ADH activities (hatched) and quantified hybridisation signals (solid) in adults in the three low ADH activity alleles and normal activity control alleles *AC5*^{F} and *AC8*^{S}. The *AC5*^{F} control allele is refered to as 100%. Standard errors are indicated by bars.

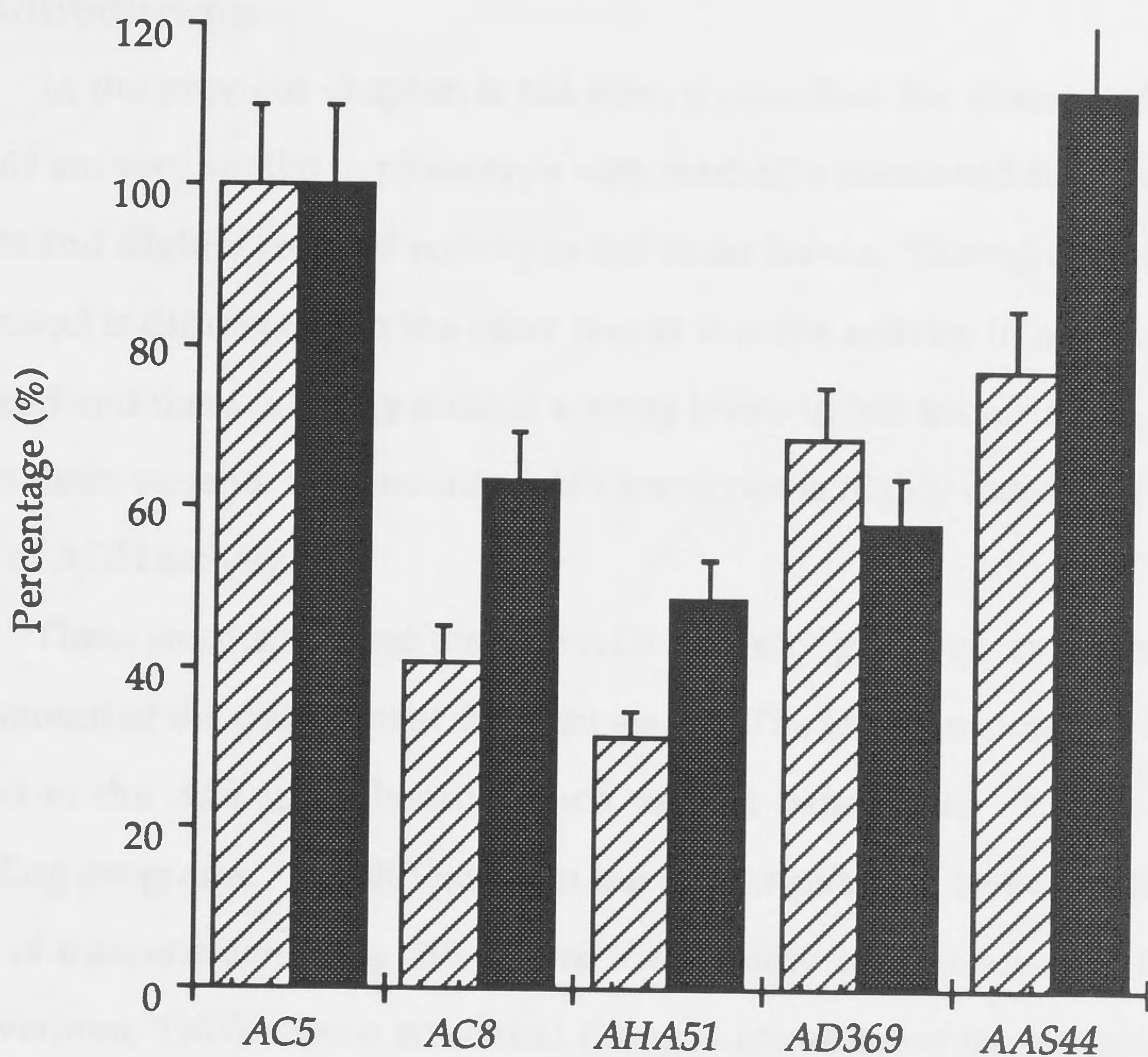


Figure 3.10 Comparison of activity and RNA hybridisation signal level in larvae

This diagram shows the ADH activities (hatched) and the quantified hybridisation signal (solid) level in larvae in the three low ADH activity alleles and normal activity control alleles *AC5*^{F} and *AC8*^{S}. The *AC5*^{F} control allele is refered to as 100%. Standard errors are indicated by bars.

Chapter 4 Molecular structures of the three *Adh* low activity alleles

4.1 Introduction

In the previous chapter, it has been shown that the alleles *AHA51* and *AD369* are very similar in phenotype with markedly decreased ADH activity in adults and slightly reduced activity in 3rd instar larvae. The third allele *AAS44* examined is different from the other two in that the activity in adults is not as reduced and there is nearly normal activity levels in 3rd instar larvae. In each of the three variants the amount of *Adh* transcript is highly correlated with the level of ADH activity.

These results suggest that the ADH activity phenotypes are related to the amount of the transcript at different stages. The factor responsible is tightly linked to the *Adh* locus because each variant behaves as an *Adh* allele in breeding programs. The difference in the transcript levels could be due to the level of transcription being affected or the stability of the transcript in adults (Brawerman, 1987). Some structural changes could cause transcription to be slower in adults, or some could affect the stability of the adult transcript, such as mutations altering the 5' leader sequences might affect the stability of the transcripts (Morris *et al*, 1986; Luscher *et al*, 1985; Piechazyk *et al*, 1985; Rabbitts *et al*, 1985).

Previous work has shown that the allele *AAS44* has an insertion in the adult leader sequence of the *Adh* gene (Jiang and Gibson, 1992a), however the sequence of neither the gene nor the insertion in *AAS44* has been obtained. The molecular structures of *AHA51* and *AD369* are not known yet. It is possible that both *AHA51* and *AD369* have modifiers like Adh^{SL}, which had a insertion approximately 3 kb 5' upstream to the distal promoter (Schott *et al*, 1988). Thus it is necessary to investigate the molecular landscape of the *Adh* gene region in each of the three variants. Using Southern blotting and hybridisation with the

*p*SAC-1 and *p*SAF-2 probes (Goldberg, 1980), it should be possible to detect whether there are any insertions or deletions present in these three *Adh* alleles in an 11.8 kb *Sac*I region.

Mutations could also occur in the 5' region which contains important motifs involved in the regulation of *Adh* transcription. Thus it is necessary to obtain the nucleotide sequence of each allele and compare it with similar data for normal *Adh* genes. The consensus sequence of normal activity *Adh* genes and their flanking regions in *Drosophila melanogaster* has been published (Kreitman, 1983; Kreitman and Hudson, 1991; Appendix I). Any putative amino acid substitution in the encoded protein as well as any change in the 5' and 3' regulatory regions of the gene could be revealed by DNA sequencing. This chapter describes the molecular landscape of *AHA51*, *AD369* and *AAS44* and their genomic sequences.

4.2 Results

4.2.1 Restriction maps of the *Adh* region in the three variants

Genomic DNA was extracted from adult flies homozygous for *AC5*, *AD369*, *AC8*, *AHA51* and from the heterozygotes *AAS44*/*Df(2L)64j* and *AC5*/*Df(2L)64j* using the method described by Chia *et al* (1985a). The DNA was digested with a variety of hexanucleotide endonucleases, including *Eco*RI, *Hpa*I, *Pvu*II, *Sal*I, *Bam*HI and *Hind*III, in single and/or double digest and then electrophoresed on an agarose gel. Southern blots were hybridised with the *p*SAC-1 and *p*SAF-2 probes, which contain *Adh* sequences (see 2.4 and Fig. 2.1). By comparing the size of the fragments produced from the each enzyme(s), restriction maps were constructed for *AD369*, *AHA51* and *AAS44*.

The hybridisation pattern of *AHA51* genomic DNA digested with *Eco*RI and probed with *p*SAC-1 showed two bands of 4.13 kb and 2.70 kb (Fig 4.1), while only one band of 4.5 kb was detected in the control *AC8*. This result indicated that there was a 2.4 kb insert with an extra *Eco*RI site located between

two *Eco*RI sites (at -2.8, +2.1) in *AHA51*. Both *Sal*I and *Pvu*II single digests confirmed the size of the insertion and revealed another *Pvu*II site in the insert. Double digests with *Eco*RI/*Hpa*I, *Eco*RI/*Pvu*II and *Sal*I/*Hpa*I verified the position of the insertion and narrowed the location to a region between the *Sal*I (-1.3) and *Hpa*I (-0.9) sites (Fig. 4.5).

Genomic DNA from *AD369* and *AC5* digested with *Eco*RI, probed with *p*SAC-1 revealed one band in each lane (Fig 4.2). The fragment in *AD369* was 6.03 kb, and that in *AC5* was 4.85 kb. This result indicated that an insertion was located between the two *Eco*RI sites (at -2.8 and +2.1) in *AD369*, and the size of the insert was about 1.2 kb. A double digest with *Eco*RI /*Hpa*I showed that the region of the insertion was 3' to *Eco*RI (-2.8) and 5' to *Hpa*I (-0.9). Further digestion with *Sal*I/*Hpa*I showed that the insert was located between the *Sal*I (-1.3) and *Hpa*I (-0.9) sites (Fig 4.5). A *Pvu*II single digest showed that there was an extra *Pvu*II site in the insert, and this was verified by a *Eco*RI/*Pvu*II double digest.

The hybridisation pattern of *AAS44* genomic DNA digested with *Eco*RI or *Sal*I , probed with *p*SAC-1 showed that there was an insertion of about 5 kb which contains an extra *Eco*RI and *Sal*I sites (Fig 4.3). *Eco*RI digestion revealed two bands, 5.72 kb and 4.09 kb in *AAS44* but one band of 4.5 kb in *AC5* . The double digests with *Eco*RI/*Hpa*I and *Eco*RI/*Pvu*II showed that the insert was located 3' to the *Hpa*I (-0.9) site and 5' to the *Pvu*II (+0.2) site. Another single digest with *Hin*dIII indicated that the insert was located 5' to the *Hin*dIII (-0.7) site. Based on the above results, the 5 kb insertion was mapped into a region between the *Hpa*I (-0.9)and *Hin*dIII (-0.7) sites (Fig 4.5).

The *p*SAC-1 probe mainly detected the 4.5 kb central *Eco*RI region surrounding the *Adh* gene (see Fig. 2.1). All of the *cis*-acting DNA sequences required for normal *Adh* gene expression *in vivo* are contained within an 11.8 kb *Sac*I restriction fragment which includes 5.5 kb 5' and 4.5 kb 3' flanking sequence (Goldberg *et al*, 1983). In order to check this wider region, the *p*SAF-2

probe, which contains an 11.8 kb *Sac*I fragment (see Fig. 2.1), was used in further studies. The hybridisation results are shown in Figure 4.4. The results were consistent with those obtained with the *p*SAC-1 probe. In *AHA51,* no other change was found, besides the 2.4 kb insert detected with the *p*SAC-1 probe. Comparison of the size of the fragments in *AD369* and *AC5* showed that besides the main difference in the 4.5 kb *Eco*RI band detected by the *p*SAC-1 probe, there was a small insertion of about 50 bp in the 1.5 kb *Eco*RI (-4.2 and -2.8) fragment located in the 5' flanking region of the *AD369 Adh* gene. The data for *AAS44* revealed a 200 bp insertion also located in the same 1.5 kb *Eco*RI fragment.

In summary, the restriction mapping showed that there were insertions in all three of the *Adh* low activity alleles (Fig 4.5). In *AHA51,* the insert was 2.4 kb and located between the *Sal*I/*Hpa*I sites (at -1.3 and -0.9). In *AD369,* a 1.2 kb insert was located in the same region and a 50 bp insertion was located in the 5' *Eco*RI (-4.2, -2.8) fragment. *AAS44* had a 5 kb insert between the *Hpa*I (-0.9) and *Hin*dIII (-0.7) sites and a 200 bp insert in the 5' *Eco*RI (-4.2, -2.8) region. The large insertions in all three variants were located in a region which included the *Adh* promoters and the adult leader sequences.

4.2.2 DNA sequence analyses of the three low activity alleles

In order to check that there were no other mutations in the *Adh* genes that could affect activity it was necessary to sequence them. PCR amplified fragments were used for sequencing and defining the position of the insertions detected by the Southern blots.

4.2.2.1 PCR amplification of the *Adh* gene

In vitro DNA amplification with *Taq* polymerase was used to produce DNA fragments spanning the *Adh* gene. The normal size of the *Adh* gene in *Drosophila melanogaster* is about 1.8 kb (Benyajati *et al,* 1981; 1983). So

overlapping fragments were produced by PCR amplification. JG88/22, JG87/83 and JG87/17 spanned the 5' flanking sequence, distal promoter and part of the first intron; JG82/17 and JG13/21 spanned a region from the intron 1 to exon 1; and JG13/48, JG14/48 and JG15/48 spanned the three exons (Fig 4.6. Sequences of primers are given in Appendix II).

PCR amplification with JG88/22 with *AD369* showed that there was a single band of size 1.9 kb, which was 1.2 kb larger than the 0.7 kb band obtained in the control *AC5* (Fig 4.7). This result is in agreement with the data obtained from the Southern analysis which showed a 1.2 kb insertion in this region. A single band amplified with JG13/48, which spanned the rest of the gene, was the same size (~1.5 kb) as the band in the control *AC5*.

I was unable to amplify a fragment that would have included the expected 2.4 kb insert in *AHA51* by using the primers JG88/22(Fig 4.8a). Instead, the primers JG87/83 (Fig 4.8) were used in the PCR with an extended annealing time (3 minutes) and an extension time increased to 3 minutes because of the size of the expected fragment. Two bands were obtained with primer JG87/JG83 (Fig. 4.8b). The large band was about 2.7 kb, which was the expected size of the fragment including the insertion. A small band of 300 bp was also present in *AHA51*, which corresponded to the amplified band in the control *AC8*, under the same conditions. Amplification with JG87/17 was also successful under the PCR conditions for a larger target (Fig. 4.8c). Two bands of 3.2 kb and 0.8 kb, were obtained in *AHA51* but no band was present in the control without DNA templet. Amplifications of the rest of the gene in *AHA51* with JG14/21 and JG15/48 produced bands of 0.4 kb and 0.8 kb respectively. Similar results were obtained in the control *AC8* (Fig. 4.8d).

Genomic DNA amplified with JG87/22 produced one band of 0.5 kb in *AAS44*, as well as in the control *AC5* (Fig 4.9). Amplification with JG13/17, which spanned the 5 kb insertion region, was not successful in *AAS44*. The rest

of the *AAS44* gene was amplified with JG14/48 and produced a 1.1 kb fragment, as well as in the control *AC5*.

4.2.2.2 Sequence results of the *Adh* gene region

PCR amplified fragments, after removing the primers, were used for direct sequencing. This technique eliminated the risk of errors that might be caused by nucleotides wrongly incorporated in the process of amplification being magnified by cloning the fragments (Innis *et al*, 1988). The sequencing technique used is described in Chapter 2.

4.2.2.2.1 *AD369*

The *Adh* gene in *AD369* was completely sequenced in both strands from position -63 (30 bp upstream from the distal TATA box) to position +1781 (100 bp downstream from the stop codon at the third exon). Nucleotide numbering follows Kreitman (1983). A few sequence gaps on each strand were always overlapped by sequence from the other strand (Fig 4.10).

The obtained sequence was first compared with the consensus sequence for normal activity *Adh* alleles (Kreitman, 1983). The results are summarised in Table 4.1 and Appendix I. There were 20 nucleotide changes in *AD369* compared to the Kreitman (1983) consensus sequence, including 2 insertions and 18 single nucleotide substitutions. One of the insertions was the 1.2 kb insertion at position -5 which will be described in detail in the next chapter. The other was a small insertion of 34 bp instead of 29 bp in the normal sequence at the position +448. This 34 bp length polymorphism has been described in all but one of the 14 Adh^F lines that have been sequenced (Kreitman, 1983). It has been postulated to be responsible for the difference in ADH protein levels between the fast and slow allozymic classes (Laurie *et al*, 1991). There were 3 single nucleotide substitutions in the 5' flanking sequence (2 transitions and 1 transversion) which are common in Adh^F alleles (Kreitman,

1983). Six single nucleotide substitutions were present in the first intron (2 transitions and 4 transversions). Five of them (2 transitions and 3 transversions) have been described in other Adh^F alleles (Kreitman, 1983; Gibson *et al*, 1992). The one which is novel was a transversion from G to A at +586. There was a single nucleotide substitution from C to T at +925 in the second intron, which has been found in an Adh^F allele by Laurie *et al* (1991). Eight single nucleotide substitutions (1 transition and 7 transversions) were present in exon 2, exon 3, and the 3' untranslated region. Most of them were in the third codon position and did not give rise to amino acid substitutions in the encoded ADH protein. However, one substitution from A to C at position +1490 resulted in an amino acid substitution from *Lys* to *Thr* which characterises *Adh* fast alleles (Kreitman, 1983).

Most of the single nucleotide changes have been described before by Kreitman (1983), Laurie *et al* (1991) and Gibson *et al* (1992), and have been found in other Adh^F alleles, but the one in intron 1 (G to A at +586) is novel. The three nucleotides at positions +175, +516 and +816 (T, C and T) are the same as the consensus for the Adh^S allele, in contrast to most Adh^F alleles which usually have A, G and G respectively at these positions (Kreitman, 1983; Laurie *et al*, 1991; Gibson *et al*, 1992).

4.2.2.2.2 *AHA51*

The *Adh* gene of *AHA51* was completely sequenced on both strands from position -63 to position +1780 (Fig 4.11). Compared with the consensus sequence of normal activity *Adh* alleles (Kreitman, 1983), there were 14 single nucleotide changes. They are summarised in Table 4.2 and Appendix I. In the 5' flanking sequence region, besides the 2.4 kb insertion at position -5, which will be described in detail in the next chapter, there were 3 single nucleotide substitutions at positions -1, -2 and -3 (two transitions and 1 transversion). These three changes are common in Adh^F alleles but not in Adh^S alleles

(Kreitman, 1983; Laurie *et al*, 1991). In the first intron, there were 9 single nucleotide substitutions (3 transitions and 6 transversions). Three of them, at positions +107, +113 and +175, are common in Adh^F alleles, but not common in Adh^S alleles (Kreitman, 1983). The other three substitutions located at +143, +169 and +173 have been shown to be present in a few Adh^F alleles (Kreitman, 1983) but not in Adh^S alleles. Another three single nucleotide substitutions at +573, +576, +578 in the first intron and one in the larval leader sequence at +687, have not been previously described (Kreitman, 1983; Laurie *et al* , 1991). A substitution from G to A at +1596 in the third exon was in the third codon position and would not result in an amino acid substitution in the encoded ADH protein. The nucleotide at +1490 was A, which is characteristic of an Adh^S allele (Kreitman, 1983).

In summary, the nucleotide changes present upstream of +174 in *AHA51* are more similar to those in an Adh^F allele, but those in the other parts of the gene are more like those of an Adh^S allele. Therefore *AHA51* has a very distinctive haplotype combining sequences from Adh^F and Adh^S alleles.

4.2.2.2.3 *AAS44*

The *Adh* gene in *AAS44* was not completely sequenced. However, a region in the adult leader sequence from +288 to +409, and the exon region from the larval leader at +650 to the 3' flanking region at +1697 were sequenced (Fig 4.12) to see if there was any single nucleotide mutation that could give rise to an amino acid substitution and affect ADH activity. Compared with the consensus sequence (Kreitman, 1983), two nucleotide substitutions were detected in the first intron at +293 and +304, and they have been previously described in other Adh^F alleles (Kreitman, 1983). Seven transvertions were detected in exons (Table 4.3 and Appendix I). Most of these mutations were located in exon 3 and have been described before by Kreitman (1983). None of them would cause an amino acid substitution, besides the one at +1490 which

substituted C for A and is characteristic of an Adh^F allele. One single nucleotide substitution occurred at +816 in exon 1 and another one at +1693 in the 3' flanking sequence; they have both been described in other Adh^F alleles (Kreitman, 1983).

4.2.2.3 The *Adh* adult enhancer (AAE) region in the three variants

The AAE region (from -615 to -473, Falb and Maniatis, 1992a) in the three variants and in the two control alleles, *AC5* and *AC8* , was sequenced (from -640 to -460) to detect any nucleotide changes, as this is an important control region for adult ADH activity (Posakony *et al*, 1985). First, the AAE region was amplified with primers JG97/JG98 which spanned nucleotides -689 to -11 and included the AAE region (Fig. 4.13). The PCR results showed that only one band was amplified in each line and they did not reveal any differences in the size of the fragments. These PCR products were subcloned into pBluescript SK(+) and then sequenced. The region sequenced was about 180 bp (from -640 to -460), which spanned the 142 bp AAE sequences (from -615 to -473). The central core region of AAE, at which the transcriptional factors AEF-1 and C/EBP competitively bind to their adjacent sites (Falb and Maniatis, 1992a), is about 25 bp (from -534 to -510).

Of the five alleles I sequenced, the two control alleles, *AC5* and *AC8*, and the *AD369* allele did not differ from each other or from the standard sequence in the 142 bp AAE region (Falb and Maniatis, 1992; Kreitman and Hudson, 1991). The *AHA51* allele had one nucleotide change at position -488, but this was outside the central core region. The other allele, *AAS44*, had two nucleotide changes at position -510 and -524 — one in the AEF-1 binding site, and the other outside the central core region.

4.3 Discussion

The Southern analyses and the PCR data revealed that in *AHA51* and *AD369* there was an insertion in the same region between the *Sal*I (-1.3) and *Hpa*I (-0.9) sites, which spanned the distal promoter (Fig. 4.5). This 390 bp *SalI/Hpa*I fragment, which includes the distal TATA box and transcription initiation site (Posakony *et al*, 1985), is crucial to the distal promoter regulating the normal expression of the *Adh* gene in adults. Although the sizes of the insertions are different in the two variants, the phenotypes of the two variants are very similar.

The 1.2 kb insert in *AD369* was amplified under normal PCR reaction conditions (see 2.8.2 and Fig 4.7). The 2.4 kb insert in *AHA51* was amplified under modified conditions. The extension parameter allows the target DNA to extend 1 kilobase per minute (Erilich, 1989; McPherson *et al*, 1991). Thus, it is possible to amplify a 3 kb fragment by extending the annealing time and the extension time to 3 minutes. Under such improved conditions, the 2.4 kb insert in *AHA51* was amplified.

An interesting result is that, using *AHA51* DNA as a target with primers JG87/83 or JG87/17, two bands were produced (Fig 4.8 b and c). The size of the large band was equal to the size of the insertion plus the size of the normal region, while the small band had the expected size of the normal region without the insertion. The smaller band was amplified in the control *AC8*, but not in the negative control without a DNA template (Fig 4.8 c), so it therefore did not arise from contamination of PCR ingredients. It is possible that one of the primers could match non-specific sequences in the insertion and then give rise to a smaller band. Another possibility is that there were two templates in the DNA sample. This will be further studied in Chapter 6.

Using *AAS44* DNA as a template, the PCR with primers JG13/17, which spanned the insertion, did not amplify. However, Southern analysis showed that the insertion in *AAS44* was between *Hpa*I (-386 from proximal promoter)

and *Hin*dIII (-219 from proximal promoter) in the adult leader sequence (Fig 4.5). This region contains elements important for *Adh* proximal promoter transcription (Corbin and Maniatis, 1990).

There were also small insertions in *AD369* and *AAS44* located in the 1.45 kb *Eco*RI (-4.2, -2.8) fragment, 1320 bp 5' to the distal promoter. This region has been reported to contain the *Adh* larvae enhancer (ALE) (Corbin and Maniatis, 1990). In this region three elements of ALE act synergistically to achieve the maximal level of *Adh* transcription in larvae, but not in adults. The 1450 bp *Eco*RI fragment contains two elements of the *Adh* larval enhancer.

The large insertions found in these three low activity alleles are at a different location to the 4.5 kb insertion in the Adh^{SL} allele, which was about 3 kb 5' to the distal promoter (Schott *et al*, 1988) between *Eco*RI sites (-4.5 and -4.2). The small insertions found in *AD369* and *AAS44* are located 3' to this region between *Eco*RI sites (-4.2 and -2.8) (Fig 4.5). Previously, deletions have been found in this *Eco*RI (-4.2 and -2.8) region, but no insertions have so far been described (Aquadro *et al*, 1986; Jiang and Gibson, 1992b). Thus, although the phenotypes of the alleles are very similar the causes must be different.

Falb and Maniatis (1992a) defined a 142 bp element, the *Adh* adult enhancer (AAE) located within 600 bp of the distal promoter, that is necessary for distal promoter function. The nucleotide sequence of the AAE region in *AD369* allele is not changed, compared to the control alleles and to previously published sequences (Falb and Maniatis, 1992a; Kreitman and Hudson, 1991), and hence is irrelevant to its low activity phenotype. The single nucleotide change in this region present in the *AHA51* allele is unlikely to affect the AAE function, as it was outside the central core region. The changes in *AAS44* allele might affect the AAE function, because one of the two is located in the AEF-1 binding site, but this needs to be investigated.

The sequence of the three *Adh* variants showed that the nucleotide changes found did not produce a substitution in the coding region that could

be responsible for the low ADH activity in adults. Kreitman (1983) compared eleven cloned *Drosophila melanogaster Adh* genes from five natural populations. 43 nucleotide changes were revealed at the *Adh* gene and some of these changes also exist in the three alleles. In total there were 19 changes in *AD369* and 14 in *AHA51* compared to the consensus sequence from normal activity alleles (Kreitman, 1983). Most of the changes are in introns or in silent positions in exons. One change, at nucleotide +1490, results in the amino acid substitution responsible for the electrophoretic variants Adh^F and Adh^S; thus *AD369* is a fast allele, and *AHA51* is a slow allele. Most of the changes are in the first intron; 7 in *AD369* and 10 in *AHA51* out of 654 bp. In exon 3 of *AD369* there are 6 changes out of 388 bp, but only one change in *AHA51*. Overall there are more transvertions than transitions, 13:5 in *AD369* and 9:5 in *AHA51*. The ratio transvertions to transitions was 25:18 for the 43 changes observed by Kreitman (1983), which does not significantly differ from the ratios observed in *AD369* and *AHA51*. In *AD369* and *AHA51* the nucleotide sites at position -1, -2 and -3 immediately 5' to the adult mRNA cap site, is CGT, which mostly occur in Adh^F alleles (Kreitman, 1983). The *AD369* variant, like most Adh^F alleles, has a sequence length polymorphism, ∇1, within the first intron of the distal (adult) transcript (Laurie *et al*, 1991). *AHA51*, an Adh^S allele, does not have this ∇1 polymorphism.

The ADH allozyme polymorphism has been intensively studied at the molecular level. A cluster of six polymorphic sites between 1443 and 1557 show strong linkage disequilibrium -- three out of the six nucleotide substitutions distinguish all Fast alleles from all Slow alleles. One is the amino acid replacement at 1490 and the others are nearby third position silent substitutions, one at 1443 and one at 1527 (Laurie-Ahlberg and Stam, 1987). More sequence comparisons eliminated the 1527 silent substitution as responsible for the activity difference as few slow alleles with typical Slow-like low expression showed the Fast-typical C at this site (Laurie *et al*, 1991). The

1490 amino acid replacement affects ADH activity but not CRM level; the 1443 silent substitution has no detectable effect on activity or CRM level. Additional sequence data have shown that a unique sequence length polymorphism ∇1 within the adult intron is strongly associated with the amino acid replacement polymorphism and with ADH CRM level (Laurie *et al*, 1991).

The changes present in the sequence upstream of position +175 in *AHA51*, which included the consecutive varying sites in the 5' flanking region and the 6 single nucleotide substitutions in the first intron, are more like those of an Adh^F allele. Whilst the changes that occurred downstream of position +175 in the rest of the gene are more in common with other Adh^S alleles. Particularly, the lack of ∇1 in the first intron and the A instead of C at position +1490, which would encode *Lys* at amino acid residue 192, characterises *AHA51* as an Adh^S allele. Four changes at position +573, +576, +578 and +687 in the first intron have not been described before in any other Adh^S allele. It is possible that recombination occurred between the *AHA51* gene and the *CyO* gene, as the *AHA51* allele was maintained heterozygous with *CyO* (See 2.1). The recombination could also have occurred before the *AHA51* allele was isolated from natural population, and this seems more likely.

The nucleotide changes present in *AD369* are similar to Adh^F alleles, including the consecutive varying sites in the 5' flanking region , the ∇1 in the first intron and the nucleotide substitution from A to C at position +1490 which gives *Thr* at amino acid residue 192 (Benyajati *et al*, 1981). A substitution at +423 from A to T has been detected in some Australian Adh^F alleles (Gibson *et al*, 1992). Three sites at +175, +516 and +816, are the same as the consensus, although they are usually changed in Adh^F alleles. The change at position +586 has not been described in other Adh^F alleles (Kreitman, 1983; Laurie *et al*, 1991).

Molecular structure analyses have revealed insertions at the *Adh* locus in each variant. Therefore, it is reasonable to propose that the decreased ADH

activity in these variants is caused by the presence of these insertions. Speculation on the mechanisms of these effects will be considered later when the insertions have been investigated in detail. However, insertions have been shown to be the cause of reduced activity in other systems. One *Adh*^{S} line has been shown to have a 5.2 kb *copia* element insertion about 250 bp 5' to the distal cap site of *Adh* and appears to be associated with a low level of ADH protein (Aquadro *et al*, 1986; Strand and McDonald, 1989). Another *Adh*^{F} line, which was derived from a natural population, has been shown to have a 1.5 kb insertion 2 kb from the 5' end of the coding region and have lower ADH activity (Jiang and Gibson, 1992b). Insertions that cause altered enzyme expression have been found at other gene loci in *Drosophila*. For instance, a defective *hobo* at *Sgs-4* locus (McGinnis *et al*, 1983); the KP element insertion at *Gpdh* locus (Reed and Gibson, 1993); fourteen different transposable element insertions at the *white* locus (Bingham *et al*, 1981; Rubin *et al*, 1982; Collins and Rubin, 1982; Zachar and Bingham, 1982; Levis and Rubin, 1982) and the *gypsy* transposable element present in 5' region of the *yellow* gene and in an intron of the *forked* gene (Parkhurst and Corces, 1985).

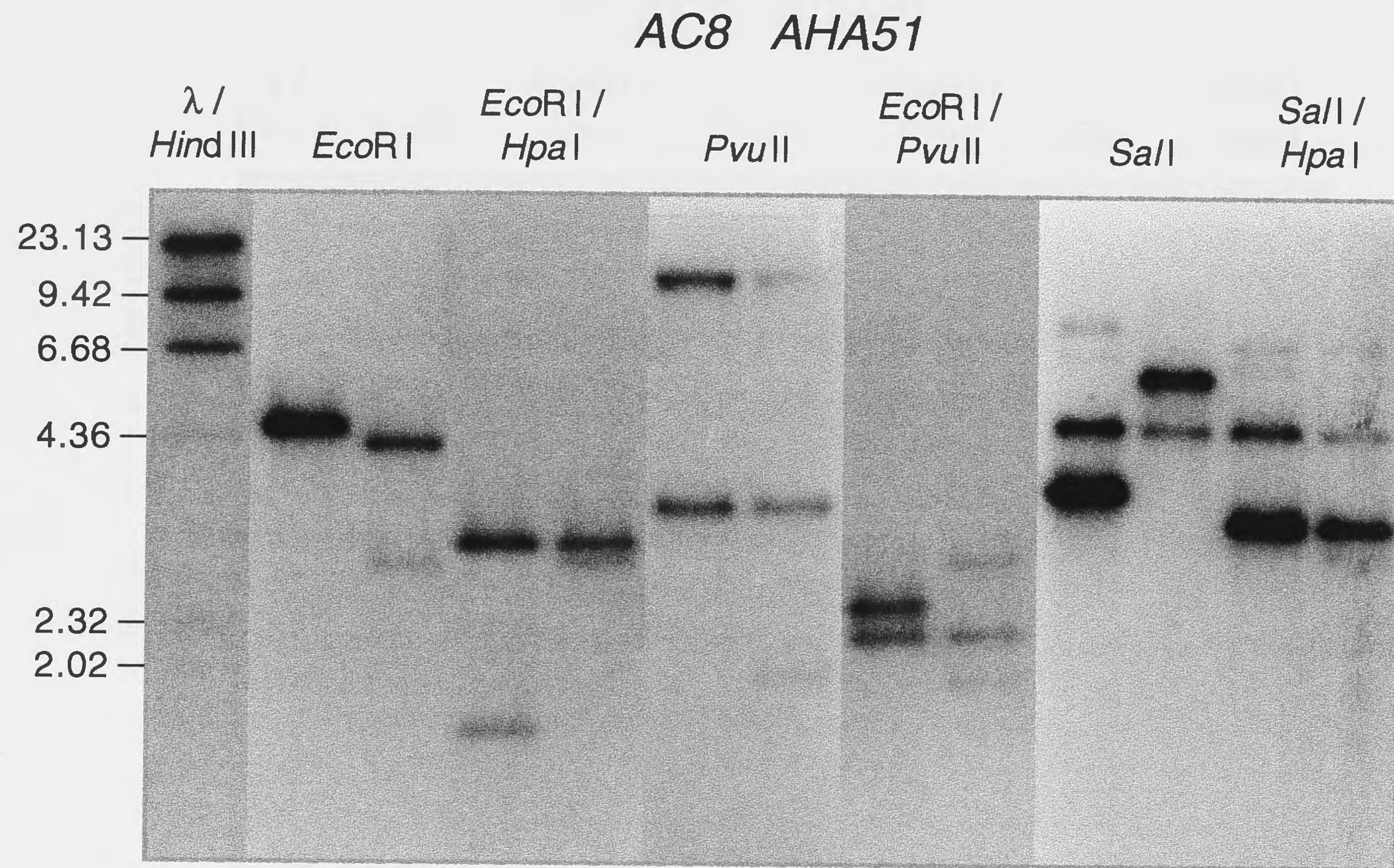


Figure 4.1 Southern blots for *AHA51* and *AC8*

DNA samples were digested with the enzymes indicated. DNA from *AC8* was loaded in the left lane, and DNA from *AHA51* in the right lane of each pair. The samples are taken from different gels run for the same time, and arranged in this way for clarity. Only the informative digests are shown. Lambda DNA digested with *Hin*dIII was used as a size marker and the sizes are in kilobase pairs. The probe used was *pSAC1* (Goldberg, 1980).

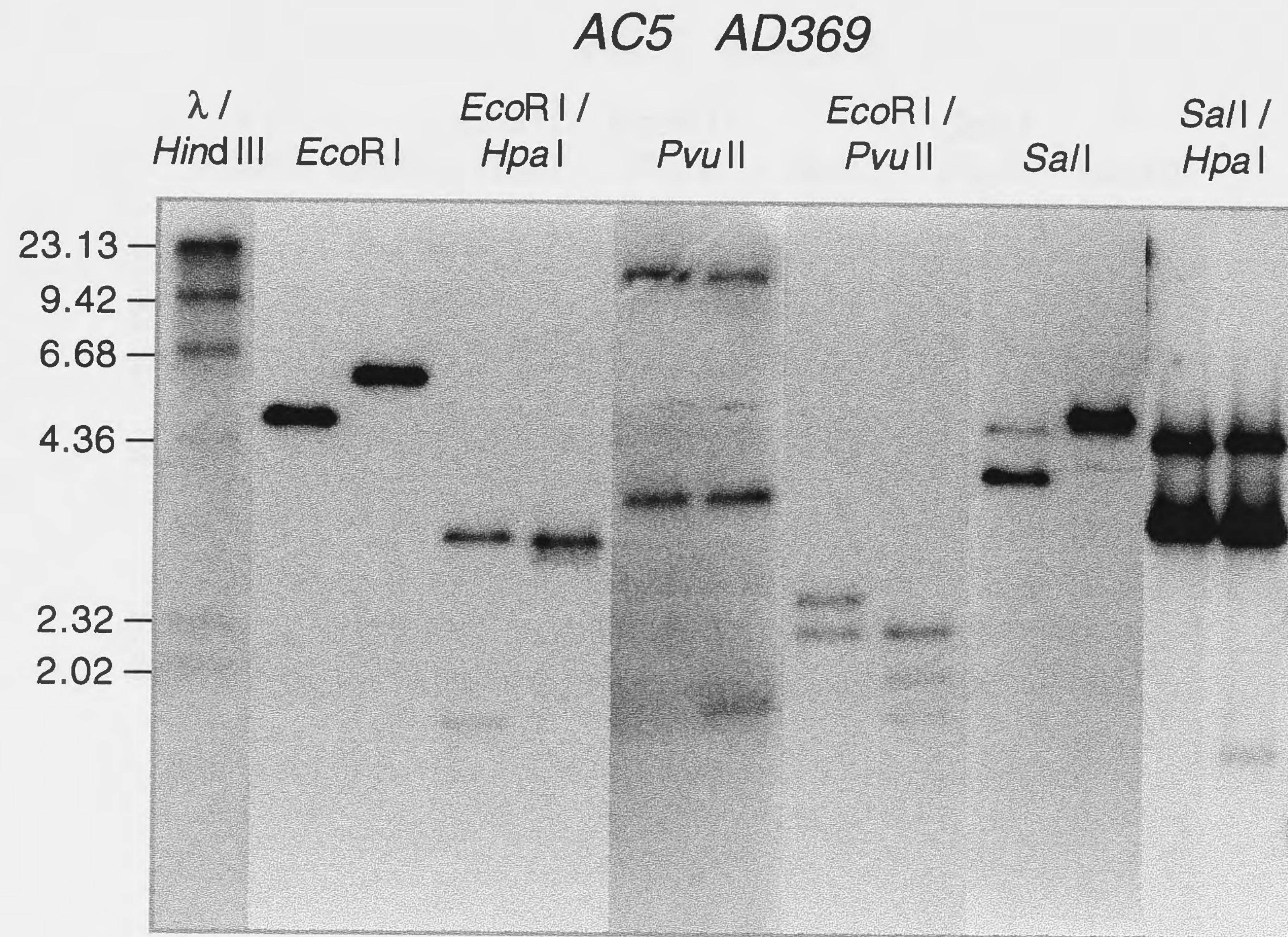


Figure 4.2 Southern blots for *AD369* and *AC5*

DNA samples were digested with the enzymes indicated. DNA from *AC5* was loaded in the left lane, and DNA from *AD369* in the right lane of each pair. The samples are taken from different gels run for the same time, and arranged in this way for clarity. Only the informative digests are shown. Lambda DNA digested with *Hind*III was used as a size marker and the sizes are in kilobase pairs. The probe used was *pSAC1* (Goldberg, 1980).

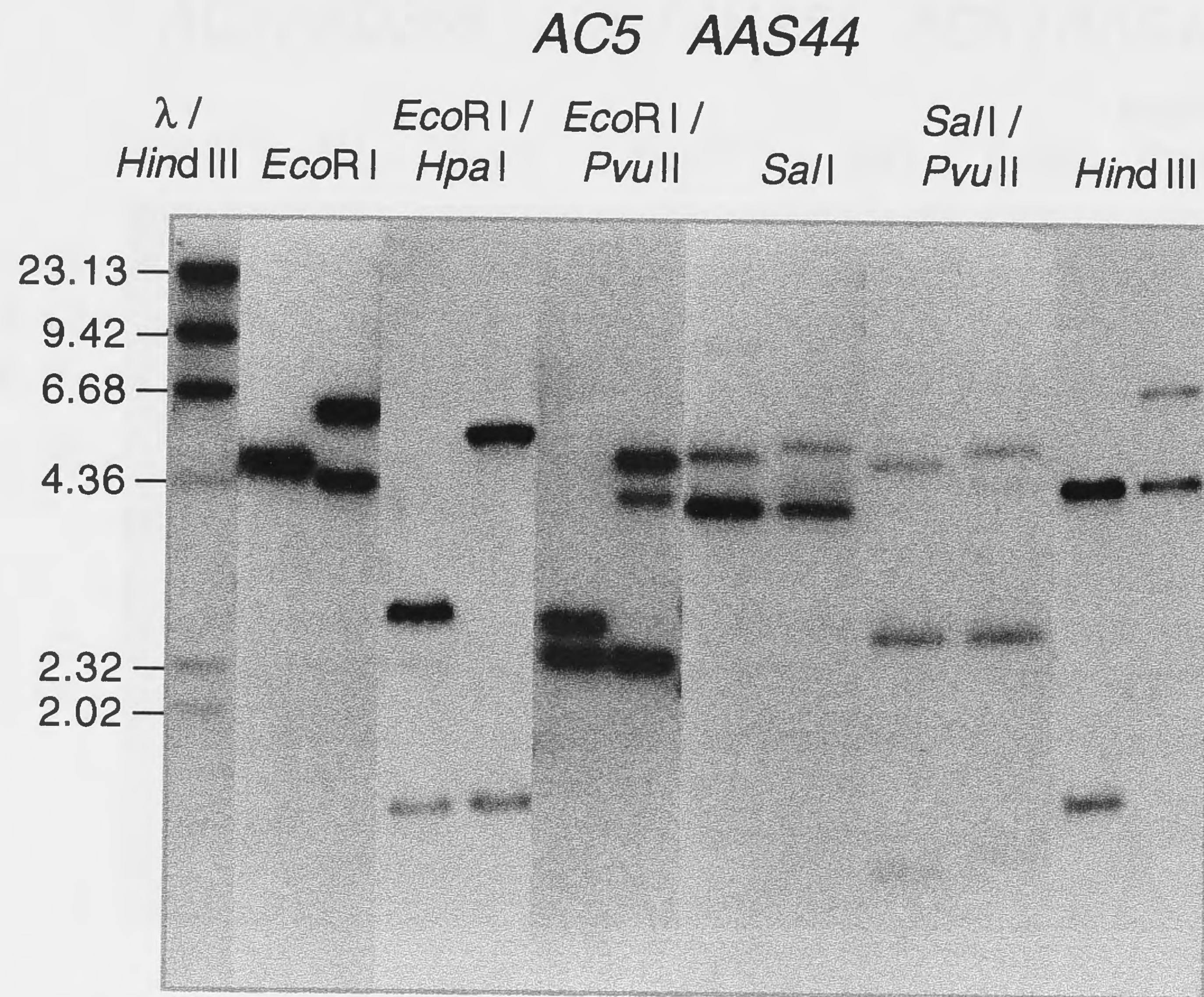


Figure 4.3 Southern blots for *AAS44* and *AC5*

DNA samples were digested with the enzymes indicated. DNA from *AC5* was loaded in the left lane, and DNA from *AAS44* in the right lane of each pair. The samples are taken from different gels run for the same time, and arranged in this way for clarity. Only the informative digests are shown. Lambda DNA digested with *Hin*dIII was used as a size marker and the sizes are in kilobase pairs. The probe used was *pSAC1* (Goldberg, 1980).

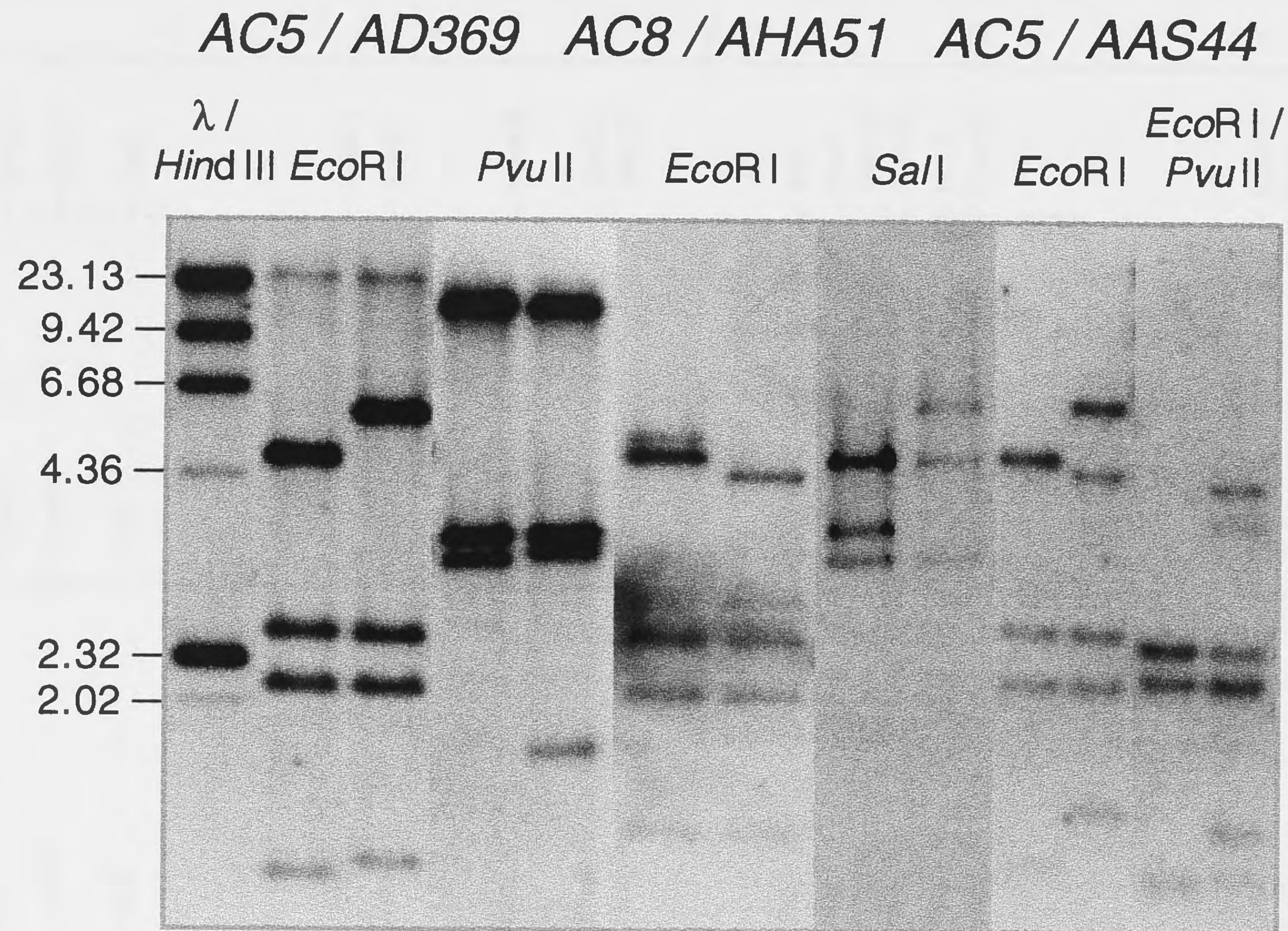


Figure 4.4 Southern blots for three variants probed with *pSAF2*

DNA samples were digested with the enzymes indicated. DNA from *AC5* or *AC8* was loaded in the left lane, and DNA from *AD369, AHA51* or *AAS44* in the right lane of each pair. The samples are taken from different gels run for the same time, and arranged in this way for clarity. Only the informative digests are shown. Lambda DNA digested with *Hin*dIII was used as a size marker and the sizes are in kilobase pairs. The probe used was *pSAF2* (Goldberg, 1980).

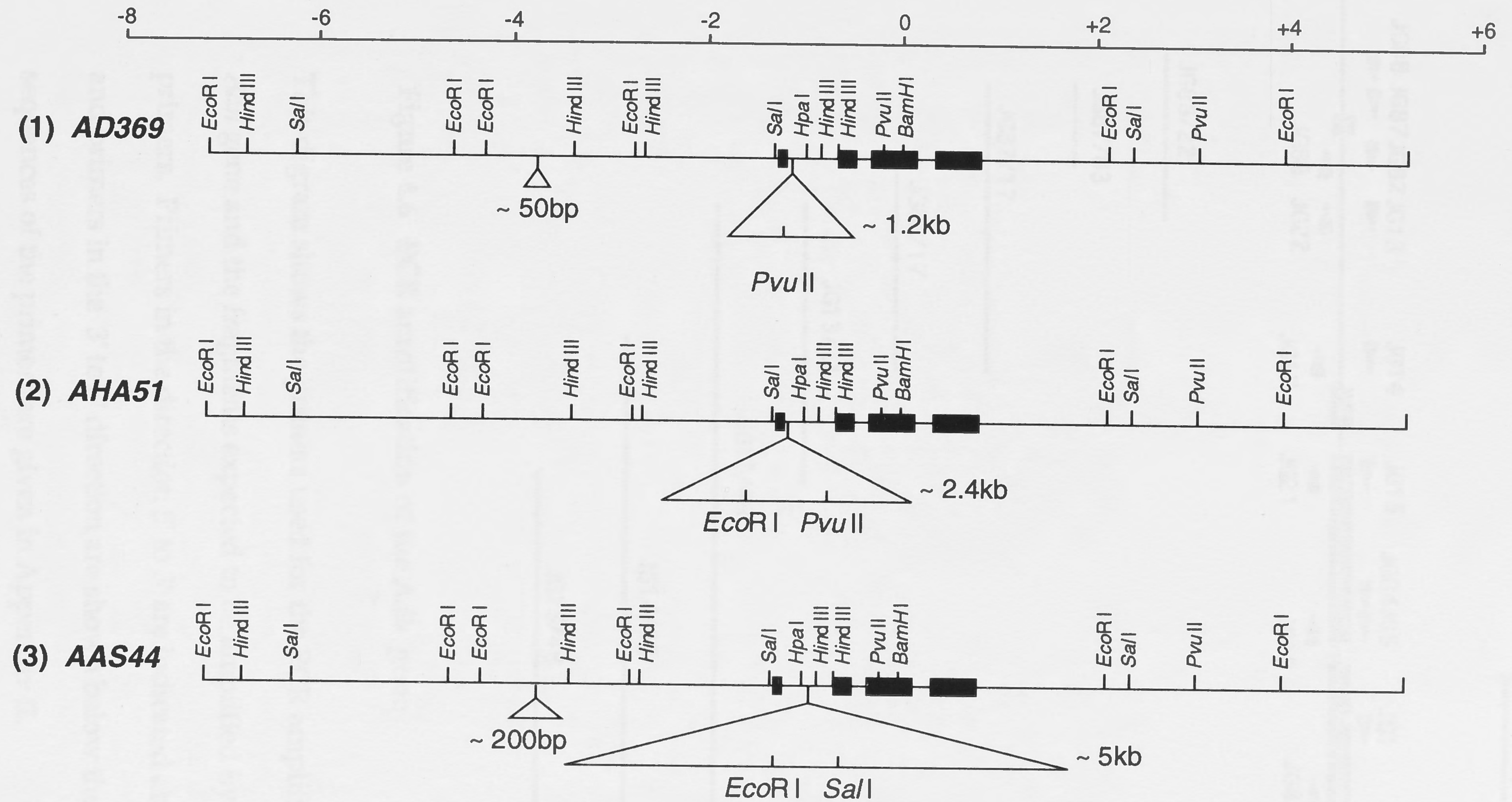


Figure 4.5 Restriction maps of the three variants.

The results obtained from the Southern analysis of the three variants are summarised as restriction maps of the three *Adh* genes. Exons are drawn as black boxes and insertions as triangles.

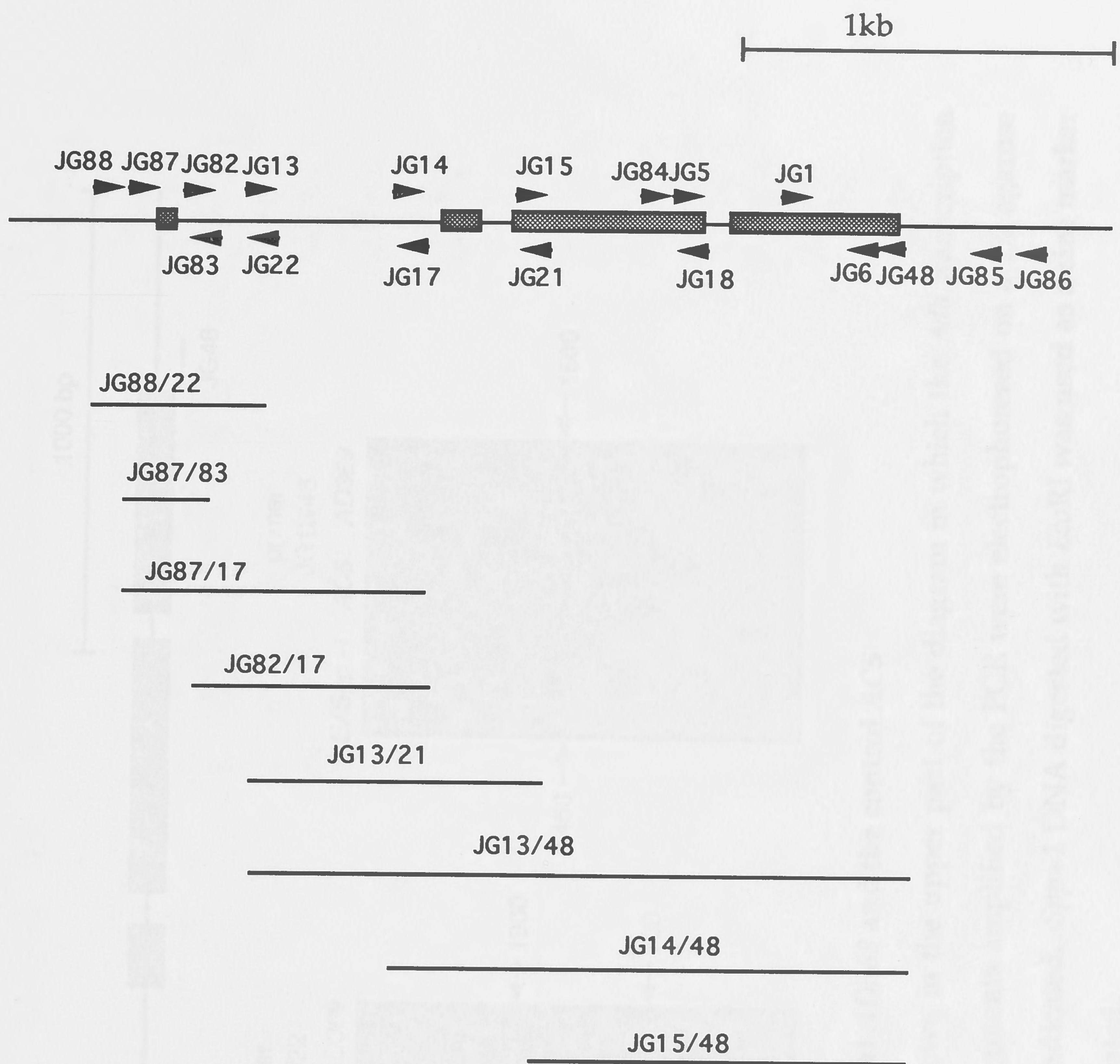


Figure 4.6 PCR amplification of the *Adh* gene

This digram shows the primers used for the PCR amplification of the *Adh* gene and the fragments expected to be amplified by each pair of primers. Primers in the direction 5' to 3' are indicated above the gene and primers in the 3' to 5' direction are shown below the gene. The sequences of the primers are given in Appendix II.

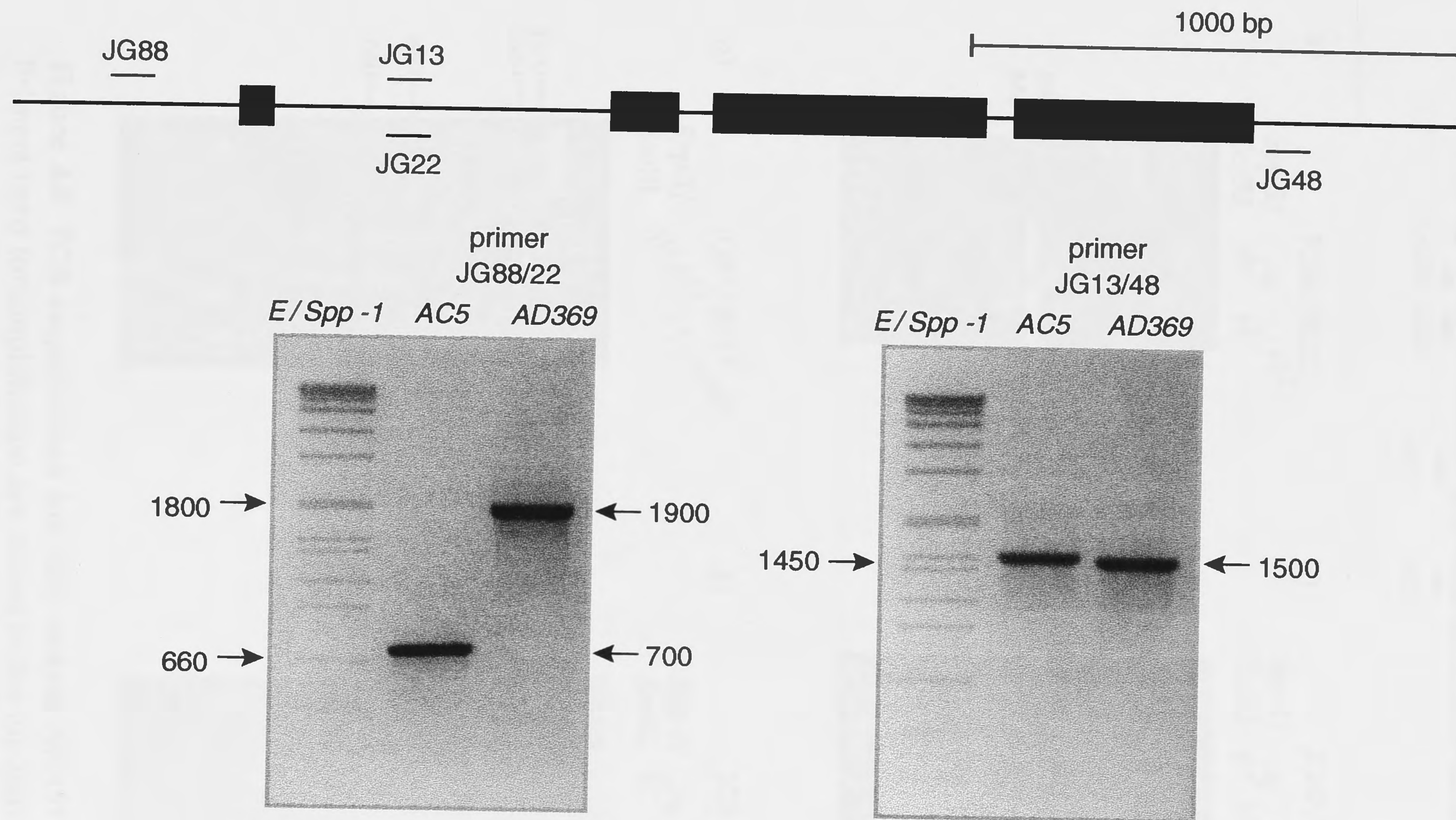


Figure 4.7 PCR amplification of the variant *AD369* and the control *AC5*

The primers used for amplification are shown in the upper part of the diagram in which the *Adh* transcription unit is indicated as black boxes. DNA fragments amplified by the PCR were electrophoresed on a 1% agarose gel. DNA samples and primers used are indicated. *Spp*-1 DNA digested with *Eco*RI was used as a size marker and sizes of the DNA in basepairs are indicated.

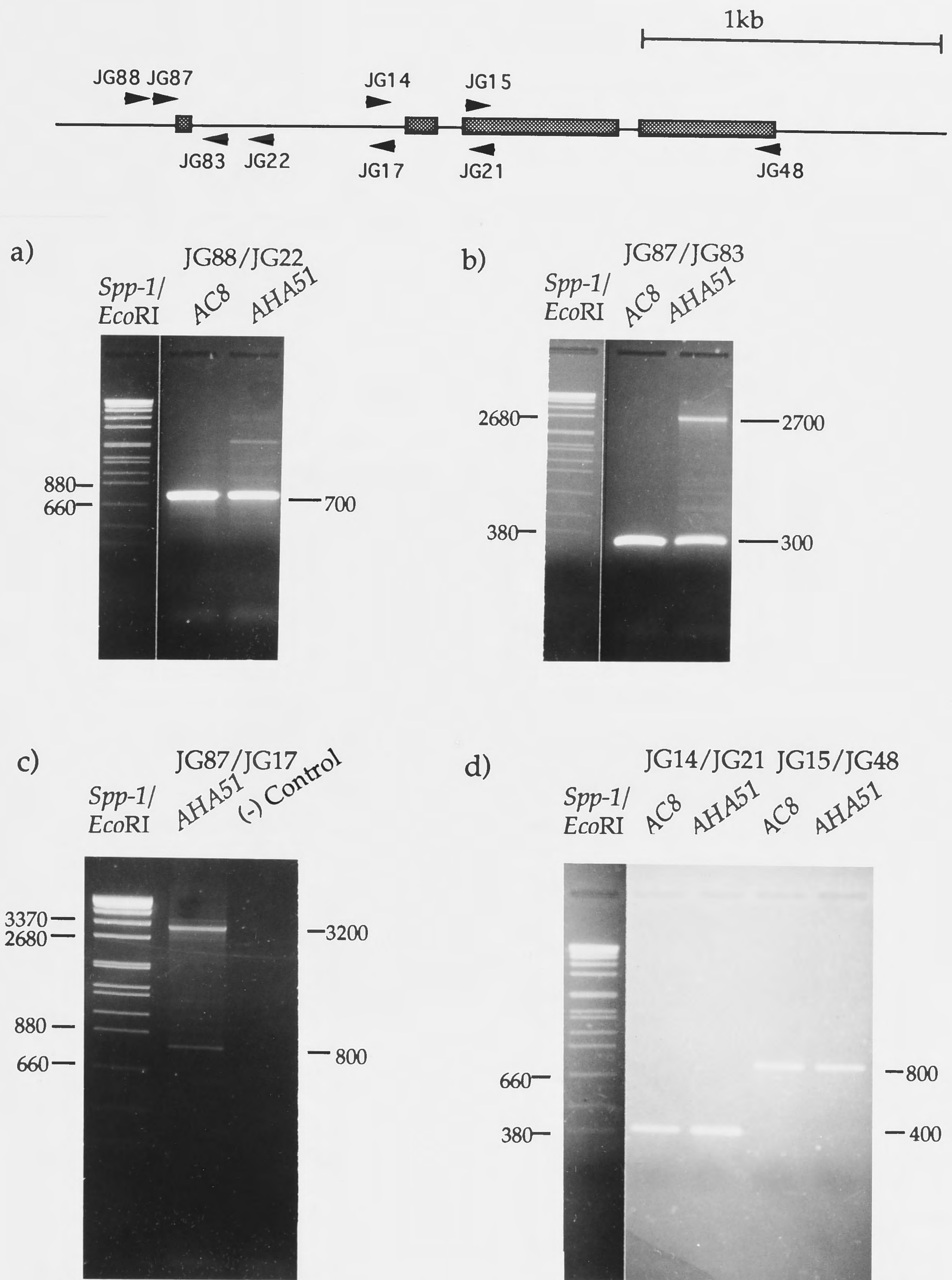


Figure 4.8 PCR amplification for the variant *AHA51* and the control *AC8*. Primers used for amplification are shown in the top part of the diagram. DNA fragments amplified by the PCR were electrophoresed on a 1% agarose gel and the estimated sizes of the fragments are indicated in basepairs (right). DNA samples and primers used are shown above each picture. *Spp-1* DNA digested with *Eco*RI was used as a size marker and the sizes of DNA are in basepairs (left).

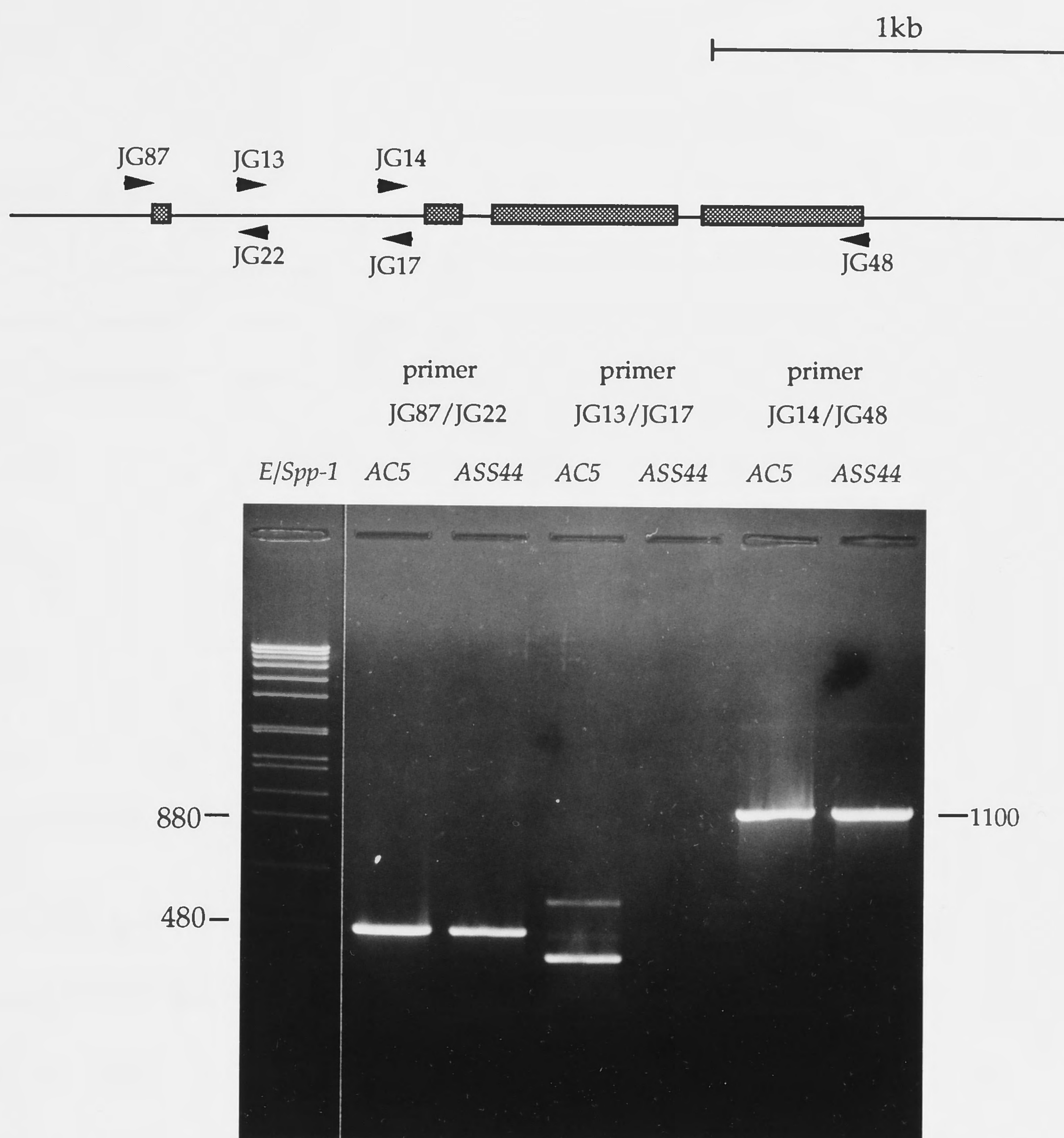


Figure 4.9 PCR amplification for the variant *AAS44* and the control *AC5*. Primers used for amplification are shown in the upper part of the diagram. DNA fragments amplified by the PCR were electrophoresed on a 1% agarose gel and the estimated size of the fragments is indicated in basepairs. DNA samples and primers used are indicated above the picture. The *Spp-1* DNA digested with *Eco*RI was used as a size marker and the sizes of DNA are in basepairs.

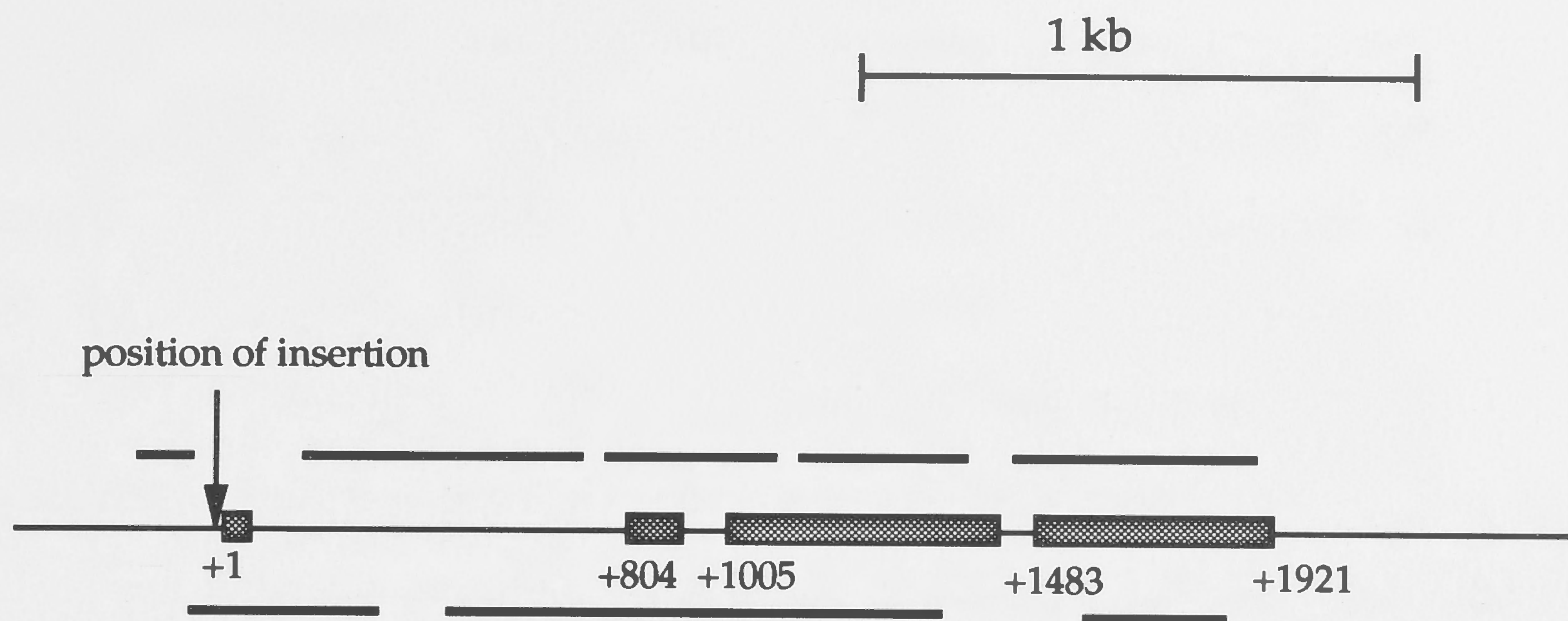


Figure 4.10 Strategy for sequencing the *Adh* gene in *AD369*

The *Adh* gene in *AD369* was sequenced in at least one direction and about 80% of the sequence was analysed on both strands from position -63 to +1781. The regions sequenced in direction 5' to 3' are indicated with a line above the diagram of the gene, whilst those sequenced in the direction 3' to 5' are indicated with a line below the gene.

Table 4.1 Sequence comparison of *AD369* with the consensus sequence from Adh^S and Adh^F (from Kreitman 1983).

In the consensus sequences for the Adh^S and Adh^F alleles, alternate nucleotides found are indicated, eg. at position -3 "T/.", it means most Adh^F are T at this site, but some are the same as the overall consensus. "." refers to a match with the consensus at the indicated nucleotide site. "*" represents the S/F allozyme site. The novel change is indicated in red. Changes in *AD369* which differ from the Adh^F consensus, but occur in other Adh^S alleles, are indicated in blue. Changes in *AD369* which are rare in other Adh^F, but common in Adh^S alleles are indicated in green.

Reference Sequence	Kreitman's No	Consensus	*Adh-S*	*Adh-F*	*AD369*
	-3	C	.	T/.	T
5' Flanking	-2	C	.	G/.	G
Sequence	-1	G	C/.	C	C
	107	C	./A	A/.	A
	113	A	./G	G/.	G
	143	A	.	./G	.
	169	T	.	./G	.
	173	A	.	./G	.
Intron 1	175	T	./A	A/.	.
(Adult Intron)	287	G	./T	.	.
	293	G	./T	T/.	T
	304	G	./C	C/.	C
	423	A	.	.	T
	447-476	29bp	.	34bp	34bp
	516	C	.	G/.	.
	550	no insertion2	.	insterion/.	.
	586	G	.	./T	A
Larval Leader	713	C	./A	.	.
and Exon 2	816	T	./G	G	.
	896	A	./G	.	.
Intron 2	925	C	./T	.	T
	1068	C	./T	.	.
Exon 3	1229	C	./T	.	.
	1235	C	.	./A	A
	1283	C	./A	.	.
	1354	G	./C	.	.
	1362	G	./A	.	.
Intron 3	1388	A	.	./G	.
	1400	A	./T	.	.
	1405	T	./A	.	.
	1425	C	./A	.	.
	1431	T	./C	.	.
	1443	C	.	G	G
Translated	1452	C	./T	T	T
Region of	1490*	A*	.*	C*	C*
Exon 4	1518	C	./T	T	T
	1527	T	.	C	C
	1557	A	./C	C	C
	1596	G	./A	.	.
3'Untranlated region	1693	A	./C	C	C

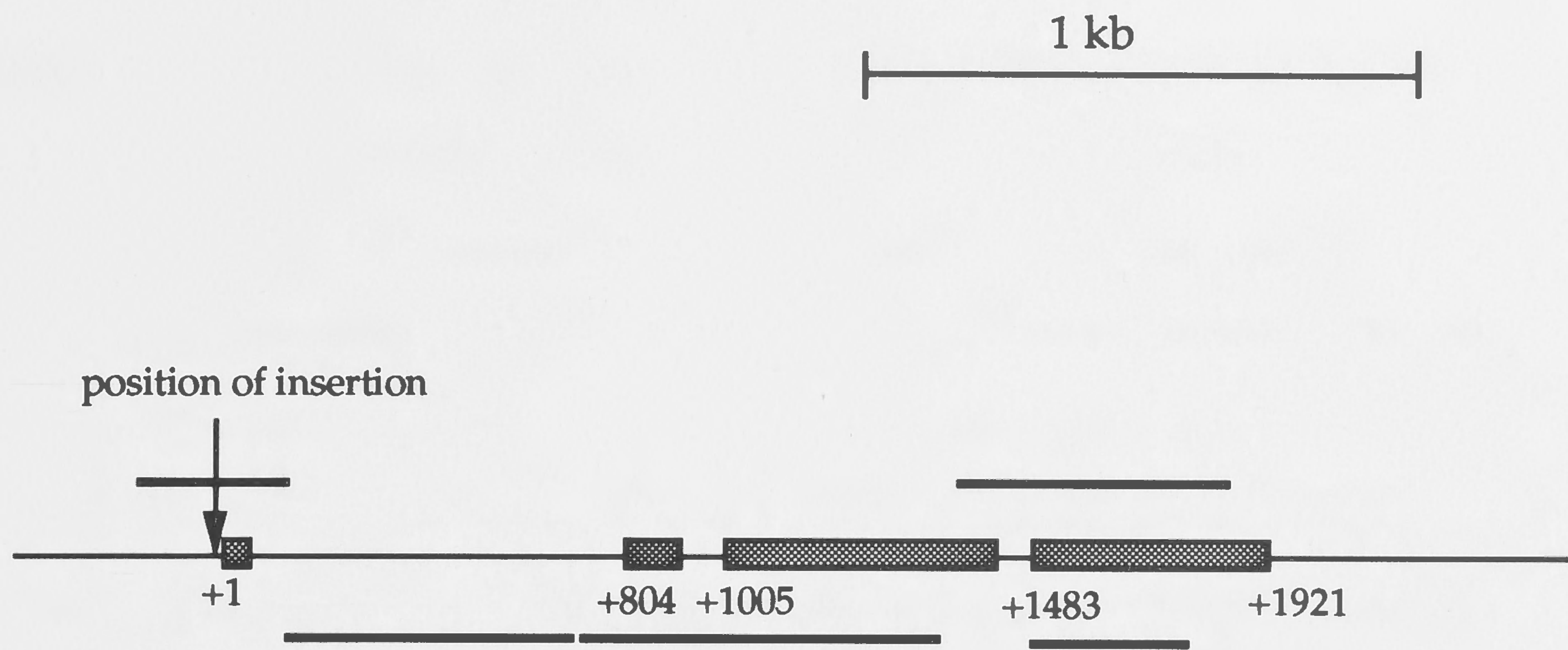


Figure 4.11 Strategy for sequencing the *Adh* gene in *AHA51*

The *Adh* gene in *AHA51* was sequenced either 5' to 3' or 3' to 5' from position -63 to +1780, and about 20% of the gene was sequenced in both directions. The regions sequenced in direction 5' to 3' are indicated with a line above the diagram of the gene, whilst those sequenced in the direction 3' to 5' are indicated with a line below gene.

Table 4.2 Sequence comparison of *AHA51* with the consensus sequence from Adh^S and Adh^F (from Kreitman 1983).

In the consensus sequences for the Adh^S and Adh^F alleles, alternate nucleotides found are indicated, eg. at position -3 "T/.", it means most Adh^F are T at this site, but some are the same as the overall consensus. "." refers to a match with the consensus at the indicated nucleotide site. "*" represents the S/F allozyme site. The novel changes are indicated in red. Changes in *AHA51* which differ from the Adh^S consensus, but occur in other Adh^F alleles are indicated in bule. Changes in *AHA51* which are rare in other Adh^S, but common in Adh^F alleles, are indicated in green.

Refer.ence Sequence	Kreitman's No	Consensus	*Adh-S*	*Adh-F*	*AHA51*
	-3	C	.	T/.	T
5' Flanking	-2	C	.	G/.	G
Sequence	-1	G	C/.	C	C
	107	C	./A	A/.	A
	113	A	./G	G/.	G
	143	A	.	./G	G
	169	T	.	./G	G
	173	A	.	./G	G
Intron 1	175	T	./A	A/.	A
(Adult Intron)	287	G	./T	.	.
	293	G	./T	T/.	.
	304	G	./C	C/.	.
	423	A	.	.	.
	447-476	29bp	.	34bp	.
	516	C	.	G/.	.
	550	insertion2	.	insterion/.	.
	573	G	.	.	A
	576	G	.	.	C
	578	C	.	.	G
	586	G	.	./T	.
	687	G	.	.	A
Larval Leader	713	C	./A	.	.
and Exon 2	816	T	./G	G	.
	896	A	./G	.	.
Intron 2	925	C	./T	.	.
	1068	C	./T	.	.
Exon 3	1229	C	./T	.	.
	1235	C	.	./A	.
	1283	C	./A	.	?
	1354	G	./C	.	?
	1362	G	./A	.	?
Intron 3	1388	A	.	./G	?
	1400	A	./T	.	.
	1405	T	./A	.	.
	1425	C	./A	.	.
	1431	T	./C	.	.
	1443	C	.	G	.
Translated	1452	C	./T	T	.
Region of	1490*	A*	.*	C*	.*
Exon 4	1518	C	./T	T	.
	1527	T	.	C	.
	1557	A	./C	C	.
	1596	G	./A	.	A
3'Untran. region	1693	A	./C	C	.

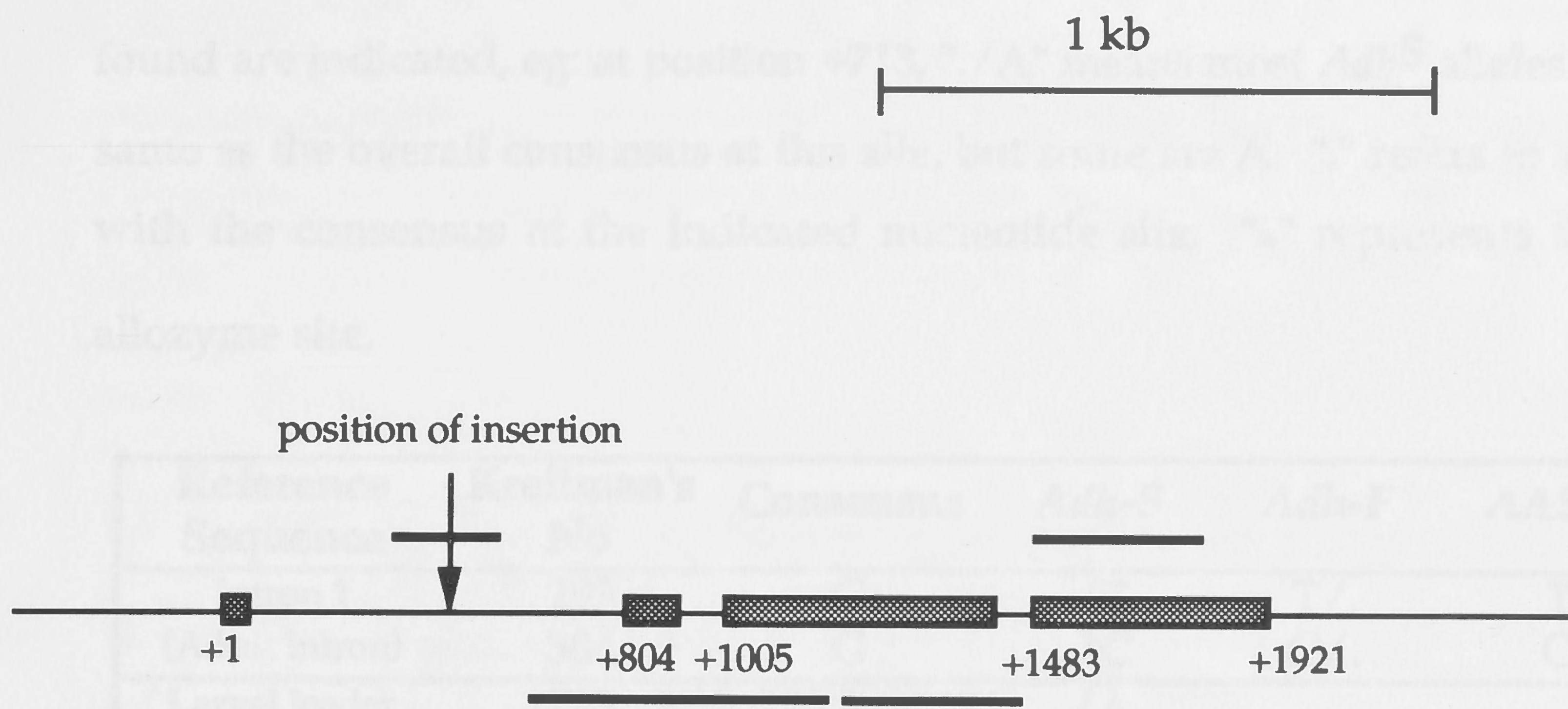


Figure 4.12 Strategy for sequencing the *Adh* gene in *AAS44*

The *Adh* gene in *AAS44* was partialy sequenced either 5' to 3' or 3' to 5' in the region which spanned the exons from position +288 to +409 and +650 to +1697. The regions sequenced in direction 5' to 3' are indicated with a line above the diagram of the gene, whilst those sequenced in the direction 3' to 5' are indicated with a line below the gene.

Table 4.3 Sequence comparison of *ASS44* with the consensus sequence from *Adh*^{S} and *Adh*^{F} (from Kreitman 1983).

In the consensus sequences for the *Adh*^{S} and *Adh*^{F} alleles, alternate nucleotides found are indicated, eg: at position +713, "./A" means most *Adh*^{S} alleles are the same as the overall consensus at this site, but some are A. "." refers to a match with the consensus at the indicated nucleotide site. "*" represents the S/F allozyme site.

Reference Sequence	Kreitman's No	Consensus	*Adh-S*	*Adh-F*	*AAS44*
Intron 1	293	G	./T	T/.	T
(Adult Intron)	304	G	./C	C/.	C
Larval leader	713	C	./A	.	.
Exon 2	816	T	./G	G	G
	896	A	./G	.	.
Intron 2	925	C	./T	.	.
	1068	C	./T	.	.
Exon 3	1229	C	./T	.	.
	1235	C	.	./A	.
	1283	C	./A	.	
	1354	G	./C	.	
	1362	G	./A	.	
Intron 3	1388	A	.	./G	.
	1400	A	./T	.	.
	1405	T	./A	.	.
	1425	C	./A	.	.
	1431	T	./C	.	.
	1443	C	.	G	.
Translated	1452	C	./T	T	T
Region of	1490*	A*	.*	C*	C*
Exon 4	1518	C	./T	T	T
	1527	T	.	C	C
	1557	A	./C	C	C
	1596	G	./A	.	.
3'Untranlated region	1693	A	./C	C	C

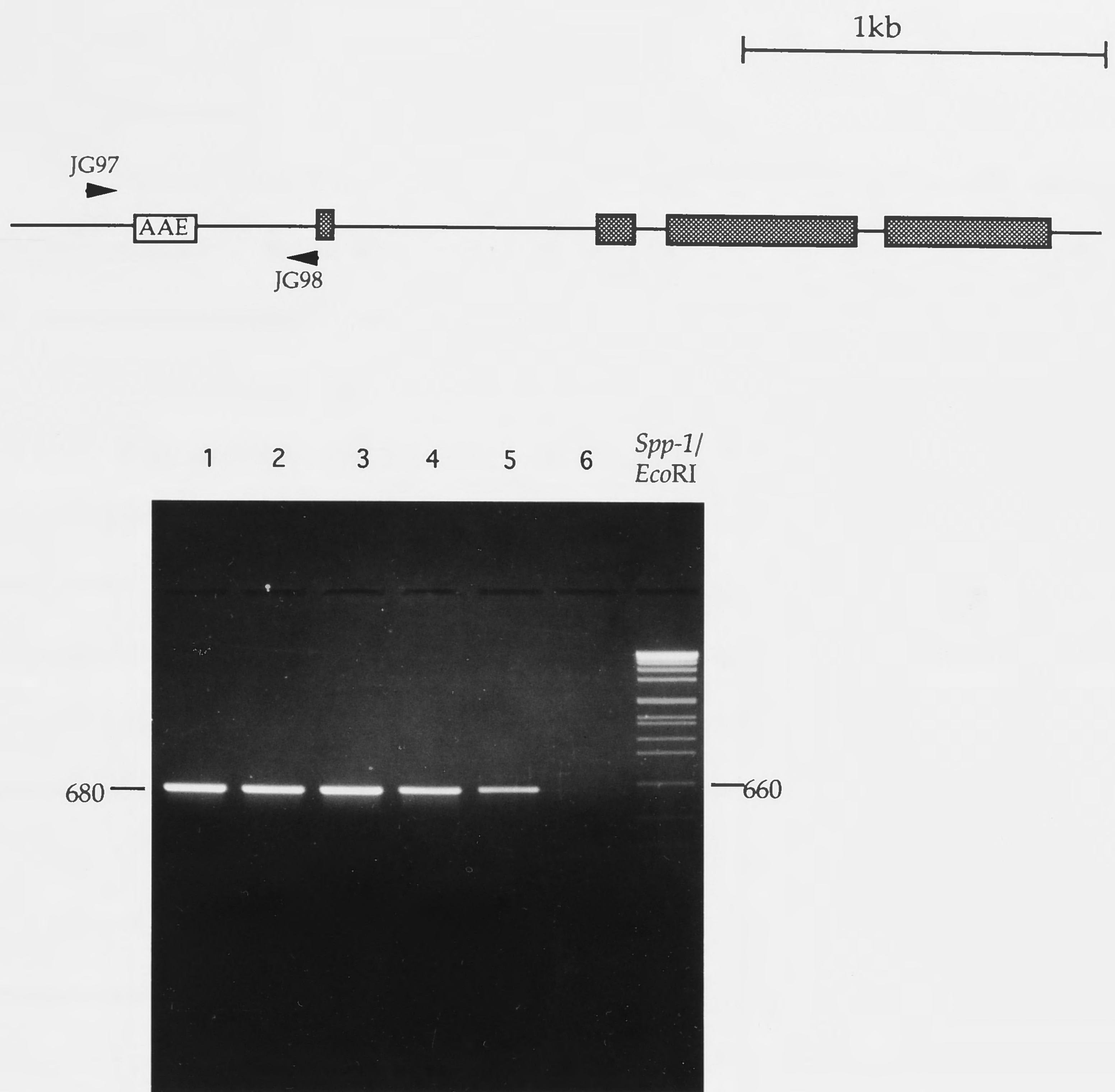


Figure 4.13 PCR amplification of the *Adh* adult enhancer (AAE) region in the low activity variants and in the control alleles.

The AAE region (Falb and Maniatis 1992) was amplified with the primers JG97/JG98 indicated in the above diagram. DNA fragments amplified by the PCR were electrophoresed on a 1% agarose gel and the estimated size of the fragments is indicated in basepairs. The DNA samples for the amplification were: lane 1, *AC8;* lane 2, *AHA51;* lane 3, *AD369;* lane 4, *AAS44;* lane 5, *AC5* and lane 6, no DNA control. *Spp-1* DNA digested with *Eco*RI was used as a size marker and DNA sizes are in basepairs.

Chapter 5 Molecular characterisation of insertions at the *Adh* locus in the three low activity variants

5.1 Introduction

The three low ADH activity variants studied in this work, *AD369*, *AHA51* and *AAS44*, were all isolated from natural populations. In surveys of restriction endonuclease map variation in the *Adh* region of chromosome from natural populations of *Drosophila melanogaster*, a large number of insertions have been found, but not all have been identified (Langley *et al*, 1982; Cross and Birley, 1986). Aquadro *et al* (1986) demonstrated that seven inserts of different sizes observed in 11 lines had sequence homology to four different transposable elements of *copia*, *B104*, *2161* and *F101*. Jiang and Gibson (1992a) revealed ten different insertions at the *Adh* region in 104 lines from Australian and 90 lines from Chinese natural populations of *Drosophila melanogaster*. These insertions were sub-cloned and probed with DNA sequences from each of seven known mobile elements. The results showed that two of the cloned insertions shared homology with the known mobile elements *B104B* and *F101*. But the rest of the insertions have not been identified. Some of the insertions in the *Adh* region were shown to be associated with lowered ADH activity (Jiang and Gibson, 1992b). For example, the variant studied in this work, *AAS44* , with a 5 kb insertion in the first intron, has been shown to have low ADH activity (Jiang and Gibson, 1992b). Another Adh^F line with a 1.5 kb insertion, located 2 kb from the 5' end of the coding region, also has lower ADH activity. An *Adh* variant with a *copia* retroviral-like transposable element inserted 240 bp upstream from the distal (adult) *Adh* transcriptional start site has reduced *Adh* transcript levels in late third instar larvae and in adults (Dunn and Laurie, 1995).

It is possible that the large insertions that I have found at the *Adh* locus are partial duplications of the gene. Such a situation has been observed in one

line (from Tamar) that has a partial duplication of the *Adh* coding region together with a duplication of a region approximately 4 kb upstream to the coding region (Jiang and Gibson, 1992b). Another case reported is the *white-ivory* (w^i) spontaneous mutation which is caused by a 2.95 kb internal duplication of the *w* locus (Karess and Rubin, 1982).

The insertions that I have demonstrated also could be some sort of transposable element, because most spontaneous mutations in *Drosophila* are caused by the insertion of transposable elements (Finnegan and Fawcett, 1986; Boeke, 1989; Smith and Corces, 1991). A large variety of transposable elements have been described in *Drosophila* and they can be classified into two main categories, retrotransposons (for reviews, see Bingham and Zachar, 1989) and non-retrotransposons. The non-retrotransposons include P-elements (O'Hare and Rubin, 1983), *hobo* (Streck *et al*, 1986), *mariner* (Jacobson *et al*, 1986) and foldback (FB) elements (Truett *et al*, 1981). These transposons differ in their size, structure, mechanism of transposition, integration specificity and mechanism of mutagenesis (Smith and Corces, 1991). Some of the transposable element insertions have been shown to affect the activity of the loci into which they insert. Defective P element insertions in the 5' region of the *Gpdh* locus cause a distinctive low-activity (Reed and Gibson, 1993). A variant at the glue protein gene *Sgs-4* locus has an insertion of a 1.3 kb defective *hobo* element just upstream of the TATA box, leading to reduced gene expression with multiple new transcription start sites and alterations in chromatin structure (McGinnis *et al*, 1983). One mutant with the *gypsy* element inserted 5' to the coding region of the *yellow* gene has reduced levels of *yellow* transcripts at a life stage in which *gypsy* elements are transcriptionally active (Parkhurst and Corces, 1986a). A retrotransposon element, *copia*, has been found inserted upstream from the distal (adult) *Adh* transcriptional start site (Strand and McDonald, 1989). None of the insertions so far identified in the *Adh* region shares homology with either the P or the *hobo* element.

The phenotypic effects caused by the insertion of a transposable element depend not only on the site of insertion — in exons, introns or regulatory regions — but also on the characteristics of the insertions (Engels, 1989). In *Drosophila,* transposable elements may insert into genes and cause mutant phenotypes by a variety of mechanisms, the nature of which will depend on the location of the insertion site with respect to the structure and functional domains of the targeted gene. For example, insertions into the protein-coding region of a gene may result in an unstable transcript or a truncated protein. Insertions into a splice consensus sequence may affect the processing of the message RNA. Those into the sequence necessary for transcription initiation may affect the transcription level of the gene (Smith and Corces, 1991). Insertions in the promoter region have been shown to affect the expression of some genes in *Drosophila melanogaster* (Reed and Gibson, 1993; McGinnis *et al,* 1983; Strand and McDonald, 1989; Parkhurst and Corces, 1986a).

The nature of the element is important in determining the mutant phenotype, because transposons carry signals, necessary for various aspects of their expression that can act on adjacent genes or interfere with their own transcriptional machinery (Smith and Corces, 1991). Based on the size and restriction map of the insertions I have found, it is hard to define the nature of the insertion as the sequences of some transposable elements are subject to change during transposition (Di Franco *et al,* 1992a; 1992b) which alters their restriction maps. It is necessary to precisely establish the position of the insertion in the three alleles by sequencing, and then to determine the structure of the insertions and the ways in which they affect the expression of the *Adh* gene. The work described in this chapter aims to define the insertion sites in the *Adh* gene and to characterise the nature of the insertion in each allele.

5.2 Results

5.2.1 Identification of the insertion in *AD369*

PCR amplification with primer pair JG88/22 in *AD369* produced a 1.8 kb fragment which included a 1.2 kb insert (see Fig 4.7). The DNA fragment was sequenced directly with a pair of internal primers, JG87 from 5' to 3' and JG83 from 3' to 5' (Fig 5.1 a). The sequence data revealed that the insertion was located between the distal TATA box motif and the transcription initiation site. The precise position of this insertion was shown to be between position -5 and -6 in the *Adh* gene (Fig 5.2).

About 150 bp of the insertion sequence were read in the first sequencing reaction with the primer JG83. A search through the EMBL nucleotide sequence database found that the sequence had perfect homology to the 2.9 kb P element, a transposable element frequently found in *D. melanogaster* populations (O'Hare and Rubin, 1983). The size of the insertion (about 1.2 kb) indicated that it was a defective P element (O'Hare and Rubin, 1983).

PCR amplification was carried out with primers JG16b/67 and JG68/83 (Fig 5.1 b). In each pair, one primer (JG16b and JG83) specifically binds to the *Adh* gene and the other (JG67 and JG68) binds to the P element (see Appendix II and III). The PCR product was successfully obtained with DNA from *AD369* but not from the control *AC5* (Fig 5.3). This result confirms that the insert has homology to the P element. The fragment amplified with primer pair JG88/22 was further sequenced with primers which specifically bind to the P element (see Appendix III).

The sequence analysis showed that the insert was a 1154 bp KP element (Black *et al*, 1987), derived from a P element by an internal deletion from nucleotide 808 to 2561 (nucleotide numbering follows Black *et al*, 1987). The KP element in *AD369* had 31-bp inverted terminal repeats (Fig 5.4). A duplicated sequence, GTCCAAGT, which is presumably the target site (O'Hare and Rubin, 1983), was found at both ends of the insertion. This target site, which is present

in the normal *Adh* sequence at nucleotide -13 to -6 (numbering follows Kreitman, 1983), differed from that found by Black *et al* (1987) at 3 nucleotides (Fig 5.2). The sequence of the KP element in *AD369* was exactly the same as that isolated from the Krasnodar strain described by Black *et al* (1987). The orientation of the KP element insertion at the *Adh* locus in *AD369*, like that in the Krasnodar strain, was in the same orientation as *Adh* transcription.

5.2.2 Identification of the insertion in *AHA51*

The insertion in *AHA51* is about 2.4 kb and was amplified with primer pair JG87/17 (see Fig 4.8c) or JG87/83 (see Fig 4.8b and Fig 5.5a) under improved PCR conditions. Based on the size of the insertion it is possible that the insertion in *AHA51* is a duplication of a KP element (see Reed and Gibson, 1993). However, PCR amplification with a pair of primers, of which one primer binds specifically to the *Adh* gene and the other binds specifically to the KP element, was not successful in *AHA51* (data not shown). The DNA amplified with JG87/83 in *AHA51* produced two bands — a fragment of 2.7 kb which was expected to include the insertion, and a fragment of 0.3 kb that was the normal size of this region (Fig 5.5a lane 3). This unexpected result will be discussed in the next chapter. PCR amplified DNA was digested with a variety of enzymes (Fig 5.5a). The large band was digested into two fragments both with *Eco*RI and *Pvu*II (Fig 5.5a lane 4 and lane 7), but the small one was not. This result confirms that there is an *Eco*RI and a *Pvu*II site present in the insertion fragment (see 4.2.1). The hybridisation signals detected by the P-element specific probe *pπ25.1* (see 2.4, O'Hare and Rubin, 1983), did not correspond to the bands observed on the ethidium bromide-stained agarose gel (Fig 5.5b). Therefore the sequence of the insertion in *AHA51* does not have homology to the P-element.

To reveal the nature of the insertion in *AHA51*, it was necessary to sequence the insertion. A 1.6 kb *Sal*I/*Eco*RI fragment, isolated from the large fragment amplified with primers JG87/17, was first subcloned into pBS SK(+)

and sequenced with the universal primers. In the first sequencing reaction, 131 base pairs of the insertion sequence was read. A search through the EMBL nucleotide sequence database showed that the sequence of the insertion was perfect homologous to the 3.0 kb *hobo* element, a transposable element found in *Drosophila* (Blackman *et al*, 1989; Streck *et al*, 1986). The size of the insertion in *AHA51* indicates that it is a defective *hobo* element. Restriction enzyme analysis of the *hobo* element in *AHA51* revealed that an approximately 0.6 kb deletion was located between the *Hae*II (+845) and *Hae*II (+1400) sites (Fig 5.6c).

The *hobo* element from *AHA51* subcloned into pBS SK(+) was partially sequenced. The data showed that the *hobo* element in *AHA51* was a deletion derivative from the complete *hobo* element (Streck *et al*, 1986) with an internal deletion of 524 bp from +944 to +1468 (Fig 5.6b) (numbering follows Streck *et al*, 1986). In place of the deletion, there was a 37 bp AT rich sequence (Fig 5.6b). The *hobo* element in *AHA51* contained the 12 bp inverted terminal repeats flanked by a 8 bp duplicated sequence (GTCCAAGT) at both ends (Fig 5.6b). The target site was present in the normal *Adh* sequence at positions from -13 to -6 (Fig 5.6a), which is exactly the same site where the KP element was inserted in *AD369*. There were three perfect tandem repeats from nucleotide 1876 to 1965 in the *hobo* insertion in *AHA51* instead of the ten tandem repeats in the complete *hobo* element. A *Pvu*II site located about 1 kb 3' to the *Eco*RI site (Fig 5.6b), is possibly due to a single nucleotide change from T to C at position 2179.

5.2.3 Identification of the insertion in *AAS44*

The insertion in *AAS44* has been found to be about 5.0 kb (see Fig 4.5), which was too large to be amplified in one fragment. The strategy to identify this insertion was to amplify the 5' part of the insertion by using inverse PCR (Triglia *et al*, 1988) and then to sequence it. The restriction map of *AAS44* showed that there was a *Sal*I site in the 5 kb insert (see Fig 4.5), which could be used for inverse PCR. However, the estimated distance from the *Sal*I site to the

5' end of the insertion was over 2 kb, which is possibly still too large to be amplified by normal PCR. Therefore, a tetranucleotide sequence recognition enzyme, *Sau*3AI, was selected for use in the inverse PCR.

There is a *Sau*3AI site on the *Adh* gene about 300 bp 5' to the insertion region (Fig 5.7). The pair of primers JG83 and JG13, which are located between the *Sau*3AI site and the insertion site, was used for the inverse PCR. Theoretically, the tetranucleotide target occurs on average once every 4^4 nucleotides (Maniatis *et al*, 1989), so there might be several *Sau*3AI sites along the 5 kb insertion fragment. The *AAS44* genomic DNA was digested with *Sau*3AI and the DNA fragments were self-ligated with T4 DNA ligase to form a circle (Fig 5.7). A restriction site *Ssp*I between the primers JG83 and JG13 was used to linearize the circularised DNA. PCR amplification with primers JG13/83 produced two bands of 600 and 150 bp (Fig 5.8). The larger band was an incompletely digested fragment as it can be re-digested into two fragments with *Sau*3AI (data not shown).

The DNA produced by inverse PCR was subcloned into pBS SK(+) and sequenced with the universal primer. About 100 base pairs of the insertion sequence were read from the first sequence reaction. A search through the EMBL data bank showed that the insertion fragment was homologous to the 5' end of the 5.7 kb *Tirant* element, a transposable element first found inserted in the *extramacrochateae (emc)* locus in *Drosophila* (Garrell and Modolell, 1990). The precise position of the insertion was between +409 and +410 in the first intron of the *Adh* gene, which is 441 bp 3' to the distal promoter and 266 bp 5' to the proximal promoter (Fig 5.9). The orientation of *Tirant* in *AAS44* was in the same direction as that at the *emc* locus, which is in the 5' to 3' direction of the *Adh* transcript. The 3' end of the insertion in *AAS44* has not yet been defined.

5.3 Discussion

Sequence analysis identified the insertions in *AD369, AHA51* and *AAS44* as transposable elements. There are different transposable elements in the three variants. The insertion was a KP element in *AD369,* a defective *hobo* element in *AHA51* and a *Tirant* element in *AAS44.* P and *hobo* transposable elements are widely distributed in the *Drosophila* genome, but have not previously been found in the *Adh* locus.

The KP element is derived from a P element (Black *et al,* 1987). P elements were first discovered in *Drosophila melanogaster* (O'Hare and Rubin, 1983) and shown to be the causative agent of hybrid dysgenesis which is characterised by increased chromosome breakage, high mutation rates, sterility and male recombination (Kidwell *et al,* 1977). There are two categories of *Drosophila* strains: P strains contain autonomous and nonautonomous P elements; M strains do not contain any autonomous P elements (Kidwell *et al,* 1977). The syndrome of dysgenic traits usually only occurs in crosses between P strain males and M strain females (for review see Bregliano and Kidwell, 1983).

The P elements and their derivatives have 31-bp inverted repeats at both ends and typically create an 8-bp duplication of the target site upon insertion (O'Hare and Rubin, 1983). The terminal inverted repeats and their internal sequences are required for transposition. The P element has four open reading frames which are required for the production of P transposase (Kraress and Rubin, 1984). The P element is transcribed to yield a single poly $(A)^+$ RNA in both somatic and germ cells. Only within the germ line the third intron is spliced from the primary transcript, which is an essential step for the production of P transposase and hence P element transposition (Laski *et al,* 1986). A strong P strain is estimated to have approximately 40 to 50 P element copies inserted within the major chromosome arms, only one third of which are

copies of the complete 2.9-kb P element. The remainder are smaller P element derivatives arising from internal deletions within the complete P element.

The KP element is one kind of defective P element and was first described by Black *et al* (1987). This element was derived from a strain of *D. melanogaster* isolated from a population at Krasnodar (Russia). The element, called KP, was cloned and found to have a deletion from nucleotide 808 to 2561 extending from the second exon to the fourth exon (see Fig. 5.4). Five KP elements were sequenced and no mutational differences were found between any of them. The KP element retains the sequences required for transcription initiation and polyadenylation (Laski *et al,* 1986) and is transcribed to yield a 0.8-kb poly $(A)^+$ RNA. Potentially, the KP element is capable of translating a 207 amino acid polypeptide but this has not been identified and its function is speculative (Rasmusson *et al,* 1993). It is also known that most defective P elements cannot transpose unless the transposase is provided *in trans* from another source (Spradling and Rubin, 1983; Rasmusson *et al,* 1993; Engels, 1989). The KP element preferentially accumulates to 40-60 copies in the presence of P factors and suppresses P-induced hybrid dysgenesis in *Drosophila melanogaster* (Black *et al,* 1987).

The KP element detected in *AD369* has exactly the same sequence as that isolated from the Krasnodar strain. They also have the same transcriptional orientation. Both the *AD369* and Krasnodar KP elements show complete homology to the P element *pπ25.1* (O'Hare and Rubin, 1983). A single nucleotide replacement of an A by a T follows the 31-bp inverted repeat at the 5' end but does not result in a 32 bp repeat. This same substitution was found in a P element cloned from the American *π2* strain (O'Hare and Rubin, 1983).

There is a tendency for P elements to insert preferentially into euchromatin compared to heterochromatin (Engels, 1989). There are also marked differences in the insertion rates at different loci. For example, the *singed* locus has P element insertions at frequencies between 0.1 and 1%

(Engels, 1989; Green, 1977), while the *Adh* gene failed to acquire any P elements in a screen of 10^7 gametes from dysgenic hybrids (Engels, 1989). The KP element insertion in *AD369* is the first example of a P element insertion at the *Adh* locus, despite intensive attempts to produce mutants at this locus (M. Ashburner, personal communication; Kidwell, 1986). In Australian populations, the number of full-sized P elements per genome was found to be high in lines from the northern region but declined overall from north to south, while the number of KP elements had a reciprocal cline to that of the full-sized P elements (Boussy *et al*, 1988). The allele *AD369* was originally isolated from the All Saints population in Victoria, which is a region with high levels of KP elements.

The KP element inserted in the *Adh* locus in *AD369* has the same sequence as the KP element detected in the *Gpdh* locus in an Australian population from Tasmania, except that the transcription orientation is in the opposite direction (Reed and Gibson, 1993). The KP elements in the *Adh* and *Gpdh* loci are inserted in the same region between the TATA box and the transcription start site.

A characteristic feature of P element insertions is that they are often found in close proximity to the 5' transcription initiation site (Tsubota *et al*, 1985; Searles *et al*, 1986; Roiha *et al*, 1988). This is exemplified by the insertions in the *Gpdh* alleles, as well as the KP insertion in *AD369*. One possible explanation put forward to account for this is that in active genes the chromatin structure in this area is believed to be relatively 'loose' which in turn makes the DNA more accessible to insertions (Eissenberg and Elgin, 1987).

Ashburner (1989) collated data on the sequences flanking the inserted P elements in 31 examples and was able to formulate a consensus sequence: GNCCAGAC. While the hot spot in the *white* locus matches this consensus exactly (O'Hare and Rubin, 1983), the sites at the *singed* (Roiha *et al*, 1988) and *Notch* (Kelley *et al*, 1987) loci only show homology in 4 or 5 out of 8 positions.

Scanning the complete *AD369* sequence for this consensus sequence, there is one perfect match and eight places that match the consensus at 7 out of the 8 positions. There are another 58 sites dispersed along the gene that match 6 of the 8 consensus bases. In addition to the actual insertion site (GTCCAAGT at position -13 to -5), the sequence of the *Adh* gene contains a number of sites which have 5 out of 8 bases identical to the consensus sequence. The fact that the insertion has occurred at this specific point, instead of inserting into other similar sequences, suggests that chromatin structure may be involved in site selection. When the *AD369* sequence was scanned for other copies of the KP insertion site (GTCCAAGT), none was found. The closest matches are five stretches having homology in 7 out of 8 bases.

A defective *hobo* element was found to be inserted in *AHA51* at exactly the same site (GTCCAAGT) as the KP element in *AD369*. The *hobo* element target site consensus sequence is NNNNNNAC or CTTTNNNN. These weak consensus sequences are dispersed all along the *Adh* gene giving many possible target sites. The insertion of a *hobo* element results in the duplication of the 8-bp target site at both ends (McGinnis *et al*, 1983; Streck *et al*, 1986).

The *hobo* element was first identified in a *Sgs-4* allele in *Drosophila melanogaster* (McGinnis *et al*, 1983). Sequence analysis revealed that a complete *hobo* is a 3.0-kb element with short 12-bp inverted repeats (Streck *et al*, 1986). It contains a single 1.9-kb open reading frame between nucleotides 307 and 2250, which could encode a 75-kDa protein. Although similar in size and structure, the sequence of *hobo* is not homologous to the P element. Only the complete 3.0-kb *hobo* elements are able to provide the transposase function required for *hobo* movement (Blackman and Gelbart, 1989).

As in the P-M system, there are two kinds of *hobo* strains: H strains which contain *hobo* elements, and E strains, which do not contain *hobo* elements of any size (Streck *et al*, 1986). H strains generally contain 2-10 copies of the 3.0-kb *hobo* element and approximately 30-75 copies of defective elements

(Blackman *et al*, 1987; Streck *et al*, 1986). The complete *hobo* element is the most highly conserved *hobo* element among different strains. The deletion derivatives are more diversified. Defective *hobo* elements appear to be generated from complete elements by internal deletions (Streck *et al*, 1986).

Hobo mobilisation has been observed in H x E and E x H crosses at nearly equal rates and also in H x H crosses (for review see Blackman and Gelbart, 1989; Smith and Corces, 1991). In these crosses there is a high frequency of cytologically visible chromosome re-arrangements. These re-arrangements are usually unstable because additional re-arrangements may involve existing chromosomal breakpoints (Hatzopoulos *et al*, 1987; Lim, 1988). The *hobo* elements are also associated with small deletions , inversions, and insertions that are not cytologically visible (Blackman *et al*, 1987; Hatzopoulos *et al*, 1987; Yannopoulos *et al*, 1987; Lim, 1988).

The defective *hobo* in *AHA51* is derived from the complete *hobo* with an internal deletion of 524 bp from +944 to +1468 (nucleotide numbering follows Streck *et al*, 1986). The *hobo* element in *AHA51* contains the terminal sequences which are highly conserved for *hobo* transposition. The 524 bp deletion was replaced by a 37 bp AT rich sequence, which is different to any other *hobo* derivative (Streck *et al*, 1986). A *Pvu*II restriction site was present in the defective *hobo* in *AHA51*, and this site has been reported in the great majority of full-sized *hobo* elements and their derivatives (Pascual and Periquet, 1991). A number of repetitive sequences are present in the complete *hobo*, one of which is a 9-bp sequence in 10 perfect and 5 imperfect copies located between nucleotides 1849 and 1983 (see Fig 5.6c and Streck *et al*, 1986). The defective *hobo* element in *AHA51* had only three perfect repeats instead of ten.

Hobo elements are widely distributed in all recently collected strains but hardly detected in old collections of *Drosophila melanogaster* (Pascual and Periquet, 1991). Natural populations from eastern Australian harbour full-

sized and small *hobo* elements (Boussy *et al,* 1988). There is no latitudinal pattern in the frequencies of the small or the full-sized *hobo* elements.

The insertion found in the third allele *AAS44* is a *Tirant,* which was first detected in the *Ach* mutation at *extramacrochaetae* (*emc*) locus in *Drosophila* (Garrell and Modolell, 1990). The *Ach* mutation is a spontaneous gain-of-function *emc* allele associated with the *Tirant* insertion, which truncates the *emc* ORF. The *Tirant* element is 5.7 kb long and contains moderately repetitive DNA. Its restriction map is different from that of other elements (Finnegan and Fawcett, 1986). The *Tirant* element causes a 4-bp (CCCG) duplication upon insertion, a characteristic of many transposable elements. The sequence of the 5' end of the insert in *AAS44* is highly homologous to the 5' end of the sequence of *Tirant* and the size of the insert in *AAS44* is very similar to that of *Tirant*. The target site duplication of *Tirant* in *AAS44* could not be detected as the 3' end of the insert in *AAS44* has not been defined. The presumed target site is located on the *Adh* gene at position +406 to +409 (AGCG), which is two nucleotides different from the target site previously described (Garrell and Modolell, 1990).

The *Tirant* element is inserted in the *Adh* gene at a different position to the KP and *hobo* elements. *Tirant* is inserted 3' to the distal promoter and transcription start site, but 5' to the proximal promoter and start site.

In summary, the three variants have different transposable elements inserted in the promoter region. The variants *AD369* and *AHA51* have the KP and *hobo* element insertions targeted at exactly the same site — 5 bp 5' to the distal transcription start site. The *AAS44* allele has the *Tirant* element insertion between the distal promoter and the proximal promoter at nucleotide +409.

Other studies of restriction endonuclease variation in the *Adh* region showed that there were a large number of insertions had been found in the region of the *Adh* locus (Langley *et al,* 1982; Cross and Birley, 1986). Some of them have been identified, but most of them remained unknown (Aqudro *et al,* 1990; Jiang and Gibson, 1992a). Among the large number insertions in this

region, only a few have been shown to be associated with ADH activity variation, most of them do not have any effect on ADH activity (Jiang and Gibson, 1992a; Dunn and Laurie, 1995). For example, among the extensive variation in sequences flanking the *Adh* gene, only one length polymorphism is clearly associated with altered *Adh* expression, that is a *copia* element inserted approximately 250 bp 5' to the distal transcript start site (Aqudro *et al*, 1986; Dunn and Laurie, 1995). Another Adh^F line with a 1.5 kb insertion about 2 kb upstream to the 5' end of the *Adh* transcription unit showed a lower ADH activity (Jiang and Gibson, 1992a). The Adh^{SL} allele, which has a 4.5-kb insert approximately 3 kb 5' to the distal promoter, consistently showed lower ADH activity in larvae and adults (Schott *et al*, 1988).

The position of the insertions defined in the three low ADH activity variants are definitely all different from the Adh^{SL} allele, which is approximately 3 kb 5' to the distal promoter (Schott *et al*, 1988). These insertions are all located in the promoter region, which might be close to where the *copia* element inserted (Dunn and Laurie, 1995). The transposable element insertion in each variant is most likely the cause for the low ADH activity observed in these three variants. It can be directly tested by deleting the insertion and re-assaying the ADH activity in each allele, which will be studied in the next chapter.

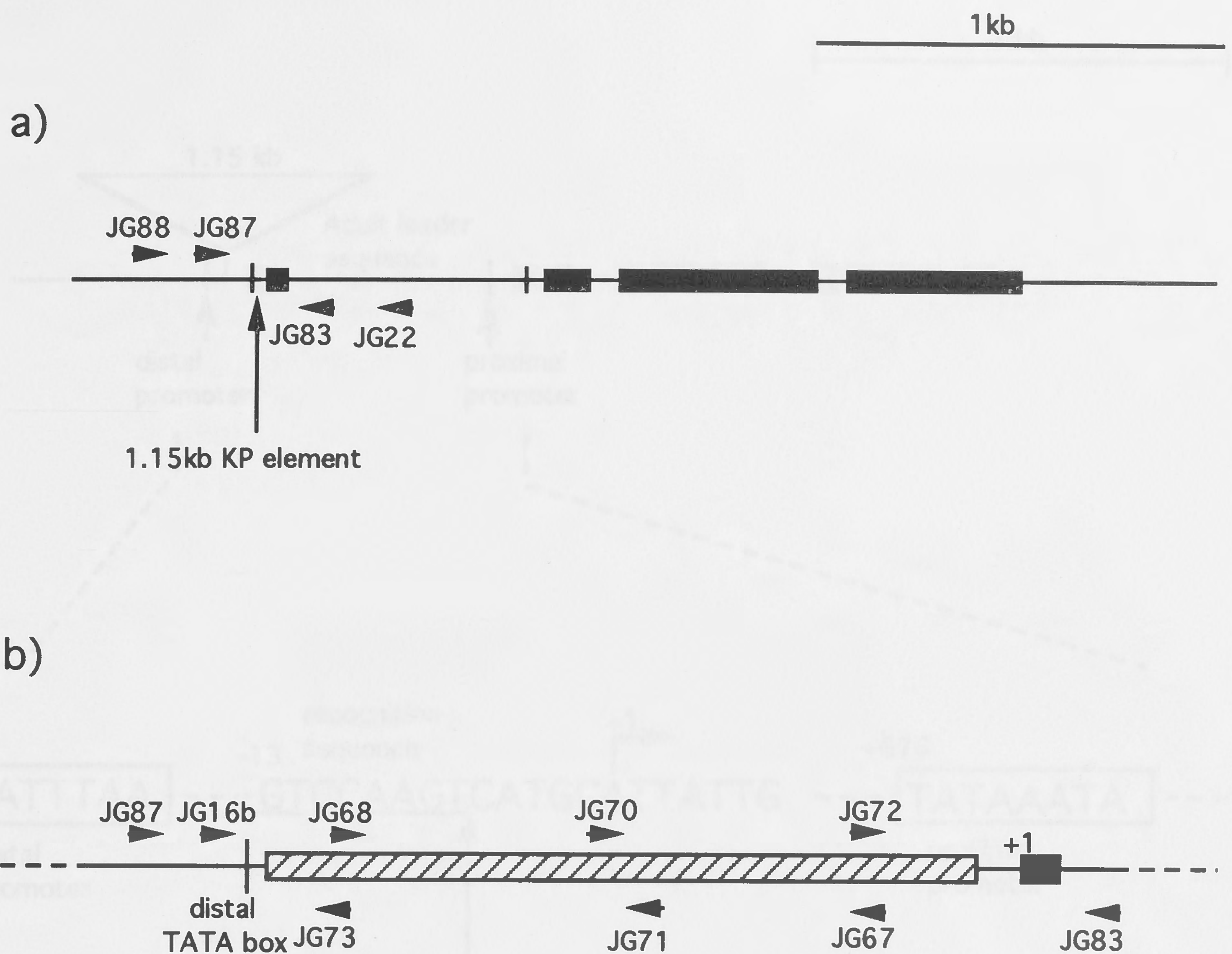


Figure 5.1 Primers used for amplification and sequencing of the KP element in *AD369*.

The positions of the primers used in the direction 5' to 3' are indicated above the diagram of the *Adh* gene and those in the direction 3' to 5' are indicated below the diagram of the *Adh* gene. In a), exons are indicated as black boxes. In b), the hatched region represents the KP element inserted between the distal TATA box and the transcription start site.

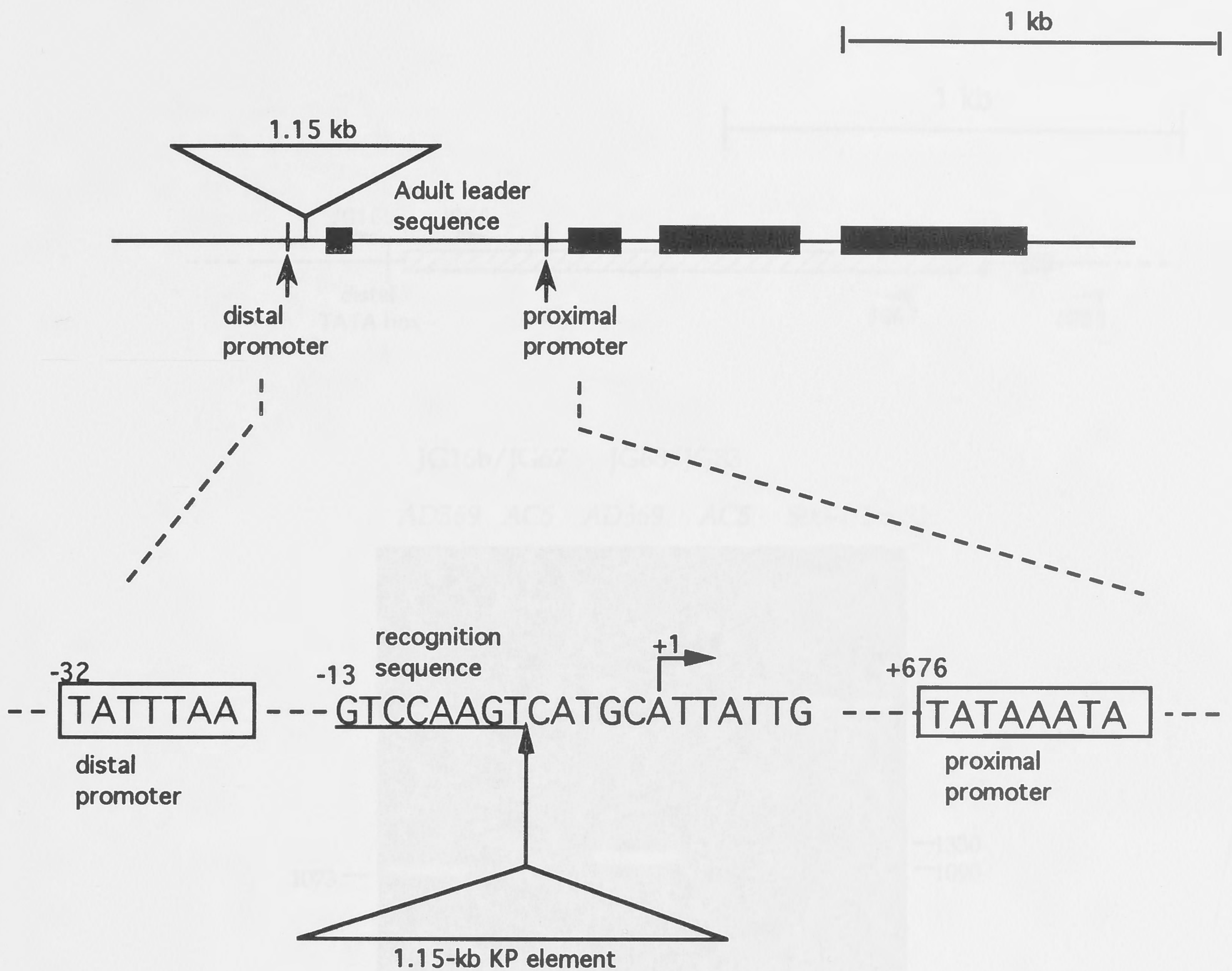


Figure 5.2 KP element insertion in *AD369*

The KP element is shown as an open trangle and the sequence of the insertion site is indicated. The exons of the *Adh* gene are shown as black boxes. In the lower part of the diagram , the two TATA boxes are boxed and the 8-bp duplication site is underlined. Nucleotide numbering follows Kreitman,1983.

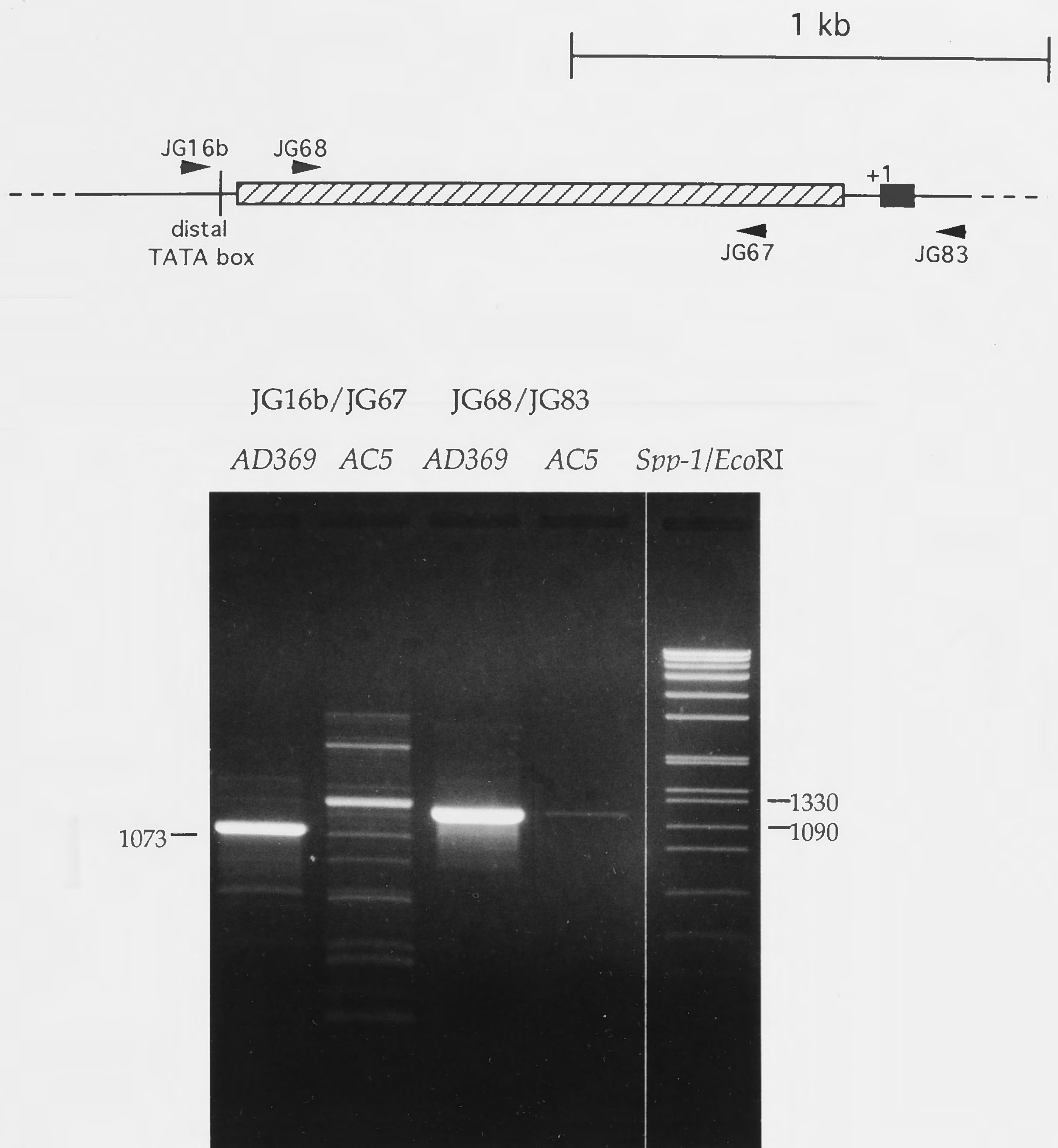


Figure 5.3 PCR amplification of the insertion in *AD369* with P element specific primers

The upper diagram shows the primers used for the PCR amplification, in which the hatched box represents the defective P element and the black box represents the first part of the *Adh* gene that is transcribed. DNA fragments amplified by the PCR were electrophoresed on a 1% agarose gel (lower picture) and the estimated sizes of the fragments are indicated in basepairs. DNA samples and primers used are shown above the picture. *Spp-1* DNA digested with *Eco*RI was used as a size marker and the sizes of DNA are in basepairs.

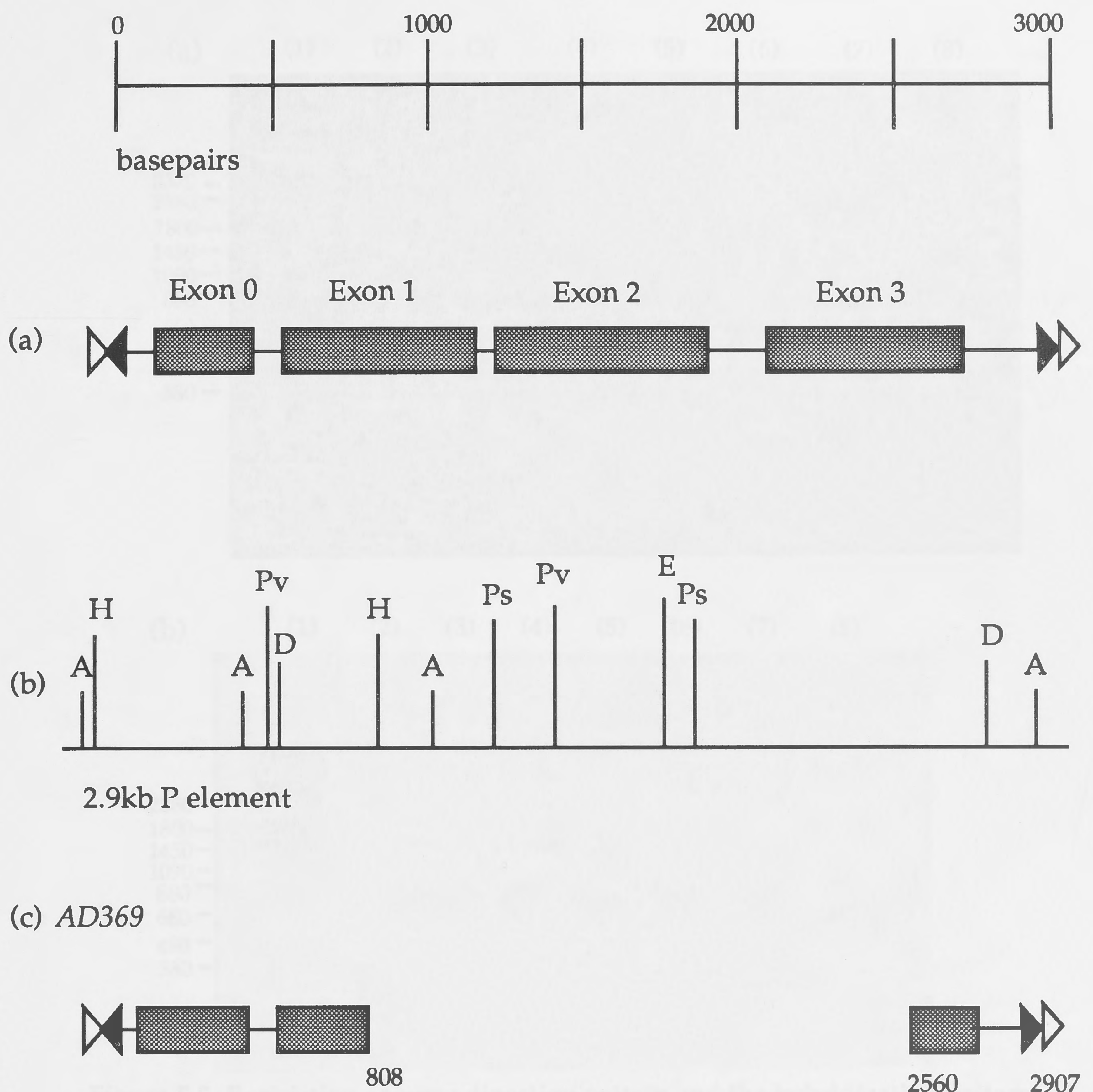


Figure 5.4 Structure of the complete P element and its derivitive in *AD369*

(a) The boxes represent the four coding regions of the complete P element. Open arrowheads represent the 8-bp target site duplication, and filled arrowheads represent the P element terminal 31-bp inverted repeats.

(b) Restriction map of the complete P element. The sites are: *Ava*II (A), *Dde*I (D), *Eco*RI (E), *Hin*dIII (H), *Pst*I (Ps) and *Pvu*II (Pv). There are three *Dde*I restriction sites at 2762, 2796, and 2814, however only one has been indicated.

(c) The structure of the KP element in the *Adh* low activity variant *AD369*. An internal deletion of 1752 bp is shown.

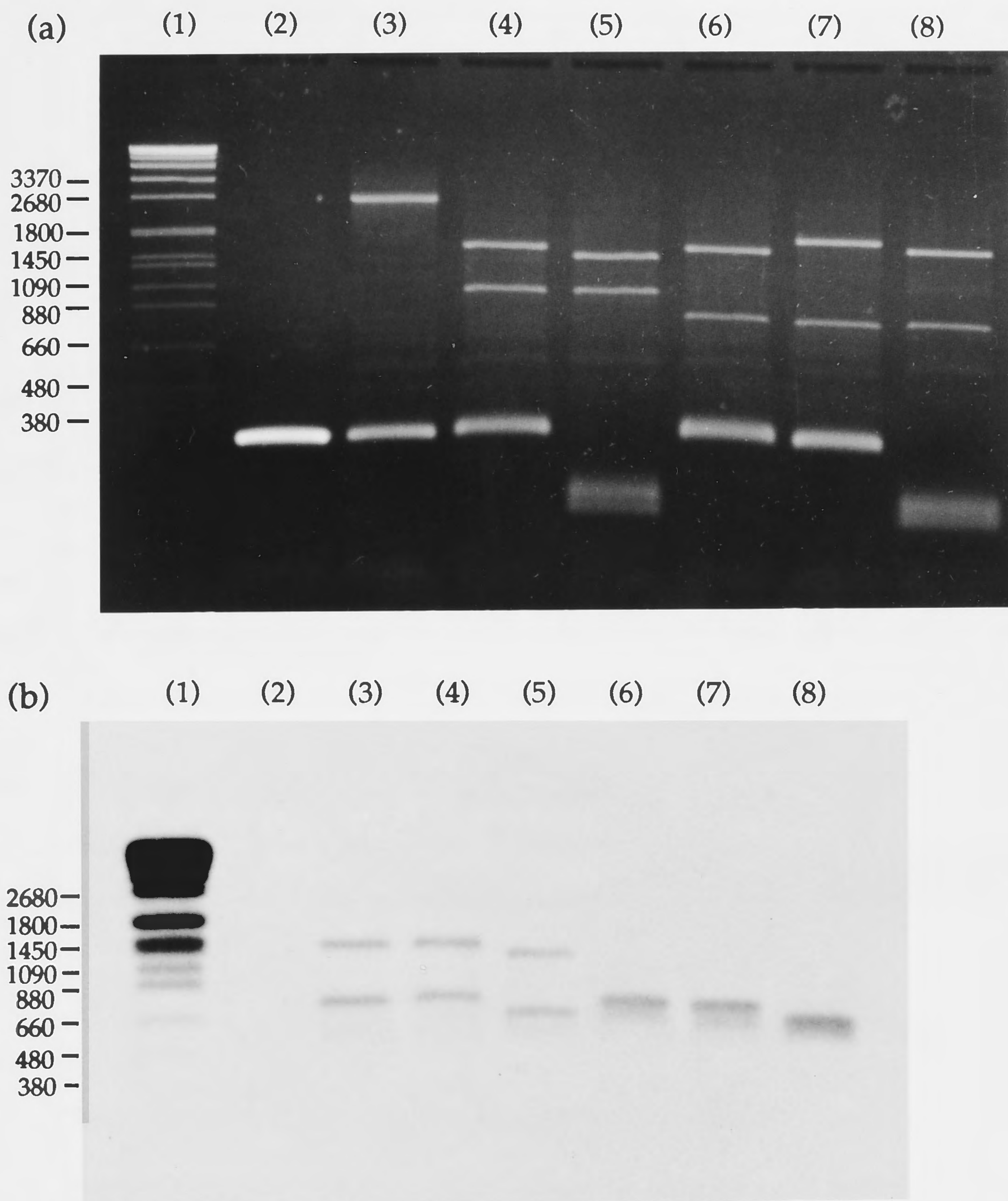


Figure 5.5 Restriction enzyme digestion pattern and the hybridization pattern of PCR amplified fragments which contained the insertion in *AHA51*.

(a) Electrophoresis pattern of fragments amplified with primers JG87/JG83 in *AHA51* and then digested with a variety of enzymes: lane 1, *Spp-1* DNA digested with *Eco*RI used as a size marker and the sizes are in basepairs; lane 2, fragment amplified in *AC8;* lane 3, fragment amplified in *AHA51;* lane 4, fragment amplified in *AHA51* and digested with *Eco*RI; lane 5, fragment amplified in *AHA51* and digested with *Eco*RI/*Sal*I; lane 6, fragment amplified in *AHA51* and digested with *Eco*RI/*Pvu*II; lane 7, fragment amplified in *AHA51* and digested with *Pvu*II; lane 8, fragment amplified in *AHA51* and digested with *Pvu*II/*Sal*I

(b) Hybridization pattern of above gel probed with ^{32}P-labelled *pπ25.1*, a P element specific probe (O'Hare and Rubin, 1983).

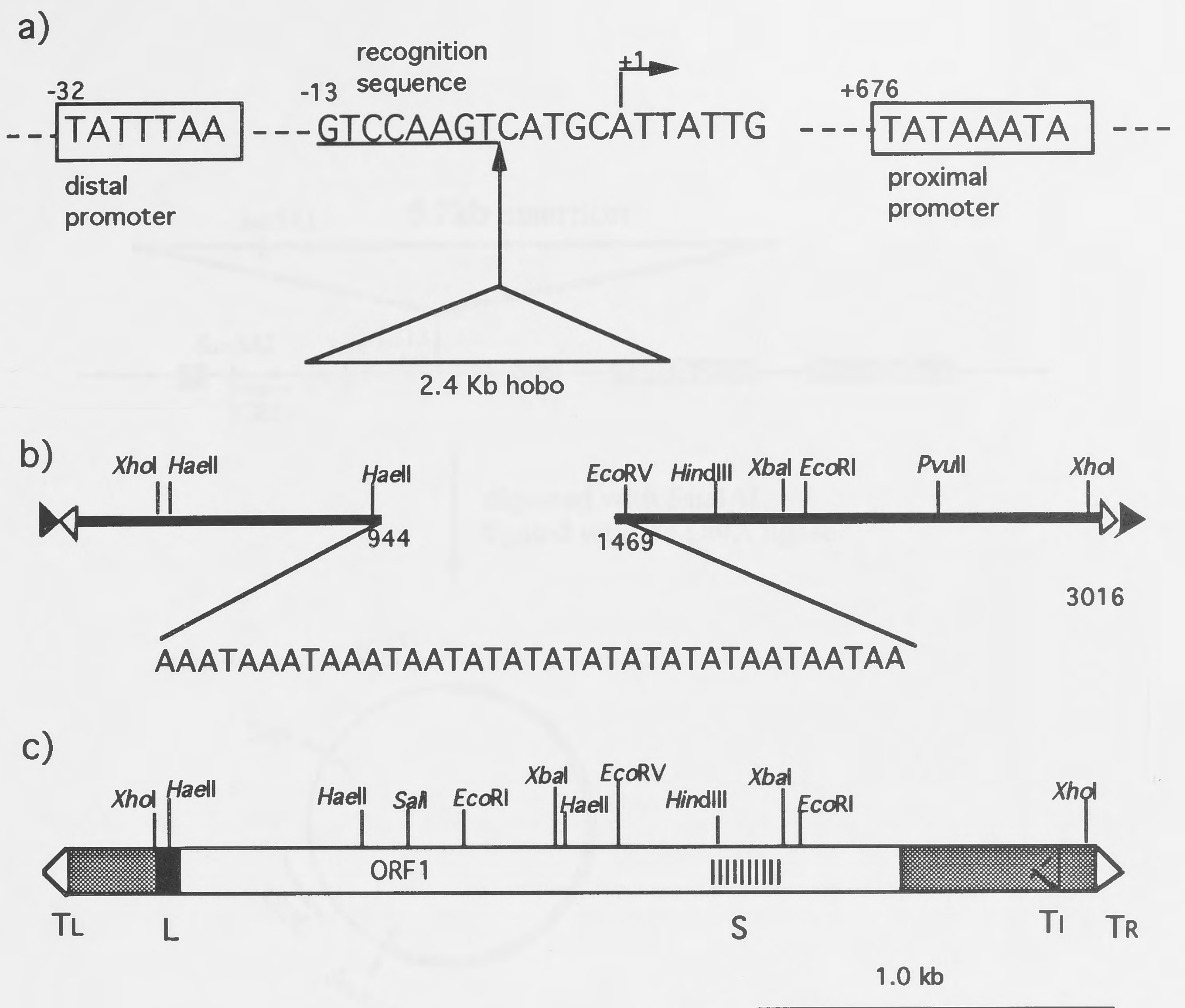


Figure 5.6 Insertion of a defective hobo element in *AHA51*

a). Insertion site of hobo in *AHA51*. The insert is shown as a large open trangle (not drawn to scale) and the site of insertion is indicated. The 8-bp target site is underlined and the two TATA boxes are boxed. Nucleotide numbering follows Kreitman, 1983.

b). Defective hobo in *AHA51*. Thick lines represent the hobo sequences. Open arrowheads represent the 12-bp inverted terminal repeats. Solid arrowheads represent the 8-bp duplication sites. The internal deleted region is indicated.

c). Restriction map and major structural features of a complete hobo element. Each terminal repeat is represented by an open arrowhead, each L repeat by a thick vertical bar and each S repeat by a thin vertical bar. The ORF 1 region is white and the rest of the element is shaded.

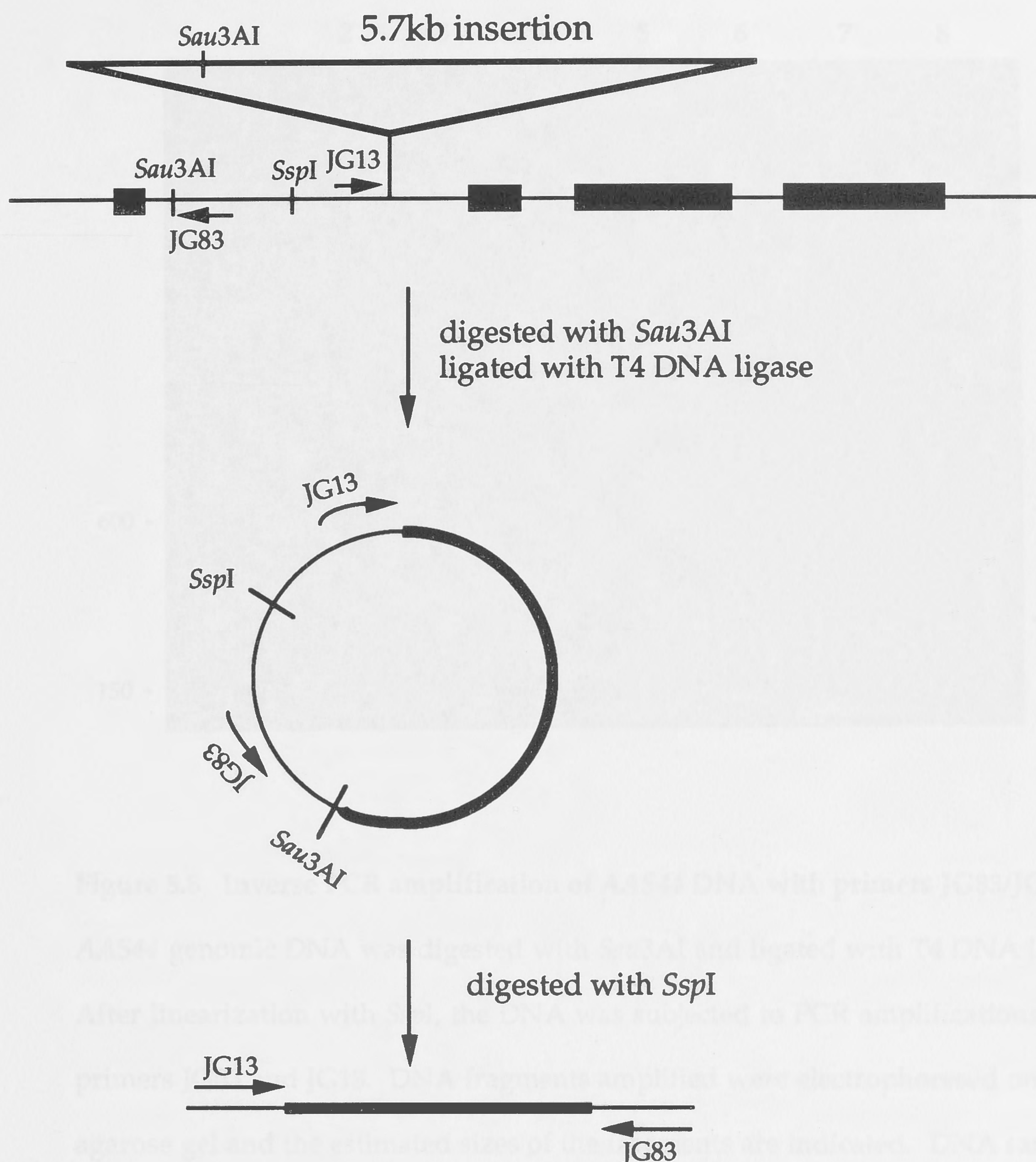


Figure 5.7 Strategy for inverse PCR used to obtain part of the insert in *AAS44*

The upper figure shows the *Adh* gene, the site of the insertion and the positions of a number of restriction enzyme sites. The structure of the ligated molecule is shown and the *Ssp*I site used for linearization of the molecule is indicated. The positions of the PCR primers are indicated and the insertion sequences are represented as a thick black line.

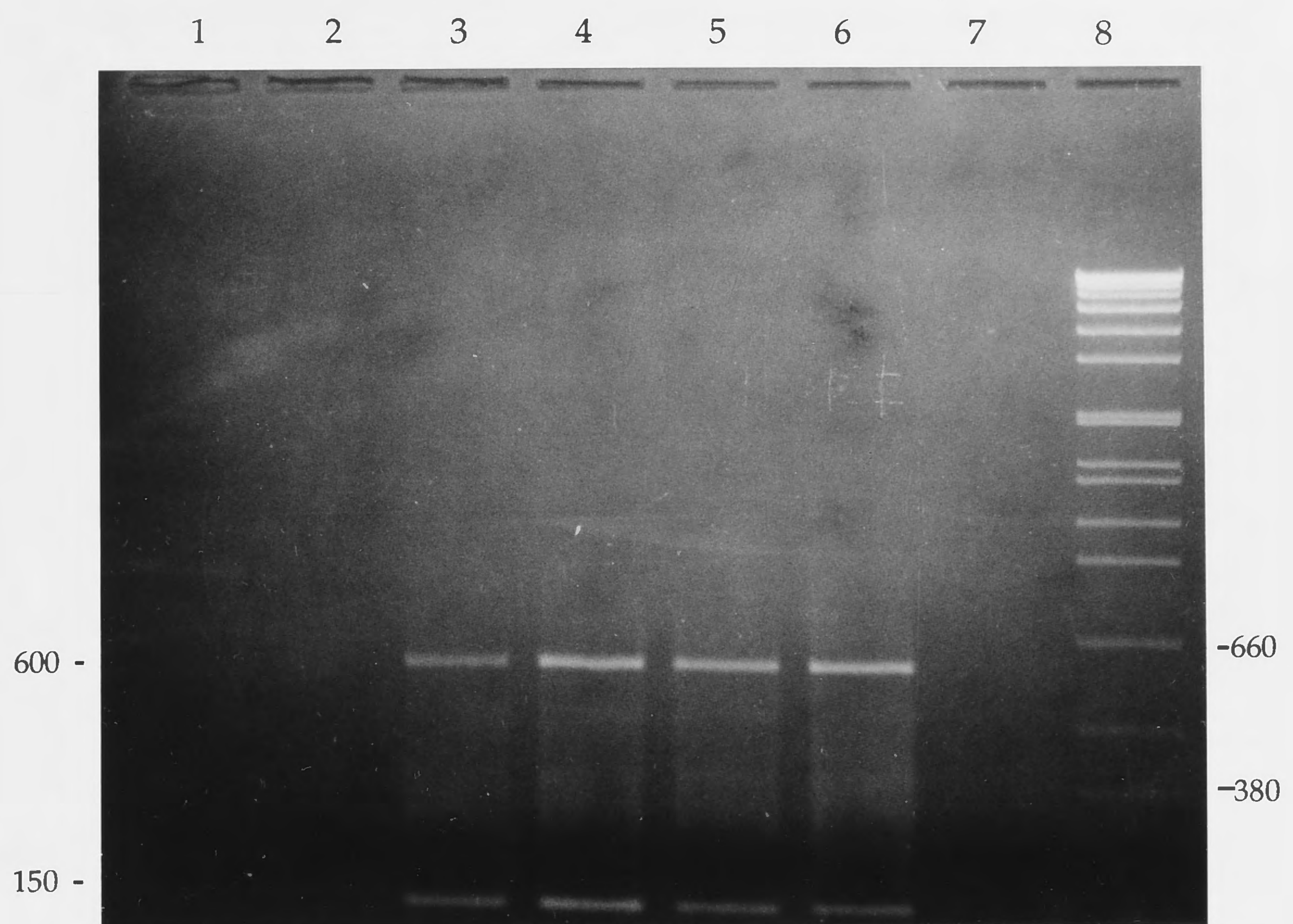


Figure 5.8 Inverse PCR amplification of *AAS44* DNA with primers JG83/JG13

AAS44 genomic DNA was digested with *Sau*3AI and ligated with T4 DNA ligase. After linearization with *Ssp*I, the DNA was subjected to PCR amplifications with primers JG83 and JG13. DNA fragments amplified were electrophoresed on a 1% agarose gel and the estimated sizes of the fragments are indicated. DNA samples were loaded as follow: lane 1, *AAS44* genomic DNA; lane 2, *AAS44* DNA digested with *Sau*3AI; lanes 3 and 4, 2µl ligation sample of *AAS44* DNA digested with *Sau*3AI; lanes 5 and 6, 5µl ligation sample of *AAS44* DNA digested with *Sau*3AI; lane 7, PCR control without DNA sample; lane 8, *Spp-1* DNA diested with *Eco*RI used as a size marker; the sizes of DNA are in basepairs.

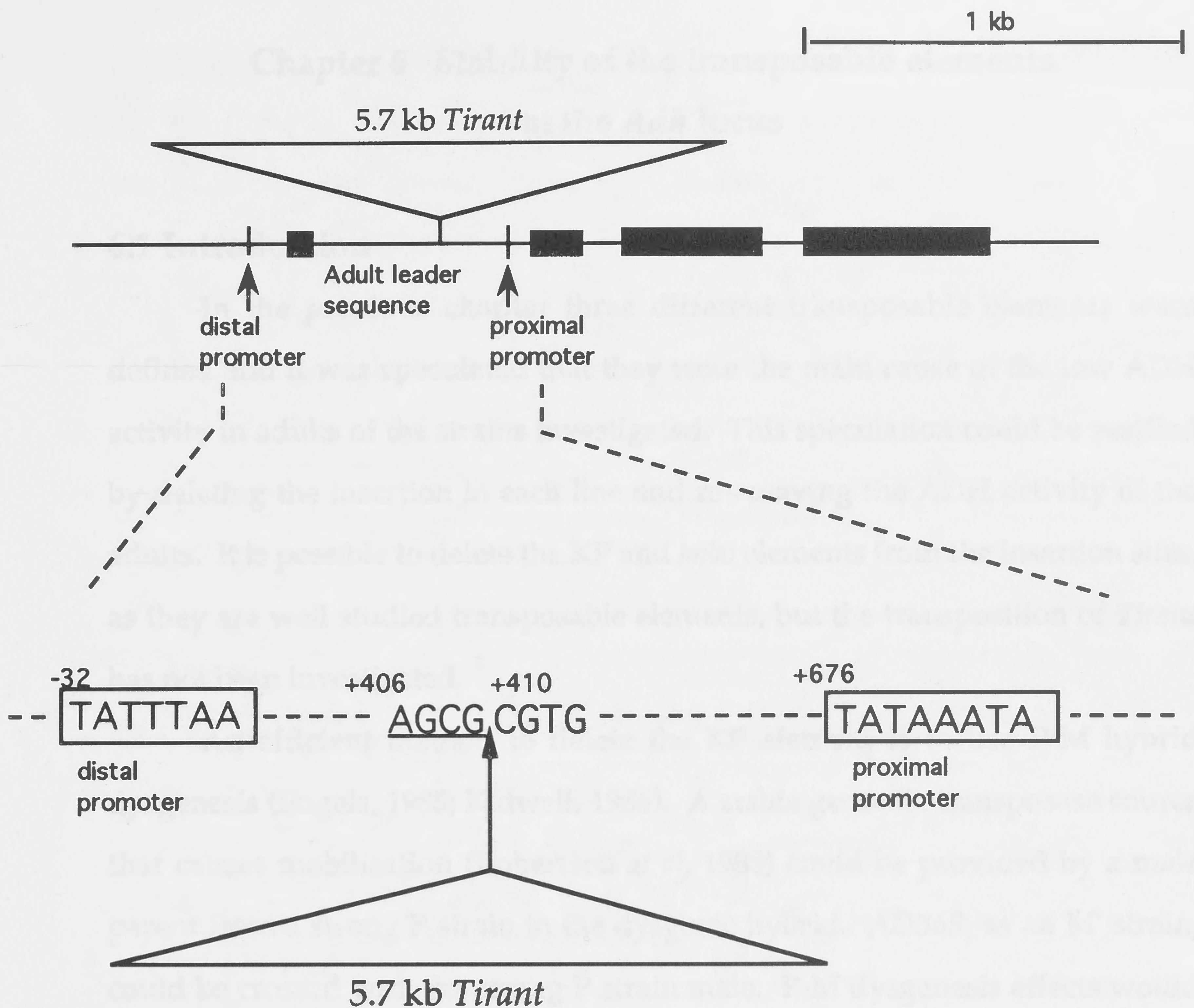


Figure 5.9 The *Tirant* element insertion site in *AAS44*

The *Tirant* insertion is shown as an open trangle (not drawn to scale) and the insert site in the *Adh* gene is indicated. The *Adh* exons are shown as black boxes. The two TATA boxes are boxed. The presumed 4-bp duplication site is underlined. Nucleotide numbering follows Kreitman, 1983.

Chapter 6 Stability of the transposable elements at the *Adh* locus

6.1 Introduction

In the previous chapter three different transposable elements were defined and it was speculated that they were the main cause of the low ADH activity in adults of the strains investigated. This speculation could be verified by deleting the insertion in each line and re-assaying the ADH activity of the adults. It is possible to delete the KP and *hobo* elements from the insertion sites, as they are well studied transposable elements, but the transposition of *Tirant* has not been investigated.

An efficient method to delete the KP element is to use P-M hybrid dysgenesis (Engels, 1985; Kidwell, 1986). A stable genomic transposase source that causes mobilisation (Robertson *et al,* 1988) could be provided by a male parent from a strong P strain in the dysgenic hybrid. *AD369,* as an M' strain, could be crossed with the strong P strain male. P-M dysgenesis effects would occur in the germ line of the F_1 progeny. In the next generation, a phenotype associated with the deletion of the KP element could be observed (Kidwell, 1986). However, KP elements have been reported to inhibit P-induced hybrid dysgenesis (Black *et al,* 1987), and this may make it difficult to delete the KP element from *AD369.*

AD369 flies in which the KP element is deleted from the *Adh* locus would be expected to have increased ADH activity. Therefore flies without the KP element in the *Adh* locus and heterozygous for the null allele *nLA248* would be expected to survive when exposed to a concentration of ethanol vapour that would kill similar heterozygotes that have the KP element insertion in the single active *Adh* allele.

When the *hobo* insertion region in *AHA51* was amplified, two bands were produced — a large band of the size expected with the *hobo* element

inserted and a small band corresponding to the size of a wild-type allele (see Fig 4.8). *Hobo* element transposition needs a transposase encoded by a complete *hobo* element, as well as a host-encoded protein (Smith and Corces, 1991). *Hobo* mobilisation has been reported to occur in the germ line (Blackman *et al*, 1989) and somatic instability has also been shown (Lim, 1979; 1981; Yannopoulos *et al*, 1983). These processes can be studied in *AHA51*.

The work described in this chapter aims to test the effect of excision of the KP element from *AD369* on ADH activity, and to investigate the stability of the *hobo* element in *AHA51* at the molecular level.

6.2 Results

6.2.1 Breeding program used to delete KP element from *AD369*

PCR amplification of *AD369* DNA with a pair of primers JG68/67, which specifically bind to the P element (Appendix III), produced a major band of 900 bp corresponding to the fragment amplified from the KP element. With the same pair of primers, an amplification product of 2700 bp was observed with *pπ25.1* plasmid DNA, which contains a normal P element (Fig 6.1). These results showed that there were no detectable complete P elements in the *AD369* genome. *AD369* is thus characterised as an M' strain.

The breeding program used to construct the lines was given in Chapter 2 and in Figure 2.4. A stable genomic source of P element transposase was provided by *P[ry*$^{+}$*Δ2-3]* (99B) (Robertson *et al*, 1988) in the F_0 male parent which was constructed with the dominant gene *Lobe* (Lindsley and Zimm, 1992) on the second chromosome as a marker. Virgin females *AD369/CyO* were used as the maternal parent in F_0 (Fig 2.4). From the F_1 progeny, male flies with *Lobe, Sb* and wild-type wings were chosen and mated to virgin females homozygous for the *Adh* null allele *nLA248*. The *AD369* allele which had been exposed to hybrid dysgenesis was labelled *AD369**. From the F_2 progeny,

heterozygotes *AD369*/nLA248* were collected for exposing to ethanol vapour to select flies with increased ADH activity.

6.2.1.1 Ethanol vapour exposure to select flies with higher ADH activity.

The technique used to expose adult flies to ethanol vapour has been described in chapter 2 and shown in Figure 2.5. With an ethanol concentration of 8% and 3% sucrose, heterozygotes for the normal activity allele *AC5* and the *Adh* null allele *nLA248* survive, but *AD369/nLA248* heterozygotes have 100% mortality (Table 6.1). Among the 216 F_2 flies tested, 26 heterozygous *AD369*/nLA248* flies survived and these were preserved by crossing in a mass culture with *nLA248* homozygotes (Fig 6.2). The male wild type progeny were aged at 20°C for 16 days and then divided into two samples (60 and 50 flies each) and exposed to ethanol vapour. In one sample 23 flies survived and in the second sample 20 flies survived (Table 6.2). Each of the surviving 43 male flies was mated individually to *nLA248* homozygous females (Fig 6.2). The wild type progeny (F_4) (ie: *AD369* + + /nLA248 cn bw*) were used for further molecular analyses and ADH activity assays.

6.2.1.2 Molecular and biochemical analysis of *AD369**

6.2.1.2.1 Single fly PCR of *AD369*/nLA248*

DNA was extracted from the wild type flies, *AD369* + +/nLA248 cn bw*, in 36 of the 43 single pair crosses. The DNA was used for PCR amplification using the primer pair JG87/83 which spanned the insertion site of the KP element (Fig 6.3). Only one band of about 300 bp was produced in each sample (27 of them are shown in Fig 6.3), while in the control *AD369/nLA248*, two bands were produced — a band of about 300 bp and a larger band of the size expected when the 1.15 kb KP insertion was present (Fig 6.3, lane 28). There was no amplification in the control sample without DNA (Fig 6.3, lane 29).

These results indicated that the KP element had been deleted from the *AD369** alleles.

6.2.1.2.2 ADH activity assay for *AD369*/nLA248* adults

In the F_4 generation (Fig 6.2), 10 males from each of the 10 single lines of *AD369*/nLA248* were collected and assayed for ADH activity. The ADH activity of the *AD369*/nLA248* adults was more than five times higher than *AD369/nLA248* but about 50% of the activity in *AC5/nLA248* adults (Table 6.3). The ADH activity of *AD369/nLA248* was about 10% of the control (see Table 3.6). Thus, although the ADH activity has been increased following the deletion of the KP element, the activity has not returned to the normal level.

6.2.1.2.3 Sequence analysis of *AD369*/Df(2L)64j*

From 8 of the *AD369*/nLA248* lines which lacked KP elements heterozygous males were crossed with virgin *CyO/Df(2L)64j* females (Fig 6.2). Flies with wild type wings, which could be *AD369*/Df* or *nLA248/Df*, were exposed to 8% ethanol vapour to remove *nLA248/Df* flies which lacked ADH activity. The surviving flies in each line were checked by PCR amplification with primer pair JG87/83 and in each sample a band of about 300 bp was produced (Fig 6.4 a). DNA from the samples of *AD369*/Df(2L)64j* in four of the eight lines was sequenced directly following amplification with the primers JG88/22 (Fig 6.4 b).

The *Adh* gene in *AD369** was sequenced from -63 to +174. In this region *AD369** has exactly the same sequence as *AD369* (see Table 4.1), except that the KP element insertion was absent. The sequence data for *AD369** revealed that the excision site was perfectly repaired and none of the KP nucleotides remained.

6.2.1.2.4 Northern analysis of *AD369*/nLA248* adults

AD369/nLA248* heterozygotes from F_4 were used for quantitative Northern analysis. The male parents from eight lines were combined for RNA extraction, after crossing them to *CyO/Df(2L)64j* females to produced progeny for sequencing. Another 50 flies from the remaining 25 cultures were also combined for RNA extraction as an additional sample. The results are shown in Fig 6.5 and Table 6.4. These data show that in both RNA samples, *AD369*/nLA248* gave the same result as the control *AC5/nLA248*, as the hybridisation signal of the normal sized *Adh* band was stronger than that of the *nLA248* band. The transcript level of *AD369*/nLA248* in adults was about 1.18 times of that of the control *AC5/nLA248*, which means that the transcript level *AD369*/nLA248* has returned to normal following deletion of the KP element.

6.2.2 The *hobo* element in *AHA51* is unstable

6.2.2.1 PCR amplification of *hobo* region in *AHA51*

PCR amplification with a pair of primers JG87/17 that spanned the insertion site produced two bands in *AHA51* homozygotes (see Fig 4.8c). The large band of 3.1 kb was of the size expected to include the insertion, while the small band of 0.7 kb was of the size expected for the normal region in *AHA51* without the insertion. Neither band was present in the PCR control without DNA.

Four possible explanations were raised for the production of the small band. 1). It was amplified from contaminating DNA in the PCR amplification. This is unlikely as the band was not present in the PCR control without DNA. 2). The wings in the *CyO/AHA51* flies might have been misclassified as wild-type so that the small band was amplified from the *Adh* region in the *CyO* chromosome. The *CyO* chromosome contained an *Adh*^{F} allele (see 2.1). 3). Germline movement of *hobo* occurred in *AHA51* to produce a mixture of two kinds of *Adh* allele, one with the *hobo* insertion and the other without. 4).

Somatic movement of *hobo* occurred in *AHA51* producing some cells that lacked the *hobo* insertion.

To eliminate the possibility of contamination with *CyO*, flies were checked for their electrophoretic phenotypes on a cellulose acetate gel. The results confirmed that there was no misclassification of *AHA51/CyO* heterozygotes as *AHA51* homozygotes — all the flies had only the Adh^S allele (data not shown). Also, genomic DNA extracted from *AHA51* homozygotes digested with *Eco*RI did not show a band with the same size as the band in the *AC8* control (see Fig 4.1), indicating that there was no detectable amount of *CyO* DNA present in the DNA prepared from the *AHA51* homozygotes.

To investigate the instability of *hobo*, DNA was extracted from ten single *AHA51* homozygote flies and amplified with the primers JG87/83; two bands were produced in each single fly sample (Fig 6.6). To clarify this result, further PCR amplifications with primer pair JG87/17 were carried out with DNA extracted from ten single flies of the genotype *AHA51/Df(2L)64j*, which contained only one copy of the *AHA51* allele in each fly (Fig 6.7). In five samples there was no amplification. Three of them produced only a band of about 700 bp –– the size expected if there was no insertion. In one of the samples, bands of 700 bp and 3100 bp were produced (Fig 6.7). These results implied that the *hobo* element in *AHA51* was possibly unstable both in the germ line and in somatic cells. However, the poor amplification of this DNA with this primer pair is not understood and this makes these results ambiguous.

6.2.2.2 Sequence analysis of the 0.7 kb small fragment

To investigate the nature of the small band (0.7 kb) amplified with the primer pair JG87/22 in *AHA51*, the 0.7 kb DNA fragment was isolated from the gel and sequenced directly with a pair of internal primers JG87 (5' to 3') and JG83 (3' to 5'). The sequence data showed that the small fragment was derived from the *Adh* gene in *AHA51*. In place of the *hobo* element, there were 6 extra

nucleotides between the 8-bp target site duplications (Fig 6.8). The extra 6 nucleotides were a palindrome to the flanking target site duplication sequence. One of the eight nucleotides next to the 6-bp insert was different from the original sequence flanking the *hobo* element (Fig 6.8).

Further sequence data were obtained when the 0.7 kb fragment was subcloned into pBS SK(+) following *Sal*I/*Hin*dIII digestion and then sequenced with the universal primers. These sequence results showed that the 0.7 kb fragment had the same haplotype as *AHA51*, which has a unique haplotype with a mixture of Adh^F (5' to +175) and Adh^S (3' to +175) (Fig 6.11 and see Table 4.2). The nucleotide changes present in the 0.7 kb fragment were compared with the sequence of *AHA51* (Table 6.5). The DNA fragments from three colonies were sequenced. One of them had a 6 bp insert at the *hobo* site, which is the same insert as that identified by direct sequencing. Two of the fragments had an insert with two repeats of the 6 bp.

6.2.2.3 Additional evidence of *hobo* instability

AHA51 genomic DNA was digested with *Xho*I and probed with a nick-translated 2.0 kb *Xho*I fragment from the defective *hobo* in *AHA51*. Two major bands of 2.6 and 2.0 kb were observed (Fig 6.9). These results indicate that complete *hobo* elements are present in the *AHA51* genome, besides the 2.4 kb defective *hobo* elements, one of which was detected at the *Adh* locus. The copy number of the complete *hobo* elements was higher than that of the defective *hobo* elements in *AHA51* as the 2.6 kb band was slightly stronger than the 2.0 kb band (Fig 6.9). The complete *hobo* elements in the *AHA51* genome would provide a source of transposase for the transposition of the defective *hobo* element.

There was no direct evidence that germline movement of the *hobo* element occurred, as the *AHA51* stock, which is kept at 20°C, appeared to be homogenous for *Adh* alleles and no fly had ever been found with normal

activity levels. To investigate whether the small band arose from somatic instability or germline movement of the *hobo* element, 12 single crosses of *AHA51/CyO* male with *CyO/Df(2L)64j* female were set up at 25ºC and the *AHA51/Df(2L)64j* flies in the progeny of each line were pooled for DNA extraction and PCR amplification with primers JG87 and JG17 (Fig 6.10). We expected that some lines would have only the large band if the *hobo* element was stable; some lines would have only the small band if *hobo* movement was confined to the germ line; and some lines would have the large and small bands if the *hobo* element was somatically unstable.

DNA from six of the twelve lines was PCR-amplified with primers JG87 and JG17. The results revealed that two bands were produced in each of the six lines (Fig 6.10), well amplified in line 2, line 5 and line 6, poorly amplified in line 1, line 3 and line 4. The ratio between the two bands differed between the lines. Line 2 showed a strong small band and a very faint large band. Line 5 showed a large band stronger than the small one, while line 6 produced two bands of similar intensity.

The *Adh* gene in line 2 and line 6 was sequenced from -3 to +586 and from +1388 to +1673 (Fig 6.11). Unexpectedly, in the region sequenced, the *Adh* gene in line 2 had nucleotides characteristic of an Adh^F allele and the *Adh* gene in line 6 was like an Adh^S allele and had the same haplotype as *AHA51*. For example, the *Adh* gene in line 2 had a "G" at nt +1490 which characterises an Adh^F allele, while the *Adh* gene in line 6 had a "C" at this position which characterises an Adh^S allele (Kreitman, 1983). A sequence length polymorphism ∇1, which is strongly associated with Adh^F alleles (Laurie *et al*, 1991), was present in line 2 but not in line 6. All the nucleotide positions which differ between *AHA51* and Adh^F alleles are listed in Figure 6.11. The comparative results show that the *Adh* gene in line 2 is the same as an Adh^F allele and the *Adh* gene in line 6 is the same as the *AHA51* allele.

6.3 Discussion

Molecular and biochemical studies of the *AD369** allele showed that the KP element excision was successful in *AD369*. The KP element has been reported to repress P element induced hybrid dysgenesis (Black *et al*, 1987). Three individual X-linked P elements isolated from wild-type strains were tested for their abilities to repress two aspects of hybrid dysgenesis: gonadal dysgenesis and mutability of a double-P element-insertion allele of the *singed* locus (sn^w) (Rasmusson *et al*, 1993). One of the KP elements was a strong repressor of sn^w mutability, another one was a strong repressor of GD sterility. The third KP element had no significant effect on either trait. These functional differences suggest that genomic position may affect the repression ability of a particular type of P element.

Molecular analysis of the *AD369** allele revealed that the KP element was precisely excised from *AD369* and the transcript level of *AD369** was the same as that of the normal activity allele. However, the level of ADH activity in the *AD369** allele was only about 50% of the control normal activity allele. The lower ADH activity of *AD369*/nLA248* heterozygotes was in agreement with the survival rate of F_3 heterozygotes *AD369*/nLA248* exposed to ethanol vapour (Fig 6.2). In this test only 40% of the heterozygotes survived, whereas 100% survival was expected if the activity was normal. Since the transcript level of the *AD369** allele was the same as the control, the lower ADH activity was possibly due to changes in the regulation at the translation level (Kozak, 1992), although I have no other experimental evidence to explain the effect.

The *AD369** allele was sequenced in the 5' region of the *Adh* gene from -63 to +175 which showed no differences from the *AD369* allele. The rest of the *AD369* Adh* gene has not yet been sequenced. It is possible that the *AD369* Adh* gene might contain a sequence polymorphism that affects ADH activity (Laurie-Ahlberg and Stam, 1987; Laurie *et al*, 1991). The chromosome background of the *AD369** is different from that of the *AD369* allele. Not only

the second chromosome, but also the X and third chromosomes have been altered and that might affect ADH activity levels. A haplotype-ADH activity association of *Bam*HI polymorphic restriction site located about 7 kb upstream from the *Adh* gene is found to be associated with the S/F substitution site (Aquadro *et al*, 1986). One fast line with the *Bam*HI site was found to have low activity and one slow line lacking the *Bam*HI site had high activity. Recombinants between the S/F substitution site and a putative regulatory site might explain the altered *Adh* expression (Laurie-Ahlberg and Stam, 1987).

Alternatively, the cause of the low ADH activity in *AD369** might be due to the small insertion located in the 1.45 kb *Eco*RI fragment (-4.2, -2.8), 1320 bp 5' to the distal promoter (see Fig 4.5). This *Eco*RI fragment contains a few elements of the *Adh* larval enhancer (Corbin and Maniatis, 1990). Whether the small insertion in this region affects ADH activity in adults needs to be investigated. However, none of the insertions previously found in this region affect ADH activity (Jiang and Gibson, 1992a; Aquadro *et al*, 1986; Cross and Birley, 1986; Langley *et al*, 1982).

PCR amplification of the *hobo* insertion in the *AHA51* allele produced two bands in most cases, a large band of the size expected for the locus with the *hobo* insertion present and a small band of the size expected for this region of a normal *Adh* gene. Similar results were obtained for DNA extracted from *AHA51/Df(2L)64j* single flies, which carries only one copy of the *AHA51* allele (Fig 6.6, 6.7 and 6.10). DNA from three single *AHA51/Df(2L)64j* flies were amplified to give only one small band. Overall, the results suggest that *hobo* is unstable both in the germ line and in somatic cells, however these results are not conclusive as the PCR amplification was variable with primers in the relevant region. Also, it is known that small fragments are preferentially amplified (Saiki *et al*, 1985; 1988).

Three different molecular structures at the breakpoint have been found when the *hobo* element excised. The first type had additional nucleotides at the

deletion breakpoint. The second type had no additional nucleotides at the breakpoint and the breakpoint was perfectly repaired from the *AHA51* allele. The third type had no additional nucleotides at the breakpoint but the breakpoint was perfectly repaired from an Adh^F allele. The occurrence of these three different structures following *hobo* excision implies different mechanisms of post-excision repair.

In the first situation, the extra nucleotides added during the excision/deletion process appeared to be palindromic to the target site duplication sequence originally flanking the *hobo* element. Atkinson *et al* (1993) reported that in plasmid constructs, *hobo* excision or post-excision repair results in the addition of approximately 8-20 additional nucleotides at the empty donor site, which appear to be related to the sequences originally flanking the *hobo* element. A possible mechanism for the creation of the additional nucleotides at the deletion breakpoints is summarised in Fig 6.12 (W. Warren, personal communication). In this model it is presumed that *hobo* is excised with one basepair 5' overhang. Then, the target site duplication sequences flanking the breakpoint form a hairpin structure. This structure is subsequently opened and extended to form a palindrome when the target site is broken on one strand (Fig 6.12). This extension step gives rise to the alteration at the positions next to the additional nucleotides on the target site duplication (Fig 6.8). The extended nucleotides might match each other to form a 6 bp extra nucleotide addition (Fig 6.12 a), or might not match each other and then both ends would join together to create 12 bp extra nucleotide addition (Fig 6.12 b). This mechanism of *hobo* excision/repair explained the additional nucleotide at the deletion breakpoint and the 1 bp altered target site duplication. However, the *hobo* element breakpoints were sometimes found to be repaired perfectly in *AHA51*, which suggests that the mechanism described above is not the only one that existed *in vivo*.

Hobo is an AC-like element having short inverted sequence repeats at their termini (Berg and Howe, 1989). These elements are thought to transpose by a DNA-only mechanism. The P elements in *Drosophila* are transposed by a cut-and-paste process which leaves a double-strand gap (Engels *et al*, 1990). The repair of gaps results in the transfer of information from a homologous template sequence to the site of P element excision. When the individual is heterozygous for the P insertion, the corresponding site on the homologous chromosome is used as the template, resulting in a precise loss of the P element. This cut-and-paste process could be speculated to occur during *hobo* transposition.

In the second situation, the *hobo* element was deleted precisely and the breakpoint was repaired perfectly from the *AHA51* allele as shown in line 6 (Fig 6.11). In this case, *hobo* element excision must occur in the somatic cells during the development of *AHA51/Df(2L)64j* flies (Fig 6.10). The breakpoint of the *hobo* deletion could not be repaired from the other chromosome as it carries a deficiency for the *Adh* locus, but the ends of the breakpoints might join together after deleting one duplication site.

The somatic movement of the *hobo* element could occur at different developmental stages. If the *hobo* element was deleted in an early developmental stage, a large proportion of cells would lack *hobo* elements and give rise to a strong small band during PCR amplification (Fig 6.10, lane 6). If *hobo* was deleted at a late developmental stage, only a small proportion of cells would lack *hobo*, so a strong large band would be produced during PCR amplification (Fig 6.10, lane 5).

The third structure observed during *hobo* excision was that the *hobo* element was excised perfectly and the breakpoint was repaired from an Adh^F allele as shown in line 2 (Fig 6.11). The only possible origin of the Adh^F allele was from the *CyO* chromosome which carries Adh^F and was used as a balancer to maintain the *AHA51* stock. In this situation the *hobo* element excision must

have occurred before meiosis in the gametes. When *hobo* excised, the breakpoint would be repaired from the Adh^F allele on the *CyO* chromosome.

Movement of the *hobo* element in the germ line has not been detected in the *CyO/AHA51* stock, as no one fly has ever been found with a normal ADH activity level. The stock is maintained at 20°C, but the flies used for experiments were always raised at 25°C so that *Cy* would be fully expressed. Therefore it is possible to speculate that germ line movement of *hobo* is temperature-dependent in *AHA51*.

To investigate germline excision of the *hobo* element in *AHA51* allele, further experiments need to be carried out. As shown in Fig 6.10, twelve single pair crosses between a *CyO/AHA51* male and a *CyO/Df(2L)64j* virgin female were set up at 25°C. Then, in the next generation, *AHA51/Df(2L)64j* flies were sacrificed for PCR and cellulose gel electrophoresis. Before being sacrificed for PCR, male flies of *AHA51/Df(2L)64j* could not be mated with *CyO/Df(2L)64j* females to preserve the *AHA51* allele for further investigation of any excision revertant. Because *AHA51/Df(2L)64j* flies are sterile, as are *AHA51* homozygotes, this makes further analysis difficult. A possible approach to overcome the sterility of *AHA51/Df(2L)64j* is to use other *Adh* deficiencies covering a narrower region than *Df(2L)64j*. If the sterility factor can be separated from the *AHA51* allele, it would be possible to obtain fertile *AHA51/Df* flies for further experiments. Another alternative strategy might be to score at least one hundred *AHA51/Df* flies grown at 25°C, assay ADH activity individually and then do PCR analyses on the same single fly. According to the mutation rate of *hobo* element at the *dpp* gene in the germ line, which is 1-5% of the progeny (Blackman and Gelbart, 1989), it might be possible to detect a few flies among the one hundred, in which germline excision of the *hobo* element has occurred.

The complete *hobo* elements present in the *AHA51* genome provided transposase activity essential for the mobilisation of the *hobo* element. *Hobo*

mobilisation is not limited to H x E crosses like the P-M system; it can occur in H x H crosses as well (Blackman and Gelbart, 1989). *Hobo* mobilisation is not only confined to the germ line (Blackman *et al*, 1989), but has also been observed in somatic cells (Blackman and Gelbart, 1989). There is evidence for maternal and zygotic repression mechanisms acting on *hobo* mobilisation in some strains (Ho *et al*, 1993; Yannopoulos *et al*, 1994). Several observations of chromosome aberration were found in single nuclei or sectors of nuclei in salivary gland polytene chromosome squashes (Lim, 1979; 1981; Yannopoulos *et al*, 1983). These rearrangements were believed to be *hobo* mediated because the breakpoints of these aberrations correlated well with known sites of *hobo* elements (Yannopoulos *et al*, 1987). Somatic reversion of the dpp^{d1} phenotype has been observed in homozygous individuals (Blackman and Gelbart, 1989). All these observations suggest that *hobo* may be somatically active, but no further evidence has been obtained at the molecular level. Recently Calvi and Gelbart (1994) have argued that somatic movement of *hobo* is a rare event and that previous reports of somatic movement probably involve special circumstances, such as mutations within host genes or *hobo* elements that result in somatic activity. Our investigation of *hobo* instability in *AHA51* provides direct evidence of the somatic instability of *hobo* at the molecular level, occurring sufficiently frequently to be readily detectable by PCR.

The excision of the *hobo* element from *AHA51* did not affect the ADH activity in adults, based on the level of ADH activity after electrophoresis and staining on acetate cellulose gels. It is possible that the *hobo* deletion occurred in cells which did not express ADH activity and therefore the effect of *hobo* excision on ADH activity was not detected. This speculation needs further analysis by using tissues which specifically express ADH activity (Anderson *et al*, 1991).

The experimental data in these chapter provide evidence that the KP and *hobo* element insertions are affecting *Adh* expression. The KP and *hobo* elements

located at this particular position, 5 bp 5' to the distal promoter transcription start site, could affect *Adh* expression in several ways. The insertions might increase the distance between the TATA box and transcription start site of the distal promoter, and therefore disrupt the core promoter structure of the *Adh* distal promoter. The insertions also increase the distance between the *Adh* adult enhancer (AAE) and the distal promoter. Some of the sequence motifs present in the KP or *hobo* elements might influence *Adh* expression by functioning as binding sites for transcriptional regulatory factors. All of these speculations will be further discussed in the following chapter.

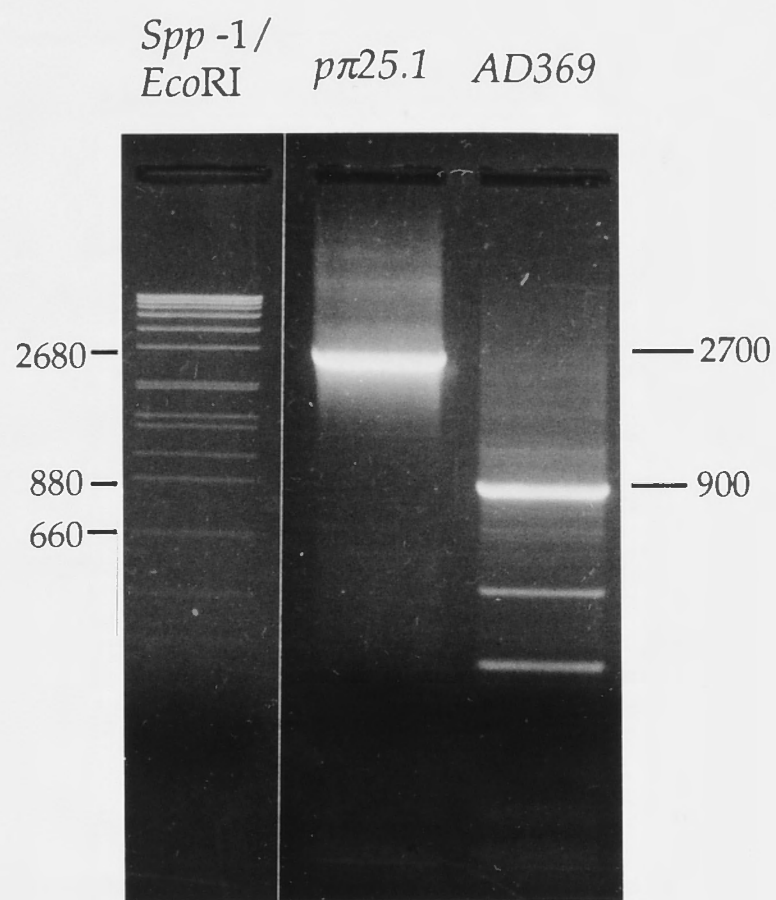


Figure 6.1 PCR amplification with P element specific primers

Plasmid DNA of *pπ25.1* and genomic DNA of *AD369* were amplified with a pair of primers JG68/67, which are complementary to P element sequences. DNA fragments amplified were electrophoresed on a 1% agarose gel and the estimated sizes of the fragments are indicated in basepairs (right side). *Spp*-1 DNA digested with *Eco*RI was used as a size marker and the sizes are in basepairs (left side).

Table 6.1 Survival of F_2 flies of the genotype *AD369*/nLA248* after exposure to ethanol. Flies were raised at 25°C according to the breeding program described in Figure 2.4.

Sample	Genotype	Age (days)	Number of flies tested^	Number of flies survived	Percentage of flies that survived
(+) control	*AC5/nLA248*	8	1 x 15	13	87%
(-) control	*AD369/nLA248*	8	1 x 15	0	0%
F_2 sample (1)	*AD369*/nLA248*	8	8 x 17	5	3.7%
F_2 sample (2)	*AD369*/nLA248*	6	5 x 16	21	26%

^ number of separate experiments and number of flies in each.

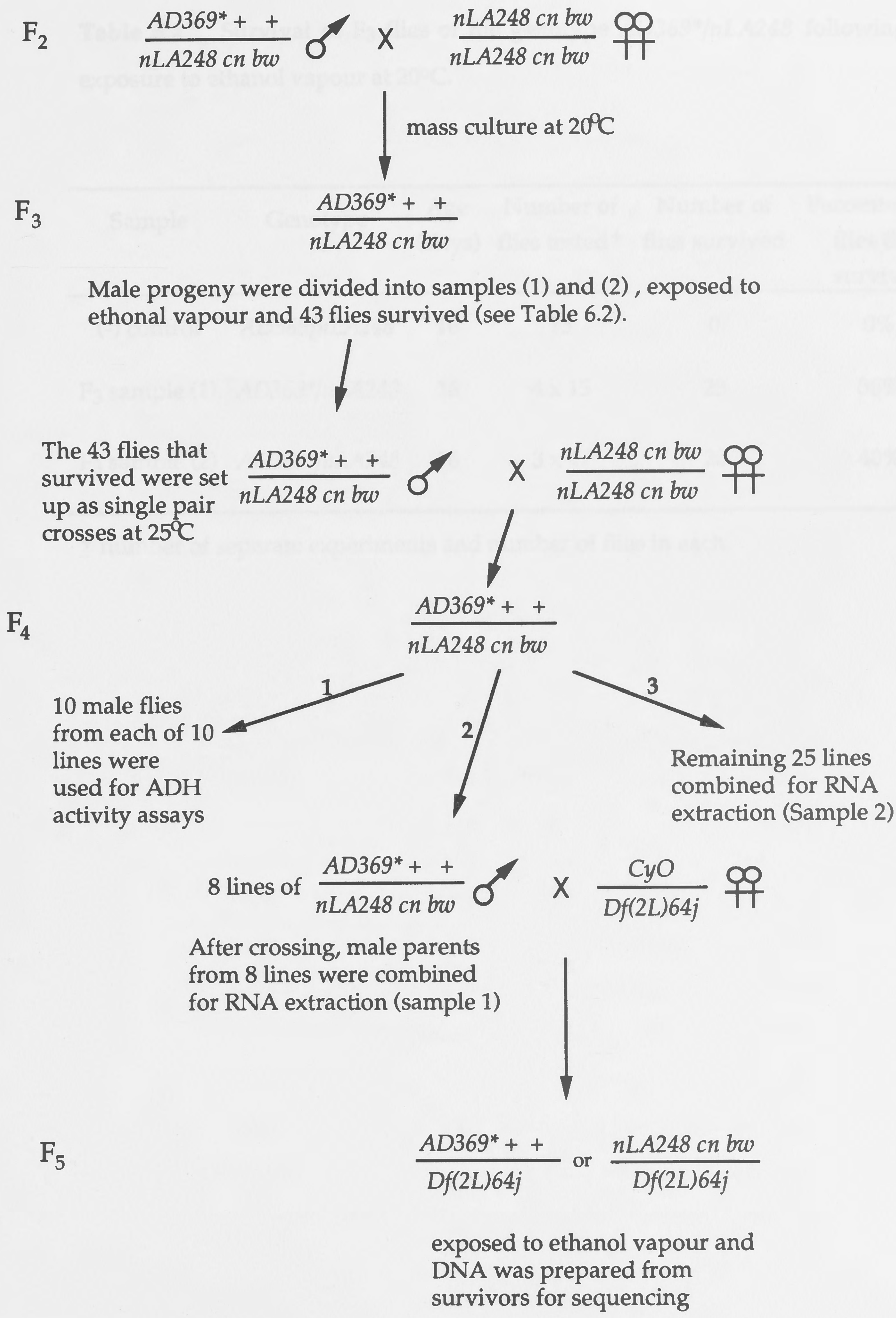


Figure 6.2 Breeding scheme for preservation and analyses of *AD369 allele**

Table 6.2 Survival of F_3 flies of the genotype *AD369*/nLA248* following exposure to ethanol vapour at 20°C.

Sample	Genotype	Age (days)	Number of flies tested+	Number of flies survived	Percentage of flies that survived
(-) control	*AD369/nLA248*	16	15	0	0%
F_3 sample (1)	*AD369*/nLA248*	16	4 x 15	23	38%
F_3 sample (2)	*AD369*/nLA248*	16	3 x 17	20	40%

+ number of separate experiments and number of flies in each

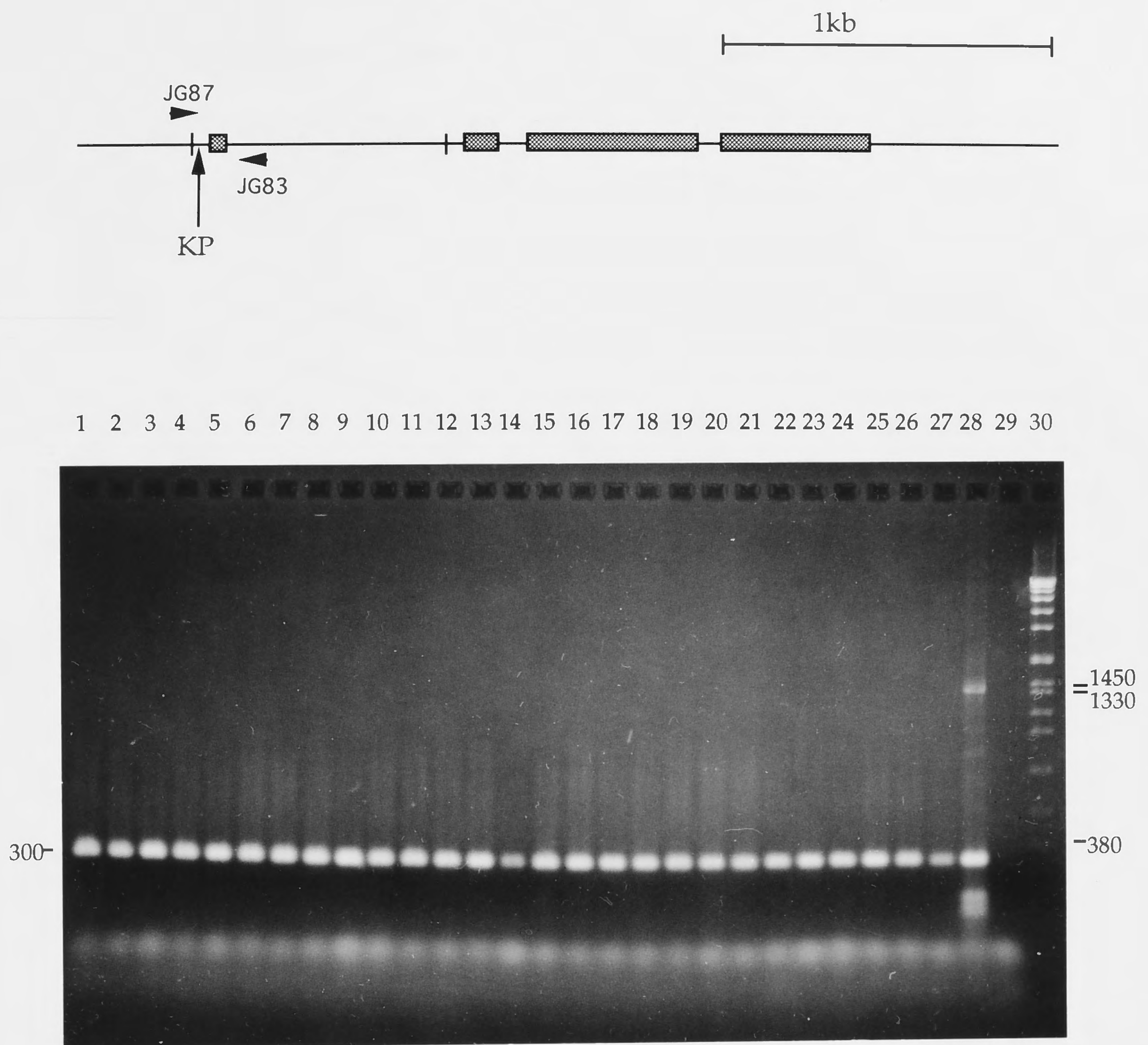


Figure 6.3 PCR amplification of *AD369*/nLA248* single fly DNA.

Twenty-seven single fly DNA extracts from *AD369*/nLA248* flies were amplified with the primers JG87/JG83, as indicated in the upper diagram. Amplified DNAs (lane 1 to 27) were run on a 1% agarose gel. Control DNA from *AD369/nLA248* was also amplified and loaded in lane 28. Lane 29 is a no DNA control . *Spp-1* DNA digested with *Eco*RI was used as a size marker and sizes are in basepairs. Estimated size of amplified band is indicated (left side).

Table 6.3 ADH activity in adult flies of *AD369*/nLA248* compared with *AC5/nLA248* and *AD369/nLA248* heterozygoutes. Each value is the mean of the assays of progeny from single pair crosses. The standard errors are shown in parenthesis.

Genotype	Number of crosses	ADH activity	Relative activity
AC5/nLA248	2	223.5(±8.5)	1
AD369/nLA248	2	19.75(±1.05)	0.088
AD369/nLA248*	10	113.6(±31.7)	0.508

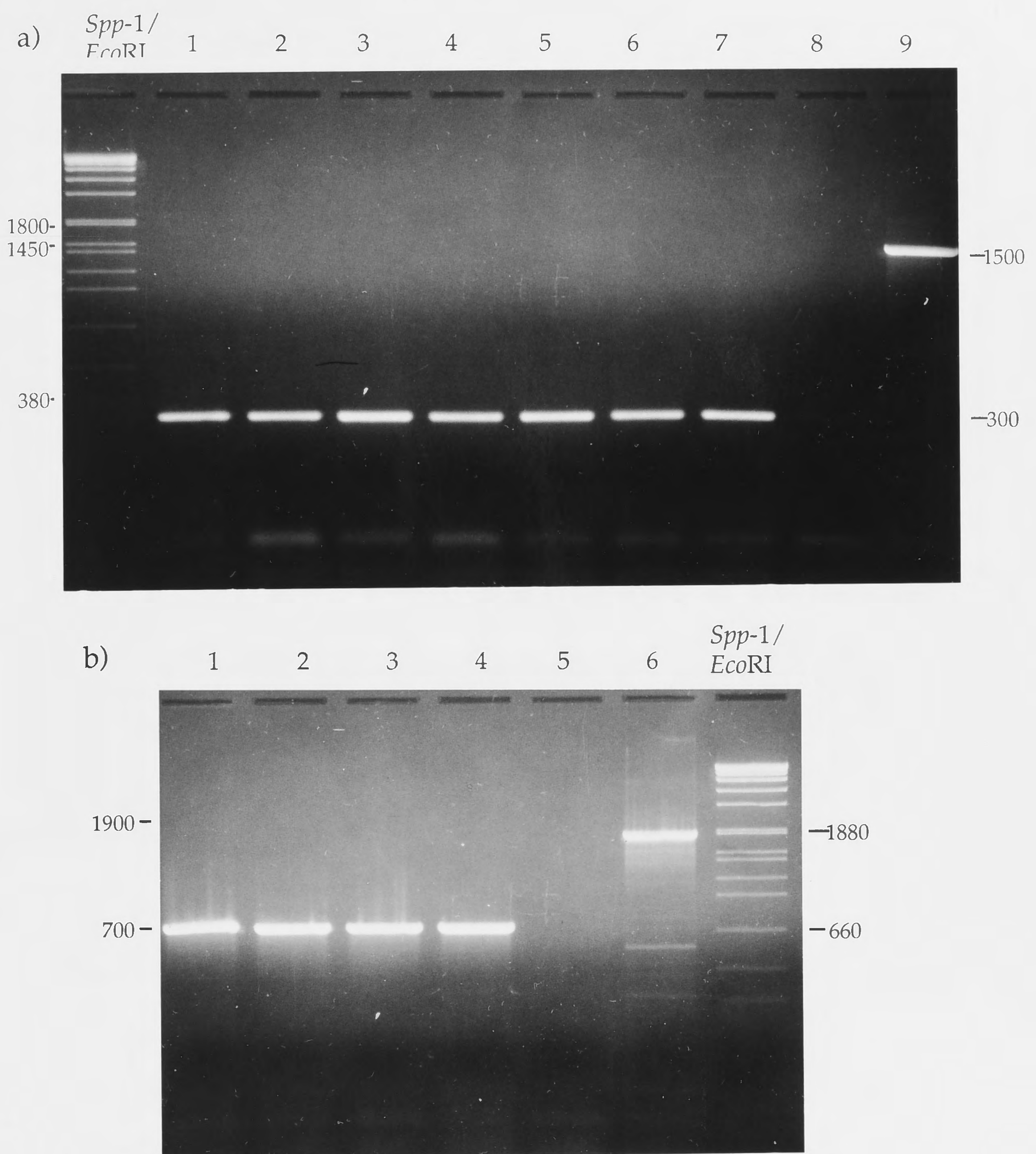


Figure 6.4 PCR amplification of *AD369*/Df(2L)64j*

Genomic DNA extracted from each of 8 lines of *AD369*/Df(2L)64j* was amplified with primers JG87/83 and seven of them are shown in (a) lanes 1 to 7. Four of the 8 lines were amplified with the primers JG88/22 for sequencing and shown in (b) lanes 1 to 4. Genomic DNA from *AD369* homozygotes was amplified and loaded in lane 9 (a) and lane 6 (b). Negative controls for PCR amplification are in lane 8 (a) and lane 5 (b). DNA fragments amplified were electrophoresed on a 1% agarose gel and the estimated sizes of the fragments are indicated in basepairs. *Spp*-1 DNA digested with *Eco*RI was used as a size marker and the sizes are in basepairs.

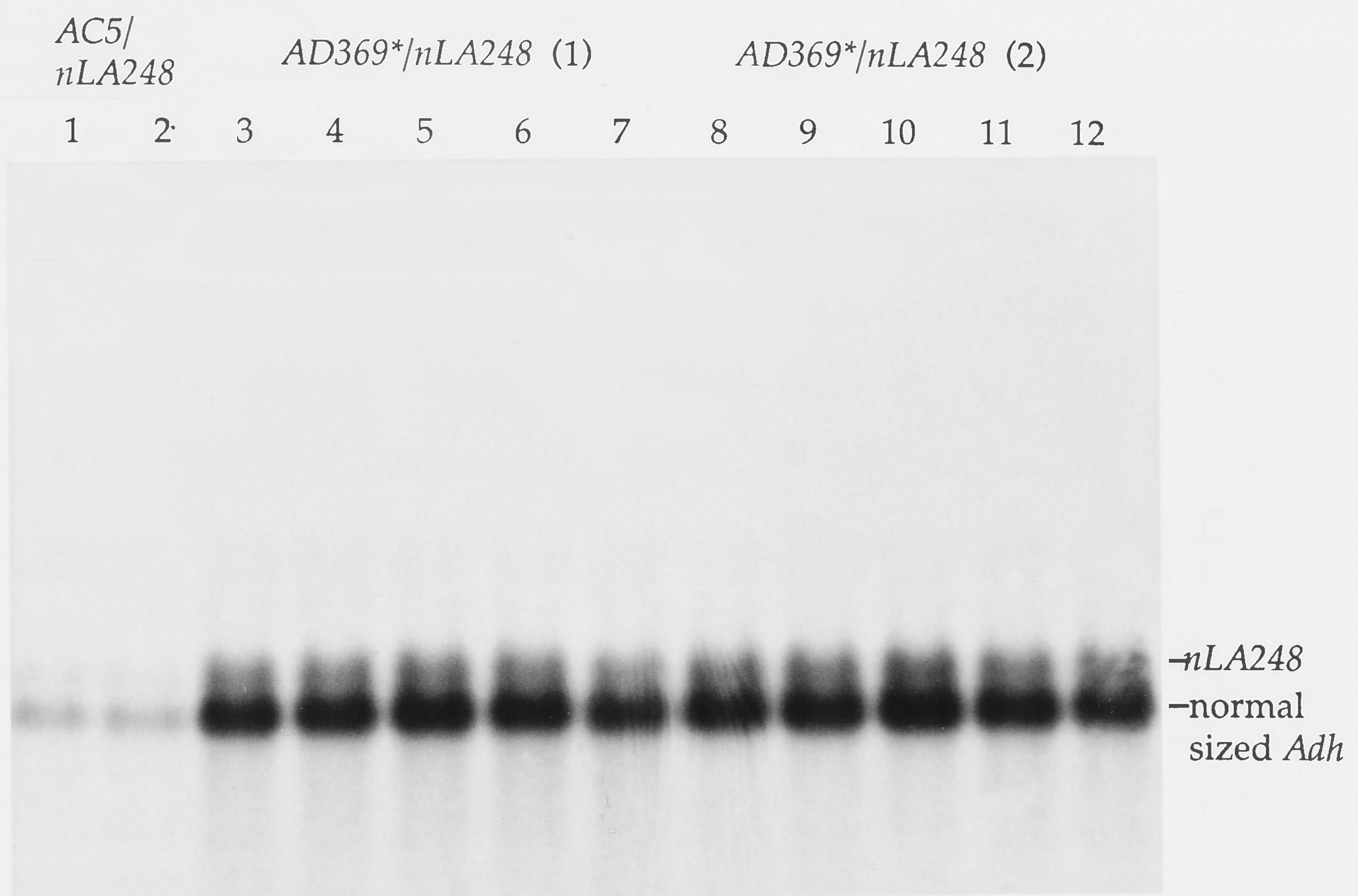


Figure 6.5 Quantitative Northern analysis for adults of *AD369*/nLA248* and *AC5/nLA248* heterozygotes

From left to right, samples were loaded as: lane 1 and 2, *AC5/nLA248*; lane 3 to 7, *AD369*/nLA248* sample (1); lane 8 to 12, *AD369*/nLA248* sample (2). The filter was probed with the nick-translated labelled *pSAF-2* (Goldberg, 1980). The upper band represents the *nLA248*-specific transcript. The lower band represents the normal sized *Adh*-specific transcript.

Table 6.4 Quantitative Northern Analyses for *AD369*/nLA248* and *AC5/nLA248* heterozygotes

The quantitative results of the Northern analysis shown in Figure 6.5 are summarised in this table. The intensity of each band was measured using the PhosphorImager instrument (Molecular Dynamics). Each value is the mean intensity of five measured samples. The relative *Adh/nLA248* specific intensity was calculated separately for each sample, to control for differences in RNA loadings. Only the average value of the relative intensity between the normal sized *Adh* band and the *nLA248* band is shown in the table. Standard errors are shown in parenthesis.

Genotype	*AC5/nLA248*	*AD369*/nLA248* (1)	*AD369*/nLA248* (2)
nLA248 1330 bases band intensity	572.7 (±13.4)	2867.2 (±987.8)	3035.8 (±688.9)
normal *Adh* 1130 bases band intensity	1712.5 (±204.4)	10128.2 (±1587.6)	10712.8 (±1412.8)
normal *Adh* : *nLA248*	2.99	3.53	3.53
*AD369** : *AC5*		1.18	1.18

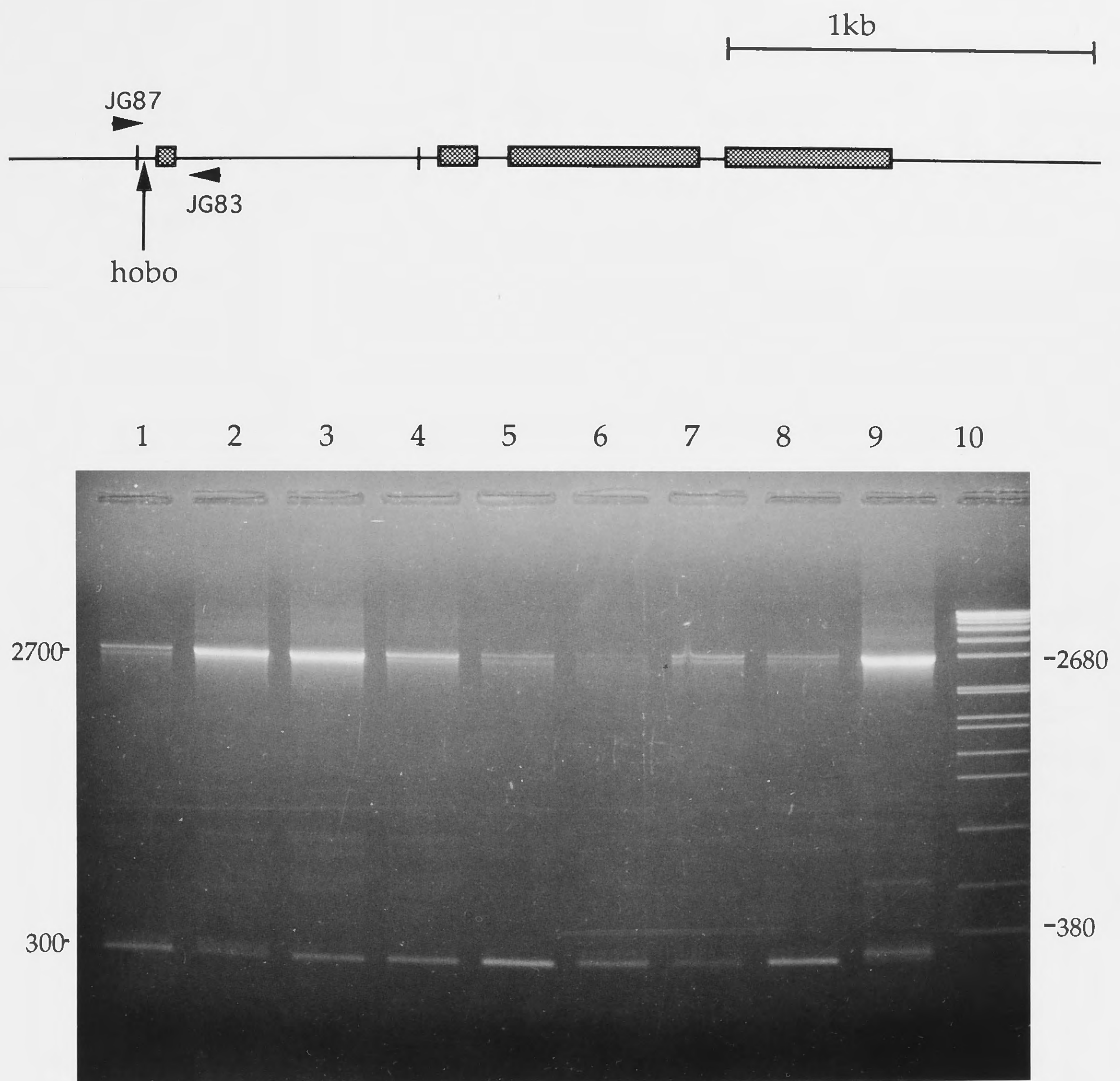


Figure 6.6 PCR amplification with single fly DNA extracted from homozygous *AHA51*.

DNA was extracted from single flies of *AHA51* homozygotes and amplified with the primer pair JG87/JG83 as indicated in the upper diagram. DNA fragments amplified by the PCR were electrophoresed on a 1% agarose gel (lane 1 to 9) and the estimated size of the fragments are indicated in basepairs (left side). *Spp-1* DNA digested with *Eco*RI was used as a size marker (lane 10) and the sizes are in basepairs.

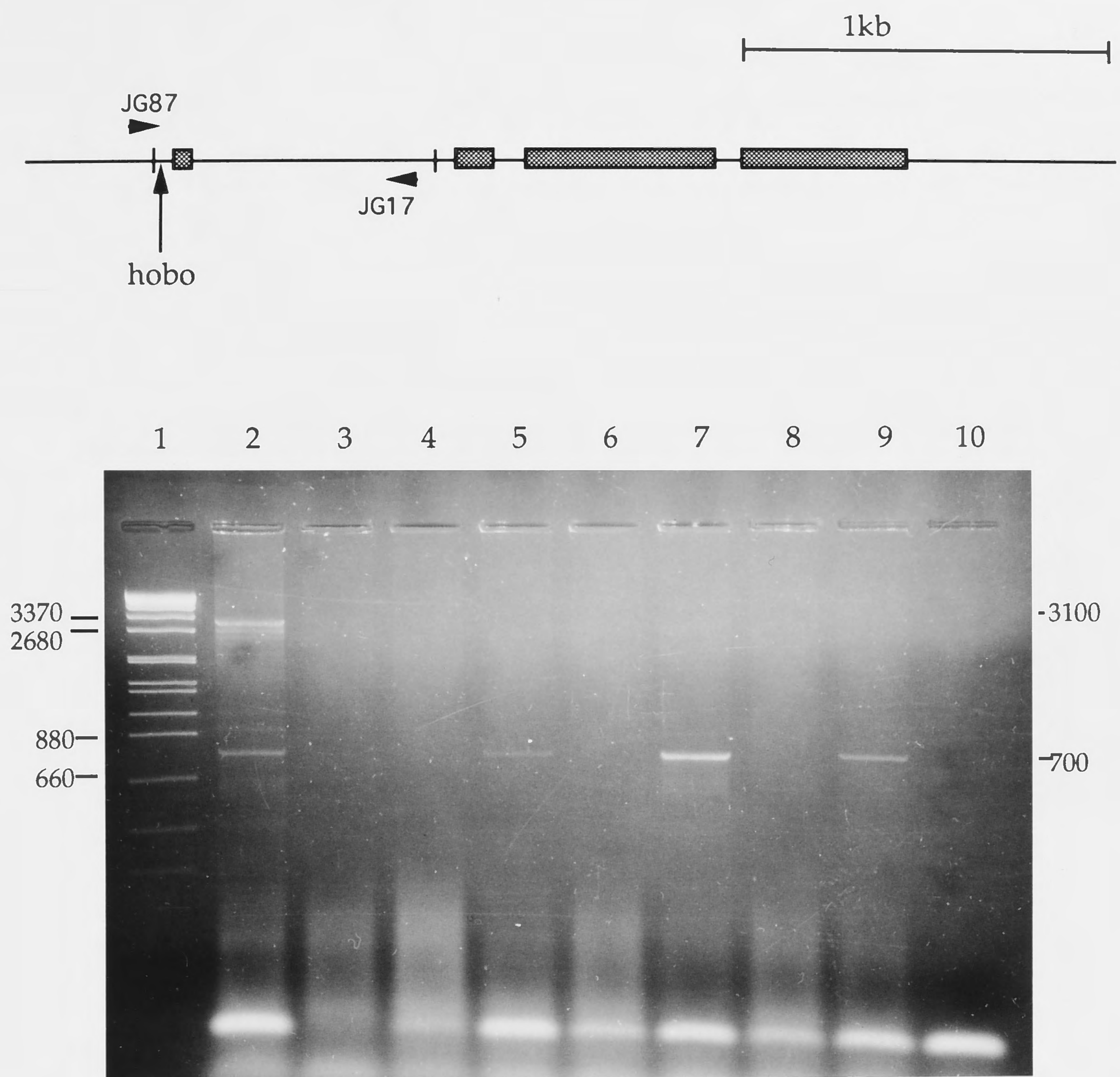
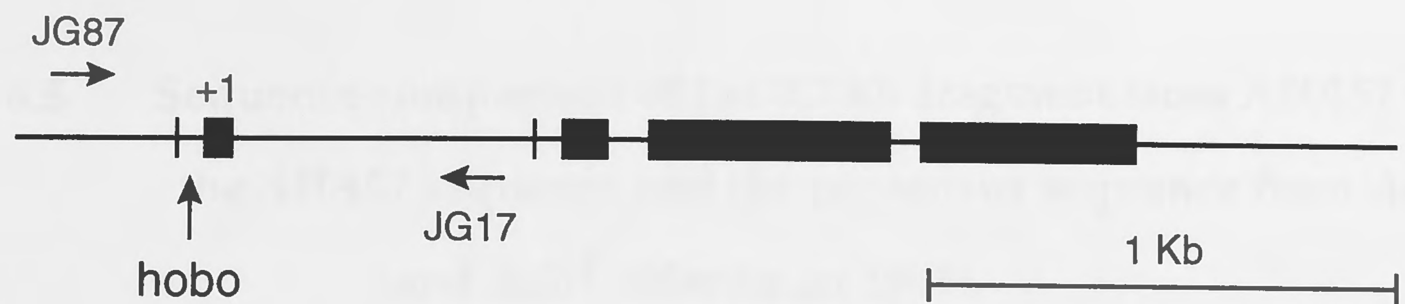


Figure 6.7 PCR amplification with single fly DNA extracted from hemizygous *AHA51/Df(2L)64j*.

DNA was extracted from single flies *AHA51/Df(2L)64j* and amplified with the primer pair JG87/JG17 as indicated in the upper diagram. DNA fragments amplified by the PCR were electrophoresed on a 1% agarose gel (lanes 2 to 10). The estimated size of the fragments are indicated in basepairs (right side). *Spp-1* DNA digested with *Eco*RI was used as a size marker (lane 1) and the sizes are in basepairs.



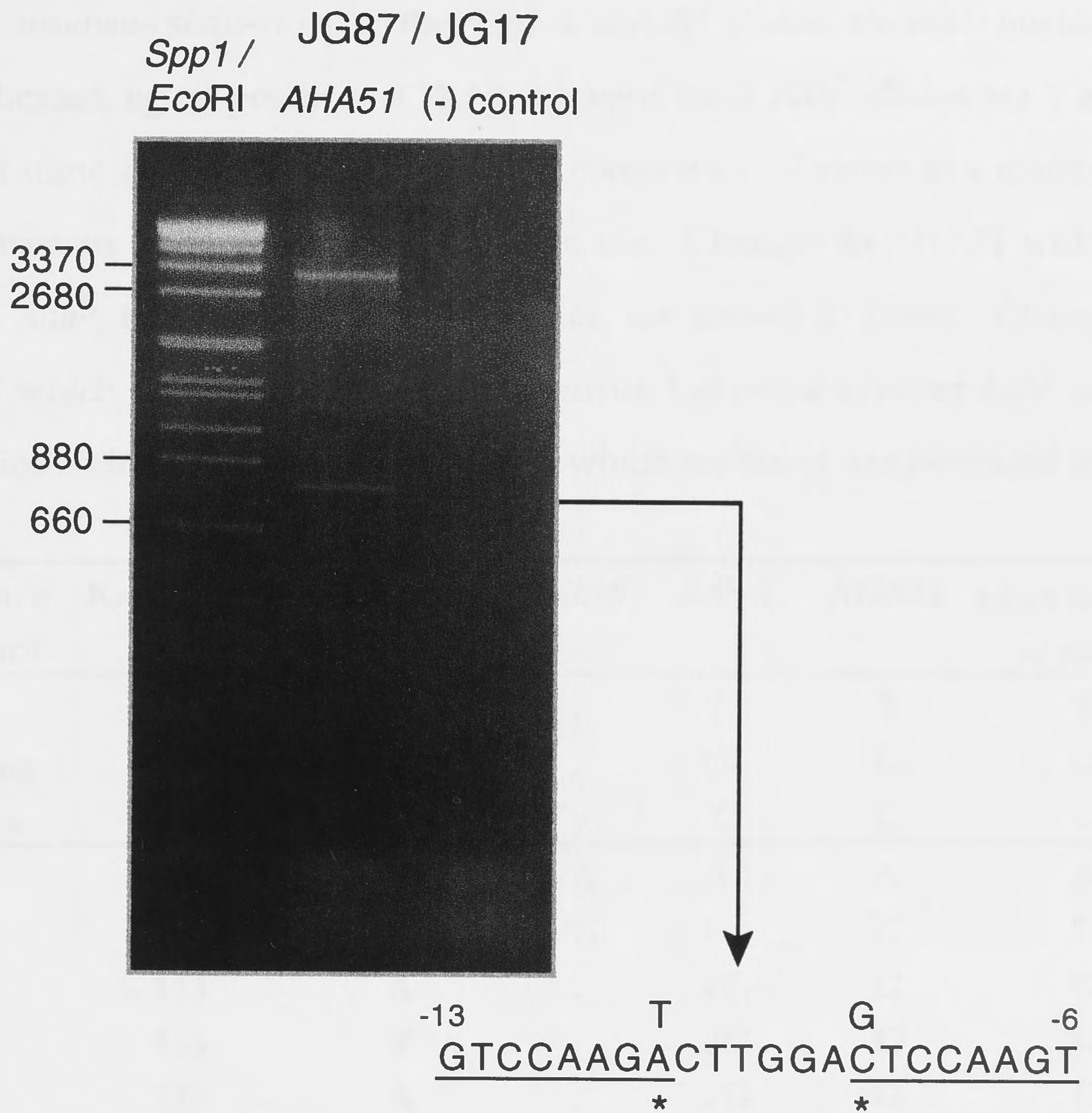


Figure 6.8 PCR amplification of DNA extracted from *AHA51* homozygotes
DNA extracted from *AHA51* homozygotes was amplified with primers JG87/17 as indicated in the upper diagram. *Spp*-1 DNA digested with *Eco*RI was used as a size marker and the sizes are in basepairs. The large band amplified in *AHA51* was of the size expected with a hobo element insertion present (see text 4.2.2.2.2), and the small band of the size expected for a normal allele (see text 6.2.2.2). The structure of the hobo excision breakpoint is shown at the bottom. The nucleotide postions marked with a star differed from the original sequence, which is indicated above.

Table 6.5 Sequence comparison of the 0.7 kb fragment from *AHA51* with the *AHA51* sequence and the consensus sequence from Adh^S and Adh^F (Kreitman 1983).

In the consensus sequences for the Adh^S and Adh^F alleles, alternate nucleotides are indicated, eg. at position -3 "T/.", it means most Adh^F alleles are T at this site but some are the same as the overall consensus. "." refers to a match with the consensus at the indicated nucleotide site. Changes in *AHA51* which are rare in Adh^S, but common in Adh^F alleles, are shown in green. Changes in *AHA51* which differ from the Adh^S consensus, but occur in other Adh^F alleles, are indicated in blue. Changes in *AHA51* which are novel, are indicated in red.

Refer.ence Sequence	Kreitman's No	Consensus	*Adh-S*	*Adh-F*	*AHA51*	0.7 kb fragment of *AHA51*
5' Flanking Sequence	-3	C	.	T/.	T	T
	-2	C	.	G/.	G	G
	-1	G	C/.	C	C	C
Intron 1 (Adult Intron)	107	C	./A	A/.	A	A
	113	A	./G	G/.	G	G
	143	A	.	./G	G	G
	169	T	.	./G	G	G
	173	A	.	./G	G	G
	175	T	./A	A/.	A	A
	287	G	./T	.	.	.
	293	G	./T	T/.	.	.
	304	G	./C	C/.	.	.
	423	A	.	.	.	.
	447-476	29bp	.	34bp	.	.
	516	C	.	G/.	.	.
	550	insertion2	.	insertion/.	.	.
	573	G	.	.	A	A
	576	G	.	.	C	C
	578	C	.	.	G	G
	586	G	.	./T	.	.

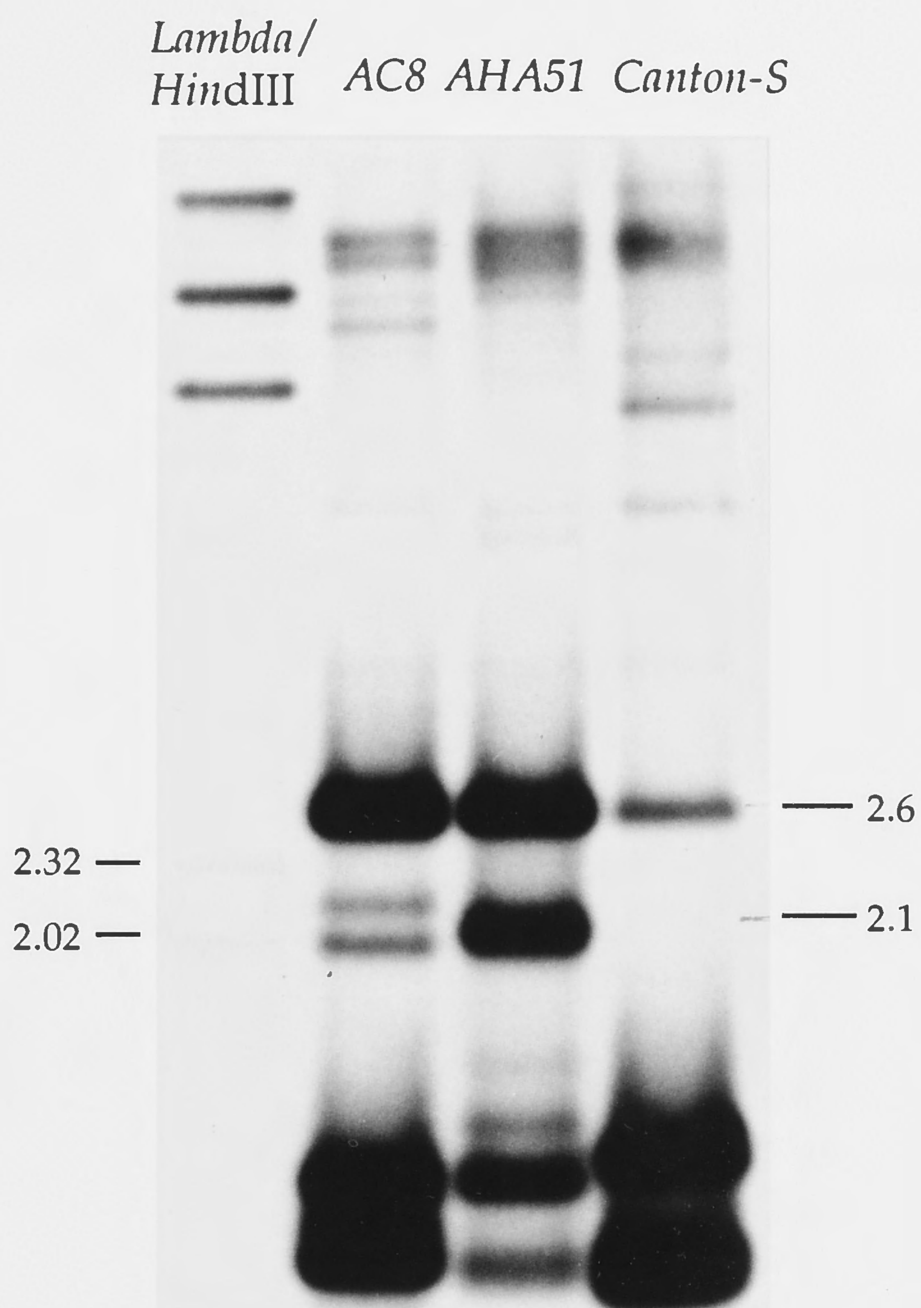
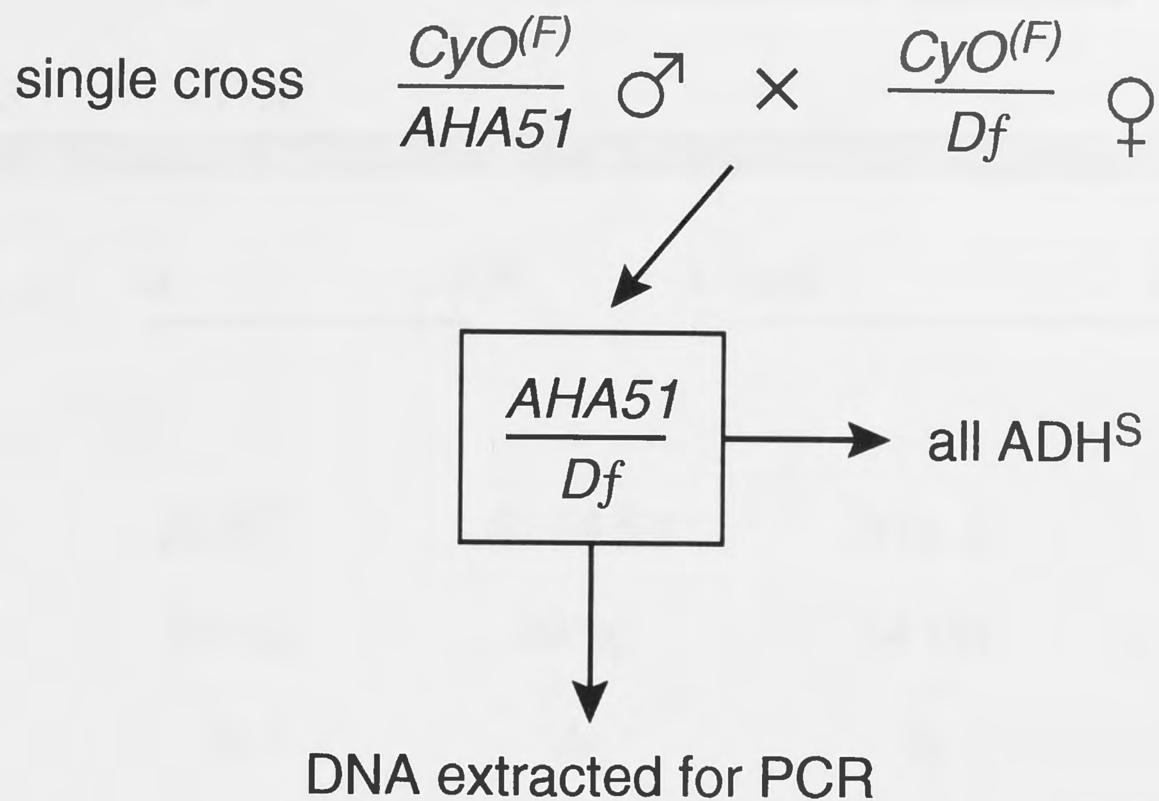


Figure 6.9 Hybridization pattern of *AHA51* genomic DNA probed with labelled hobo element.

Genomic DNA extracted from *AC8, AHA51* and *Canton-S* (loaded from left to right) was digested with *Xho*I, run on a 1% agarose gel and transfered to the nitrocellulose filter. The filter was probed with a nick-translated 2.1 kb *Xho*I fragment, which was isolated from the defective hobo element in *AHA51*. *Lambda* DNA digested with *Hin*dIII was used as a size marker and the sizes are in kilobase pairs.



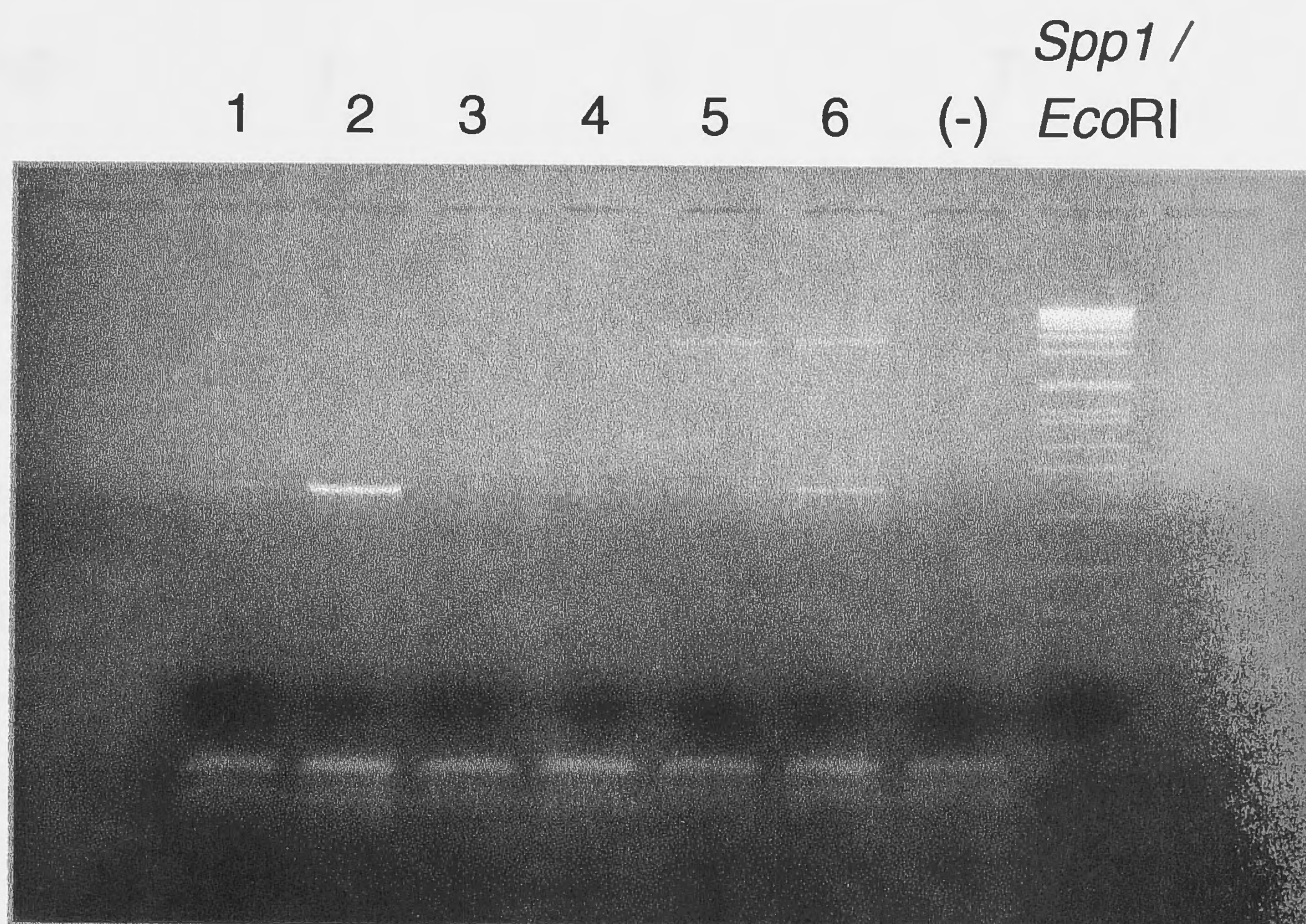
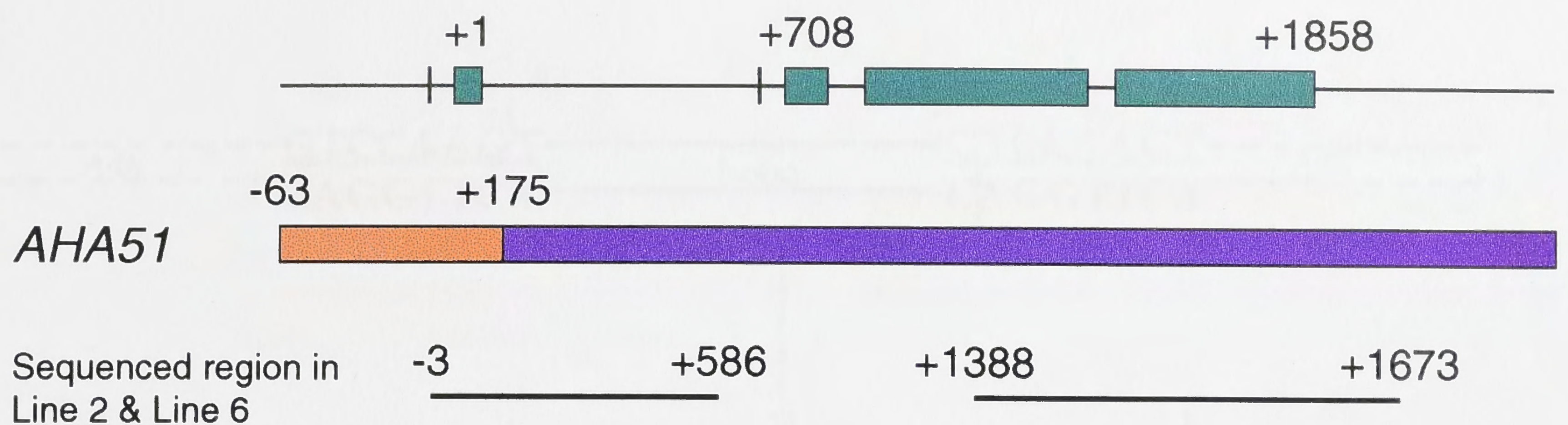


Figure 6.10 PCR amplification with DNA extracted from *AHA51/Df(2L)64j*

AHA51/Df(2L)64j flies obtained from a single pair cross shown in the diagram were collected for analysis. DNA extracted from each line was amplified with the primer pair JG87/17. The amplified DNA fragments were run on a 1% agarose gel (lane 1 to 6). A control for PCR amplification was loaded in lane (-). *Spp*-1 DNA digested with *Eco*RI was used as a size marker.



K. no	Adh^F	*AHA51*	line 2	line 6
447 -476	34 bp	29 bp	34 bp	29 bp
573	G	A	G	A
1443	G	C	G	C
1452	T	C	T	C
1490^+	C	A	C	A
1518	T	C	T	C
1527	C	T	C	T
1557	C	A	C	A
1596	G	A	G	A

Figure 6.11 Comparison of nucleotide differences between the *Adh* alleles from line 2 and line 6

The *Adh* gene of two particular lines, line 2 and line 6, from the single cross shown in Fig 6.10 were sequenced in the regions indicated. The results show that line 2 has the same sequence as an Adh^F allele (red columns) and line 6 has the same sequence as the *AHA51* allele (blue columns). The nucleotides which differ between the Adh^F allele and the *AHA51* allele are listed in the table. The upper diagrams show a normal *Adh* gene with the exons indicated as green boxes, and the unique haplotype of the *AHA51* gene in which the region 5' to nt +175 is like an Adh^F allele (shown as the orange box) and 3' to nt +175 is like an Adh^S allele (shown as the purple box).

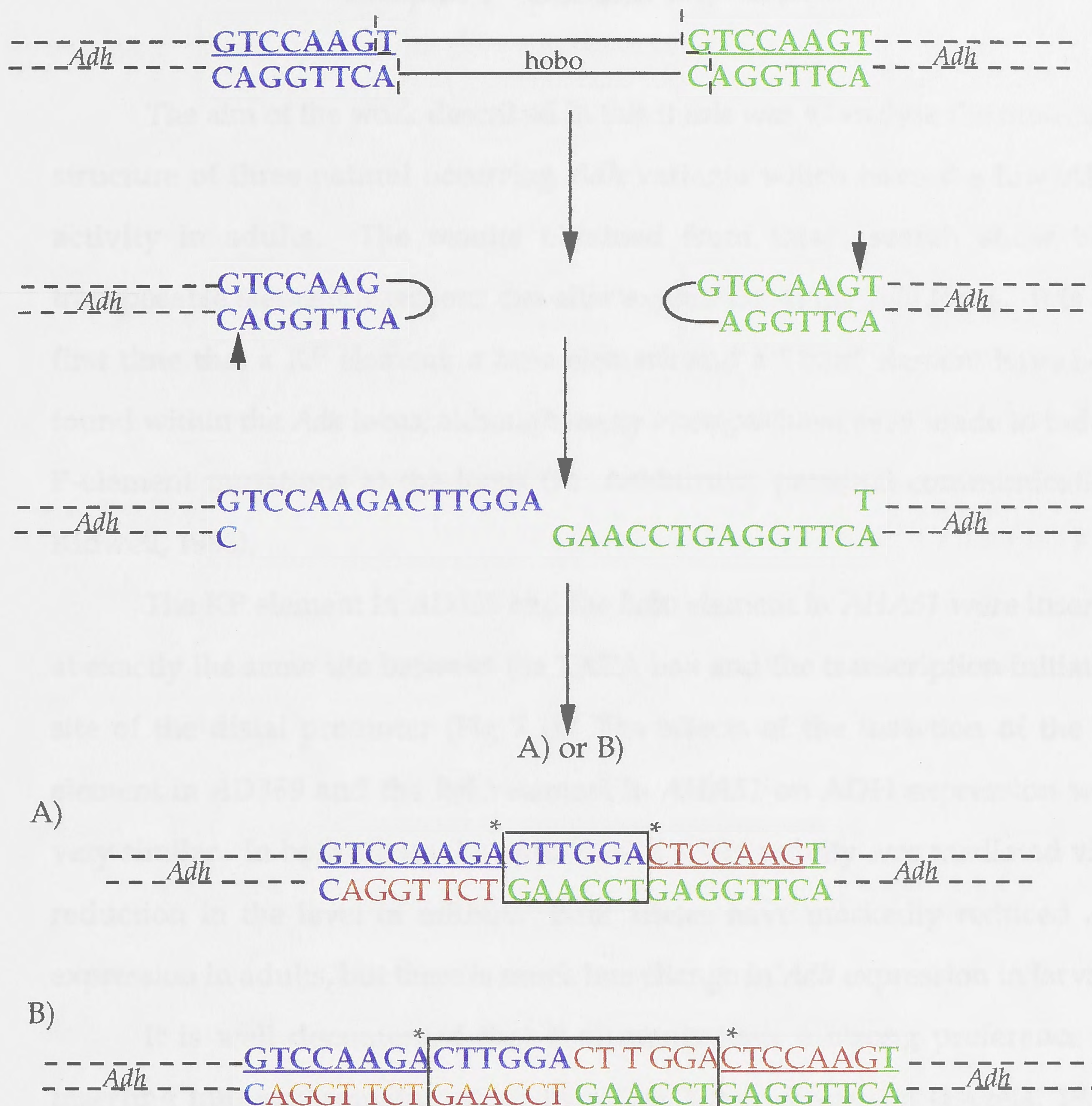


Figure 6.12 Possible mechanism for hobo excision and post-excision repair. The structure of the hobo element insertions and the structure after hobo excision in *AHA51* are shown in the diagram. Double dashed lines represent the *Adh* sequences and double solid lines represent the hobo sequences. The 8 bp target site duplications are underlined and the hobo was excised with a 1-bp 5' overhang as indicated by dashed broken lines. The broken sites in the hairpin structures are indicated with arrows. The sites that differed from the original sequence in the target site duplications are marked with stars. The newly formed inserts are boxed with newly filled nucleotides in red colour.

Chapter 7 General Discussion

The aim of the work described in this thesis was to analyse the molecular structure of three natural occurring *Adh* variants which have the low ADH activity in adults. The results obtained from this research show how transposable element insertions can alter expression at the *Adh* locus. It is the first time that a KP element, a *hobo* element and a *Tirant* element have been found within the *Adh* locus, although many attempts have been made to induce P-element mutations at the locus (M. Ashburner, personal communication; Kidwell, 1986).

The KP element in *AD369* and the *hobo* element in *AHA51* were inserted at exactly the same site between the TATA box and the transcription initiation site of the distal promoter (Fig 7.1). The effects of the insertion of the KP element in *AD369* and the *hobo* element in *AHA51* on ADH expression were very similar. In both alleles the reduction of ADH activity was mediated via a reduction in the level of mRNA. Both alleles have markedly reduced *Adh* expression in adults, but there is much less change in *Adh* expression in larvae.

It is well documented that P elements have a strong preference for inserting into euchromatin rather than into heterochromatin (Engels, 1989). Both KP and *hobo* elements have a tendency to insert into the 5' regions of transcription units which are normally involved in the regulation of gene expression (McGinnis *et al*, 1983; Tsubota *et al*, 1985; Chia *et al*, 1986; Searles *et al*, 1986; Kelley *et al*, 1987; Roiha *et al*, 1988; Reed and Gibson, 1993). It is interesting to find both the KP and the *hobo* elements located at exactly the same site between the TATA box and the transcription start site of the distal promoter at the *Adh* locus. It has been reported that *hobo* elements are distributed differently with respect to P elements. Comparison of *hobo* and P insertion patterns on third chromosomes in a *hobo* enhancer trapping mutagenesis experiment showed that P elements tend to be clustered in the

proximal region of the right arm, while the *hobo* elements tend to be found in the more distal locations of each arm (Smith *et al*, 1993).

There is no evidence of any association between the insertion sites of the two elements (Smith *et al*, 1993). The 8-bp host genomic sequences duplicated upon the insertion of P and *hobo* elements share very little similarity. The common insertion site (GTCCAAGT) in *AD369* and *AHA51* matches the consensus sequence for the KP element insertion (GNCCAGAC) in 5 out of the 8 nucleotides, and it also matches the sloppy consensus sequence for the *hobo* element insertion (NNNNNNAC). There is no homology between the P transposase and the putative *hobo* transposase. In addition, the transposition mechanism in the germline is different for *hobo* and P elements (Calvi and Gelbart, 1994). Therefore, it can be speculated that the common insertion site for KP and *hobo* elements at the *Adh* locus is coincidental and might be related to some specific chromatin conformation of the *Adh* distal promoter in the germline.

The chromatin structure of the *Adh* gene in the proximal promoter, distal promoter and AAE region has been shown to be accessible during *Adh* expression (Felsenfeld, 1992; Jackson and Benyajati, 1992; 1993). Such an open chromatin structure is hypersensitive to DNase I cleavage and is nucleosome free and accessible to non-histone DNA binding proteins. The region between the distal promoter and AAE contains a nucleosome positioning element (NPE) which is occupied by Adf-1 and the TATA binding factor TFIID in cells actively transcribing distal *Adh* transcripts. While in cells not transcribing *Adh*, this NPE region is occupied by a positioned nucleosome and Adf-2 (Ewel *et al*, 1990; see Chapter 1).

The alteration from condensed chromatin structure with positioned nucleosome during the stages not-transcribing *Adh* to the accessible DNase I hypersensitive chromatin structure during transcribing might facilitate the insertion of transposable elements at the distal promoter region. A portion of

yeast transposable element Ty917 was known to be responsible for regulation of the expression of the adjacent *HIS4* gene. The chromatin at the *Ty/HIS4* junction region was shown to be accessible to micrococcal nuclease (Feaver and Pearlman, 1991). It has been suggested that P elements preferentially insert into 5' nontranscribed or 5' transcribed, untranslated region, which may be due to an altered chromatin structure as detected by the presence of DNase I hypersensitive sites (Smith and Corces, 1991).

The sequence of the KP element inserted in *AD369* is identical to the KP element isolated from the Krasnodar strain (Black *et al*, 1987), as well as to that found at the *Gpdh* locus in Australian populations (Reed and Gibson, 1993). The KP element insertion in *Gpdh* is also located between the TATA box and the transcription start site, and interferes with the normal transcriptional process and results in decreased levels of *Gpdh* mRNA transcript (Reed, thesis 1993). It has been shown that the KP element in the *Gpdh* allele does not affect the transcription start site.

Whether the KP element insertion affects the utilisation of the distal promoter, or/and alters the transcription start site in *AD369* needs to be investigated by more detailed study, such as primer extension and S1 mapping. Although the size of the *Adh* transcript in *AD369* is the same as that of the wild-type transcript, the RNA analyses are not accurate enough to decide whether it was generated from the normal transcription start site of the distal promoter. The distal-specific probe *BSSBN4* did not hybridise in the experiments described in Chapter 3, therefore, it is hard to tell if the decrease of transcripts level in adults is due to the reduction of the transcripts from the distal promoter. Some other DNA fragments, for example a primer which specifically binds to the distal promoter, can be used as a probe to establish whether the distal promoter is being utilised properly (see Appendix V).

When the KP element was deleted from *AD369*, ADH activity did not completely return to the normal level. Possible explanations for why the

activity remained lower than expected in *AD369*, but not the level of *Adh* mRNA, have been discussed in Chapter 6. One possible explanation is that the small insertion located in the upstream region of the *Adh* gene has an effect on activity (see Fig 4.5).

The *hobo* element insertion in *AHA51* affected the expression of the *Adh* gene resulting in a lower level of transcript, but did not alter the size of the *Adh* mRNA. Similar phenomena caused by *hobo* element insertions have been observed at other loci. A defective *hobo* element insertion into the TATA box of the *Sgs-4* allele decreases the expression of the glue protein 50- to 100-fold, but does not affect the developmental pattern of expression of this gene (McGinnis *et al*, 1983). In addition, a *hobo* element insertion into the region of *gypsy*, to which the *suppressor of Hairy-wing [su(Hw)]* protein binds in y^2 flies, results in a partial y^2 revertant phenotype (Geyer *et al*, 1988b).

The *hobo* insertion in *AHA51* was at exactly the same site as the KP element in *AD369*, and showed similar effects. This result suggests that the actual sizes and intrinsic properties of the two types of element are not as important for the effect as the insertion site.

There are several possible mechanisms to explain how the KP and *hobo* elements affect *Adh* expression from the distal promoter. First of all, it is possible that the reduction of the *Adh* transcript level is due to the fact that the KP and *hobo* element insertions increase the distance between the TATA box and the transcription start site. This change in orientation could affect the interaction between the RNA polymerase and other transcription factors binding to the promoter. Another example of a KP element insertion located between the TATA box and transcription start site has been found at the *Gpdh* locus in *Drosophila melanogaster* (Reed and Gibson, 1993). A H.M.S. Beagle element splitting the promoter region of cuticle protein 3 (CP3) gene is evidently responsible for abolishing *CP3* gene expression (Sydney *et al*, 1982).

The TATA box together with upstream sequences are necessary for efficient transcription *in vivo* (McKnight *et al*, 1981; Grosschedl and Birnstiel, 1980). The TATA box directs the RNA polymerase to the initiation site. Both the TATA box and the initiation site are in direct contact with the RNA polymerase and the accompanying initiation complex (composed of numerous transcription factors), therefore the spacing between these two elements is crucial for efficient transcription (Hahn *et al*, 1989; Johnson and McKnight, 1989; White and Jackson, 1992). The TATA box is usually located 25 nucleotides upstream of the transcription start site (Alberts *et al*, 1994). When a linker is inserted between the TATA sequence and the transcription start site of the rabbit β-globin gene in mouse cells, RNA polymerase II initiates at a constant distance of 25 to 29 bp downstream from the TATA box (Dierks *et al*, 1983). Evidence from experiments increasing and decreasing the spacing between the TATA box and the transcription start site of the trout protamine gene (TPG-3) has shown that if the distance between the TATA box and transcription start site is increased or decreased by up to 3 base pairs, there is a change in the initiation site (Kovacs and Butterworth, 1986). Mutations in the spacer region between the TATA box and transcription start site show that a 4-bp insertion reduces the transcription activity to 30% of the wild-type $tRNA^{val}$ expression in *Methanococcus vannielii* (Hausner *et al*, 1991).

The KP element insertion can introduce a new TATA sequence instead of the normal one. There is a TATA-like sequence, TTATGTTATTT, in the 3' terminal inverted repeat of KP element. This sequence could be used as an alternative TATA box for *Adh* distal transcript initiation, which might give rise to a new initiation site a few base pairs different to the original one. The alternative TATA box and initiation start site are more likely to affect transcription from the distal promoter. The selective utilisation of the distal promoter is not only dependent on the upstream enhancer, but also is highly dependent on the core promoter elements of the *Adh* gene. The presence of the

distal initiator sequence (ATTATT) is linked to high levels of distal transcription. When the distal initiator sequence was substituted by the proximal motif (AACAAC), transcription from the distal template was strongly reduced relative to the original distal promoter (Hansen and Tjian, 1995). Also, mutation of the TATA box motif, TATTT, into the sequence CCTTG abolished distal promoter activity (Hansen and Tjian, 1995). Further investigation can be carried out by experiments of primer extension, S1 mapping or RACE to analyse the 5' termini of the adult transcript and establish the transcription start site in *AD369* adult transcripts.

The insertion of KP and hobo elements between the TATA box and the transcription initiation site might affect the assembly of the initiation complex. In the basal transcription apparatus, TFIID is the first component to assemble on the core promoter and then it initiates the recruitment of RNA Pol II and other GTFs (see Fig 1.0; Heard, 1995; Roeder, 1996). Binding of TBP to the TATA element forms a stable complex. Then TFIIA binds to this complex through direct contacts with TBP and upstream DNA sequences. TFIIB then binds through direct interaction with TBP and downstream sequences of TATA elements. The contact sites of TFIIB are located at positions -23 to -14. Following the TFIIB binding, pre-formed TFIIF-RNA Pol II complex contacts with TFIIB and DNA sequences that lie between the TFIIB contact site and position +17. TFIIE binds through direct interactions with RNA Pol II. TFIIH binds by direct contacts with TFIIE and completes assembly of the pre-initiation complex (PIC). Protein-DNA contacts extend to position +30 in the complete PIC formed on the adenovirus Major Late Promoter (MLP) by general transcription factors and RNA polymerase II (Buratowski *et al*, 1989; Van-Dyke *et al*, 1988).

It has been reported that the TAF fraction, not the TBP, of the TFIID complex plays a critical role in the *Adh* distal promoter selection process (Hansen and Tjian, 1995). The TFIIA fraction was required for maximum *Adh*

distal promoter activity. If the KP and hobo element insertions do induce new TATA box and transcription start sites to the *Adh* distal promoter, the assembly of TFIID and TFIIA could be affected because of the new TATA box motif. The alternative TATA motif and its surrounding sequences differ from the original core promoter of the *Adh* gene, and might interfere with the TBP recognising TATA sequences and the following TFIIA association. Another TFIIF-RNA Pol II complex, which contacts with the DNA sequence between -14 to +17, is more likely to be affected by the insertion of KP or hobo element at position -5/-6.

It should be possible to replace the KP or *hobo* element insertion by unrelated DNA of a similar length to see if it gives rise to similar effects on *Adh* expression. As the distal promoter mainly affects *Adh* expression in the fat body, both in larvae and adults, it would be useful to isolate tissues from *AD369* and *AHA51* adults that specifically express *Adh* to analyse the transcript levels.

The KP and *hobo* element insertions also increase the distance between the *Adh* adult enhancer (AAE) and the distal transcription start site. The AAE contains the binding site of a few activators and repressors which interact with the distal promoter (see Chapter 1). Usually, DNA sequences between the enhancer and the promoter loop out to allow the proteins bound to the enhancer to interact directly either with the general transcription factors or with RNA polymerase (Alberts *et al*, 1994). The KP or *hobo* element may interfere with this process so that the factors binding to AAE will not interact efficiently with the polymerase to initiate transcription.

A similar mechanism has been postulated for a *copia* element insertion located between the adult enhancer and the two promoters, which causes the reduction of *Adh* expression in the *RI42* allele (Dunn and Laurie, 1995). This low activity allele contains a *copia* insertion 243 bp upstream of the distal promoter (Strand and McDonald, 1989). It has been clearly shown that the *copia* insertion makes a contribution to the lower *Adh* expression in *RI42*, but not by

simply disrupting the sequences which it inserted (Dunn and Laurie, 1995). The *copia* insertion is located between the *Adh* enhancers and the distal promoter, but has different effects on the distal and proximal promoters. Possibly, the *copia* insertion reduces the *Adh* distal transcription rate by transcriptional interference from itself, or by displacing and inactivating the enhancers (Dunn and Laurie, 1995).

There are a number of activators and repressors bound to the *Adh* distal promoter and AAE regions that regulate *Adh* expression (see Chapter 1). The functions of the activators bound to the distal promoter regions could be interfered with due to structural changes resulting from the KP or *hobo* element insertions. The KP or *hobo* element is inserted close to the Adf-1 binding site (-85 to -47) and the distal promoter (-32 to -25) (Fig 7.2). It is possible that the insertions affect binding of Adf-1 and TFIID, which are transcriptional activators binding to the distal promoter region during *Adh* transcription. On the other hand, the KP or *hobo* element insertion might activate the repressors in the distal promoter region, such as Adf-2 (Alberts *et al*, 1994).

The KP and *hobo* element insertions might also change chromatin structure and thus affect *Adh* expression. It has been observed that accessible chromatin structure is present when the *Adh* gene is actively transcribed (Jackson and Benyajati, 1992; 1993). It is possible that the insertions change the active accessible chromatin structure into condensed heterochromatin and interfere with interactions between proteins bound in the accessible regions of the distal promoter and AAE. Therefore *Adh* transcription from the distal promoter is interfered with. Whether any chromatin structural changes are caused by the KP or *hobo* element insertions could be investigated by analysing the sensitivity to DNase I or MNase of nuclei extracted from adult tissues of these lines.

There could be factors that bind to the KP and *hobo* elements that inhibit *Adh* transcription. A repressor, such as the steroid receptor DHR39 which

binds to AAE (Ayer *et al,* 1993), might have binding sites in the KP and *hobo* elements. Therefore, when it binds to the KP or *hobo* element, it affects normal *Adh* expression. Also there could be different repressors binding to the KP or *hobo* element respectively, but giving rise to the same effects. Similar binding regions have been found in the retrotransposon *gypsy*. In the y^2 variant, the *gypsy* element is inserted about 700 bp upstream from the *yellow* promoter, which contains a binding site for the *suppressor of Hairy-Wing* [*Su(Hw)*] protein (Parkhurst and Corces, 1986a; Geyer *et al,* 1986; 1990). *Su(Hw)* protein binds to a particular sequence of *gypsy* containing twelve copies of a 10-bp sequence, 5'-PyPuTTGCATAC-3', which are separated from each other by AT-rich spacer regions (Spana *et al,* 1988; Mazo *et al,* 1989; Spana and Corces, 1990). When the *Su(Hw)* protein binds to the *gypsy* element it decreases transcription from the *yellow* gene. A partial deletion or the insertion of other transposable elements into this binding region, results in a reversion of the y^2 mutant phenotypes (Geyer *et al,* 1988a; Peifer and Bender, 1988).

The KP and *hobo* element insertions slightly affect *Adh* expression from the proximal promoter in larvae. Obviously, the insertions increase the distance between the *Adh* larvae enhancer (ALE) and the proximal promoter, and this might affect polymerase binding and interfere with the normal interaction between the ALE and the proximal promoter. This can be tested by removing the insertions and then analysing ADH activity and transcript levels in tissues specifically expressing *Adh* from the proximal promoter, such as the larval fat body, mid gut and Malpighian tubules.

The insertion of *Tirant* in *AAS44* gave rise to a different phenotype than that of the other two alleles. Although the *Tirant* element insertion did not appear to alter the size of the transcript, it decreased the level of the transcripts in adults, but not by as much as the KP and *hobo* elements. The insertion of the *Tirant* element was located in the adult leader sequence (first intron) between

the two promoters (see Fig 5.9) and how it affects *Adh* expression from the upstream distal promoter is still under speculation.

Studies have indicated that mutations within introns can alter gene expression (Bennetzen *et al*, 1984). Most of the mutants appear to alter intron processing and produce mature mRNAs containing some portion of the introns. An *Adh* null allele, Adh^{nHA52}, has been shown to contain an 8-bp insertion in intron 2 and give rise to a transcript longer than normal and at a lower level than the wild-type (Gibson and Wilks, 1989). A 1.4 kb insertion within the first intron of maize *Adh1* reduces mRNA levels in the mutant allele (Bennetzen *et al*, 1984). The *white-zeste-mottled* (w^{zm}) mutation is associated with a 6 kb insertion of a transposable element in the first large intron of the *white* gene (Zachar and Bingham, 1982). This insertion creates a sequence well matched to the consensus sequence of donor splice sites. The *white-ivory* (w^{i}) mutation is caused by a 2.95 kb internal duplication of *w* locus sequences (Karess and Rubin, 1982). It is likely the insertion disrupts the normal splicing pattern of the *white* transcripts.

Voelker *et al* (1990) explained how some of the insertions within the large 5' intron cause the mutant phenotypes at the *suppressor of sable* locus. Six of the mobile element insertions in the *Drosophila suppressor of sable* [*su(s)*] locus occur within a region 34 nt downstream from the GT donor splice signal. They might interfere with the pairing of the 5' end of the intron with the U1 RNA in the spliceosome complex (Zapp and Berget, 1989; Freyermuth *et al*, 1990; Konforti *et al*, 1993), thereby inhibiting splicing of the primary transcript and causing a reduced amount of wild-type message. A P element insertion and five other non-P element insertions that occur in the middle of the intron disrupt the potential formation of a prominent stem-loop structure (Voelker *et al*, 1990). How this kind of mechanism might apply to the *Tirant* element insertion, which is located in the middle of the first intron of the *Adh* gene, is

not known. A search for consensus sequences that might give rise to stem-loop structures or other structural motifs within the *Adh* first intron would be useful.

The insertion site of the *Tirant* transposable element in *AAS44* is 266 bp upstream of the proximal promoter — a region containing a *cis*- regulatory element which is important for the tissue- and temporal- specificity of *Adh* expression in larvae (see 5.2.3; Corbin and Maniatis, 1990). Deletion of these sequences results in the reduction of ADH activity in the fat body and midgut, and elimination of ADH activity in the Malpighian tubules. Whether the *Tirant* insertion affects the tissue specificity and temporal specificity of *Adh* expression in *AAS44* larvae is worth investigating, although it does not affect the overall level of *Adh* expression in larvae.

A further possible mechanism for the effect of the *Tirant* insertion in *AAS44* is that the insertion might cause premature termination of *Adh* transcription in the *Tirant* element. Such a mechanism has previously been reported where insertions are located in introns. The *white-apricot* (w^a) allele results from a *copia* element insertion in the second intron of the *white* gene (O'Hare *et al*, 1984). The mutant phenotype caused by the *copia* insertion is due to the aberrant termination of transcription from the *white* gene into the long terminal repeats of the *copia* element (Zachar *et al*, 1985). The majority of the w^a transcripts are polyadenylated within the *copia* sequences, but about 5% of the w^a primary transcript is spliced normally to generate wild-type *white* mRNA. When a P element is inserted in the same orientation as the *yellow* gene in the $y^{1\#7}$ allele, the transcription of the *yellow* gene is terminated within the insertion and the P element sequences could not be spliced, therefore these flies show a null y^- phenotype (Geyer *et al*, 1988c; 1991).

Sequences at the 5' end of the *Tirant* element described by Garrell and Modolell (1990) show that there is a polyadenylation signal (AATAAA) about 100 bp downstream of the 5' termini of *Tirant* (Appendix IV). In the *emc* allele *Achaetous*, a *Tirant* element integrated into the coding region leads to a short

protein terminated at this poly(A) signal. The *Tirant* element is inserted in the same orientation as the *Adh* transcript, and the poly(A) signal in *Tirant* is positioned downstream of the *Adh* distal promoter. Hence it is possible that this poly(A) signal might be functional in *Adh* transcription initiated from the distal promoter, thereby interfering with the production of the wild-type transcripts.

Recently, Thomas *et al* (1995) described a spontaneous dominant mutation at the *Serrate* locus (Ser^D) which is caused by a 505 bp *Tirant* insertion. The *Tirant* insertion is located within the 3' untranslated region and leads to a truncated but stable transcript. The hexanucleotide AAUAAA possibly serves as a polyadenylation signal, and this truncation of the Ser RNA within *Tirant* eliminates the putative RNA degradation signals downstream.

If this poly(A) signal does function in *AAS44*, the new transcripts will be about 500 bp, which would be expected to be detected in the Northern analysis. However, I didn't observe any different sized *Adh* transcripts in *AAS44* adults. There are two possible explanations. First, if the new transcripts are spliced normally, there will be only a very small transcript of about 100 bp, which might be trimmed off from the gel. Secondly, the new transcript could be unstable, because the 3' end of the new transcript is not highly conserved and this might affect the stability of this transcript (Jackson and Standart, 1990). Whether this polyadenylation signal is used as a terminator needs further investigation, as transcription termination in eukaryotes also depends on specific DNA sequences downstream (Birnstiel *et al*, 1985). Particularly in the *copia* element, polyadenylation requires distant upstream sequences (Kurkulos *et al*, 1991).

The *Tirant* insertion in the first intron of the *Adh* in *AAS44* might also affect *Adh* expression due to transcriptional interference from *Tirant* itself. Similar mechanisms have been postulated for the *gypsy* insertion in the *yellow* gene and in the *forked* locus (Parkhurst and Corces, 1986a; 1986b). The *gypsy*

element is activated in a particular development stage to decrease the amount of *forked*-encoded RNAs in the f^1 mutant, which results from a *gypsy* element insertion into an intron of the *forked* locus. The f^1 mutant phenotype can be reversed by a mutation at the *suppressor of forked* locus, mediated via an effect on the transcription of the *gypsy* element (Parkhurst and Corces, 1985). A similar situation occurred at the *suppressor of Hairy-wing* [*su(Hw)*] locus at which mutations can reverse the phenotype of the y^2 allele caused by the *gypsy* insertion at the *yellow* gene (Parkhurst and Corces, 1986a). Whether the *Tirant* insertion in *AAS44* affects *Adh* expression by activating its own transcription has not been investigated. To clarify this speculation, it will be necessary to define the *Tirant* element itself to see if it is able to be transcribed. The transcription profile of the *Tirant* element during development stages could be investigated.

Other sequence motifs affecting gene expression could exist in the *Tirant* element, such as binding sites for other proteins involved in the regulation of *Adh* expression. All the hypotheses discussed above suggest that it would be worthwhile obtaining the sequence of the *Tirant* element and defining its structure. Based on the sequences described by Garrell and Modolell (1990) and Thomas *et al* (1995), the *Tirant* element could be regarded as a *Drosophila* retrotransposon, as it carries 416-bp LTRs termination with 5 bp inverted repeats. The *Tirant* element also contains a few other structural features common to *Drosophila* retrotransposons (Bingham and Zachar, 1989; Appendix IV). First, the target site of *Tirant* insertion is 4-bp which is usually duplicated upon retrotransposon insertion; second, the size of the *Tirant* is 5.7 kb fitting the general size of *Drosophila* retrotransposons which is 5 to 9 kb; third, the *Tirant* element contains 5-bp inverted terminal repeat sequences (AGTTA at 5' termini and TAACT at 3' termini). More structural features of *Drosophila* retrotransposons, such as purine-rich stretch 5' to the 3' LTR, tRNA primer-binding sites, sequence similarity to corresponding vertebrate retrovirus etc

(Bingham and Zachar, 1989), can only be defined when the complete sequence of the *Tirant* element is obtained.

The *Tirant* element insertion could also affect transcription factors binding to the proximal promoter region. The *Tirant* element insertion is 28 bp 5' to the binding site of the P2 protein (Fig 7.2), which is a transcriptional regulatory factor of the proximal promoter, although the nature of the P2 protein has not been identified (Heberlein *et al*, 1985). The *Tirant* element insertion might also alter chromatin structure to affect normal *Adh* transcription from the distal promoter.

It's interesting to find that there are three different transposable element insertions at the *Adh* gene giving rise similar phenotype. Transposable elements are widely distributed throughout the *Drosophila* genome. One *Drosophila* genome may contain many copies of a given transposable element, but each individual site on the chromosome has very low frequency of particular transposable element insertions (Charlsworth and Langley, 1989). The low frequency of transposable element insertion mutations is consistent with the idea that transposable element insertions usually reduce the fitness of their hosts. Therefore, transposable element insertion mutations are almost always selectively neutral or selected against (McDonald, 1993). P elements mainly induce lethal and severely detrimental mutations, and also induce viability-reducing mutations at a high rate (Yukuhiro *et al*, 1985). *Drosophila hobo* elements have recently been shown to be responsible for chromosomal rearrangement mutations (Ho *et al*, 1993; Sheen *et al*, 1993).

Transposable element insertions are rarely found in the transcriptional unit of a particular gene in *Drosophila* natural populations. In a survey of transposable element distribution in *D. melanogaster* population, there was only one transposable element mapped within the intron of the *Ddc* gene with no associated phenotype (Charlesworth and Langley, 1989). Since the large insertions within or very near transcripts unit are usually deleterious, they are

rapidly eliminated from the population. Natural occurring insertions outside the transcription unit have rarely been associated with reduced gene expression (Aquadro *et al*, 1986). A large number of insertions have been effectively confined to a particular region flanking the transcription unit, and have little or no functional significance (Langley *et al*, 1982; Langley and Aquadro, 1987; Aguade *et al*, 1989; Eanes *et al*, 1989).

The three transposable element insertions at the *Adh* locus are all located within the *Adh* transcription unit. The three transposable element insertions at the *Adh* locus produce deleterious effects on ADH activity in adults, however, they have much less harmful effects on ADH activity in larvae. ADH activity levels are associated with alcohol tolerance and utilisation (Gibson and Oakeshott, 1982; Van Delden, 1982), but it has been argued that ADH activity in larvae is more important to viability and fitness than in adults (Heinstra *et al*, 1989). This might explain why alleles producing low activity in adults, but near normal levels in larvae, are not immediately eliminated in natural populations.

The three transposable element insertions at the *Adh* locus have been found only in alleles extracted from natural populations, not in laboratory experimental strains. The selection methods used in laboratory experiments to obtain P element mutations at the *Adh* locus would not preserve the "leaky" mutants like these three I have analysed, which retain between 11% and 34% of normal level of activity in adult flies. Therefore, the data presented in this thesis underline the importance of studying naturally occurring variation to gain insights into the control of gene expression.

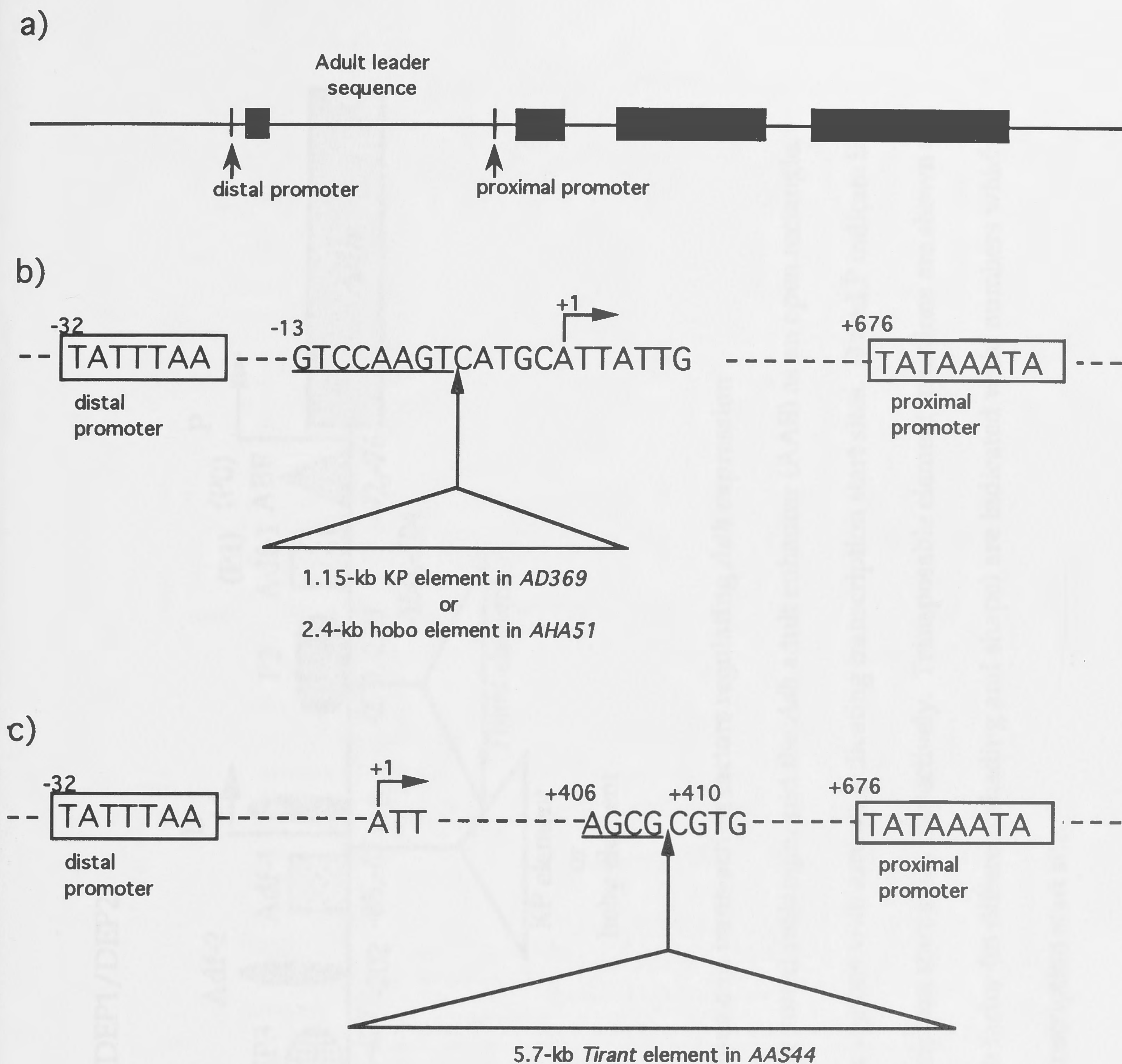


Figure 7.1 The insertion sites in three variants

The structure of the *Adh* gene is shown at a) with black rectangles representing exons and thin lines representing introns. The insertion sites in *AD369* and *AHA51* are shown at b) and that in *AAS44* is shown at c). The transposable element insertions are shown as open trangles and the sequences at the insertion sites are indicated. The two TATA boxes are boxed and the duplicated target sites are underlined. Nucleotide numbering follows Kreitman (1983) with the distal transcription start site at +1.

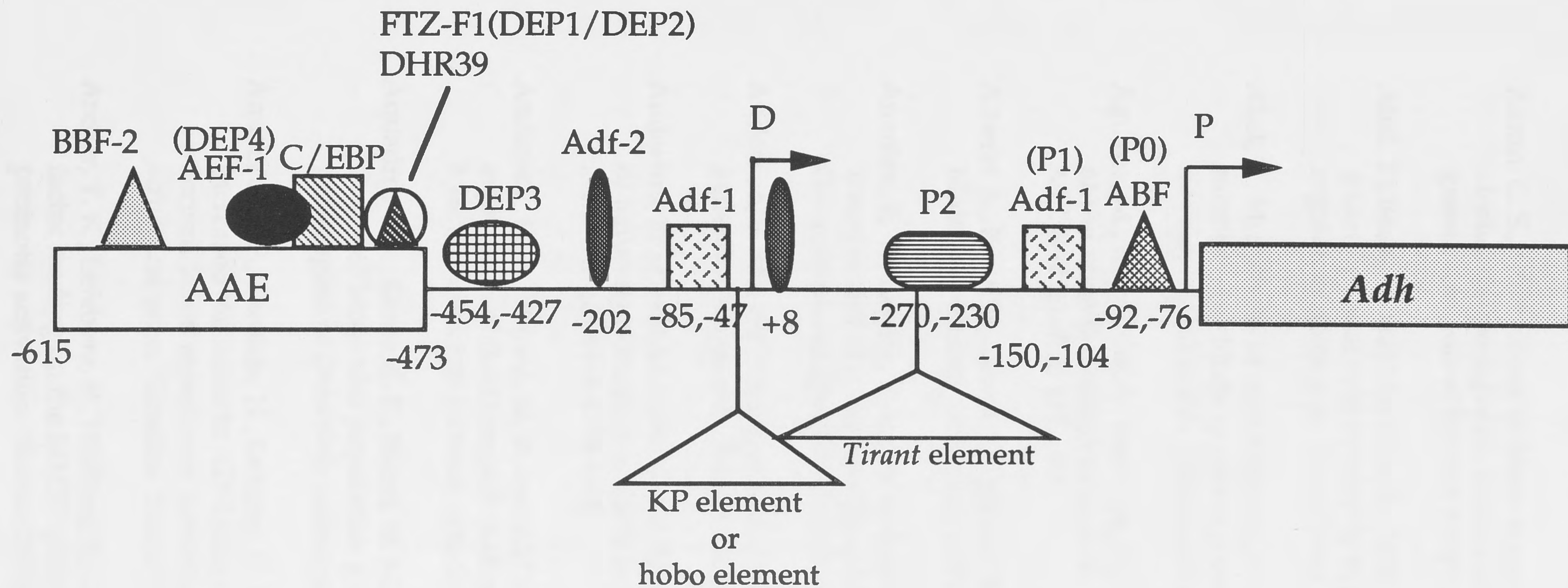


Figure 7.2 Relationship between insertions and trans-acting factors regulating *Adh* expression

The *Adh* gene coding region is shown as a dotted rectangle, and the *Adh* adult enhancer (AAE) as an open rectangle. Lines represent promoters and intergenic regions with arrows indicating transcription start sites. D and P indicate the *D. melaogaster* distal and proximal transcription start sites, respectively. Transposable element insertions are shown as open trangles. The binding sites for each factor (in different shading and shape) are indicated with numbers which refer to distances from the approriate transcription start site.

Bibliography

Aaron, C. S., (1979) X-ray induced mutations affection the level of the enzyme alcohol dehydrogenase in *Drosophila melanogaster*: frequency and genetic analysis of the null enzyme mutants *Mutat. Res.* **63**: 127-137

Abel, T., Bhatt, R. and Maniatis, T. (1992) A *Drosophila* CREB/ATF transcriptional activator binds to both fat body- and liver-specific regulatory elements. *Genes Devel.* **6**: 466-480

Abel, T., Michelson, M. and Maniatis, T. (1993) A *Drosophila* GATA family member that binds to *Adh* regulatory sequences is expressed in the developing fat body. *Development* **119**: 623-633

Aguade, M., Miyashita, N. and Langley, C. H. (1989) Restriction-map variation at the zeste-tko region in natural populations of *Drosophila melanogaster*. *Mol. Biol. Evol.* **6**: 123-130

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., and Watson, J. D. (1994) Molecular biology of the cell (third edition). *Garland Publishing, Inc.*

Ananiev, E. V., Barsky, V. E., Yu V. Ilyin and Ryzic, M. V. (1984) The arrangement of transposable elements in the polytene chromosomes of *Drosophila melanogaster*. *Chromosoma* **90**: 366-377

Anderson, S. M. (1981) Shotgun DNA sequencing using cloned DNase I generated fragments. *Nucleic Acids Res.* **9(13)**, 3015-3027

Anderson, S. M. and McDonald, J. F, (1983) Biochemical and molecular analysis of naturally occurring variants in *Drosophila melanogaster Proc. Natl. Acad. Sci. USA* **80: 4798-4802**

Anderson, S. M., Brown, M. R. and McDonald, J. F. (1991) Tissue specific expression of the *Drosophila Adh* gene: a comparison of *in situ* hybridisation and immunocytochemistry. *Genetica* **84:** 95-100

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

Aquadro, C. F., Tachida, H., Langley, C. H., Harada, K. and Mukai, T. (1990) Increased variation in ADH enzyme activity in *Drosophila* mutation-accumulation experiment is not due to transposable elements at the *Adh* structural gene. *Genetics.* **126(4)**: 915-9

Archer, T. K., Lefebvre, P., Wolford, R. G. and Hager, G. L. (1992) Transcription factor loading on the MMTV promoter: a bimodal mechanism for promoter activation. *Science* **255**: 1573-1576

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

Atkinson, P. W., Warren, W. D. and O'Brochta D. A. (1993) The *hobo* transposable element of *Drosophila* can be cross-mobilised in houseflies and excises like the *Ac* element of maize. *Proc. Natl. Acad. Sci. USA* **90:** 9693-9697

Ayer, S. and Benyajati, C. (1990) Conserved enhancer and silencer elements responsible for differential *Adh* transcription in *Drosophila* cell lines. *Mol. Cell. Biol.* **10(7)** 3512-3523

Ayer, S. and Benyajati, C. (1992) The binding site of a steroid hormone receptor-like protein within the *Drosophila Adh* adult enhancer is required for high levels of tissue-specific alcohol dehydrogenase expression. *Mol. Cell. Biol.* **12(2)** 661-673

Ayer, S., Walker, N., Mosammaparast, M., Nelson, J. P., Shilo, B. and Benyajati, C. (1993) Activation and repression of *Drosophila* alcohol dehydrogenase distal transcription by two steroid hormone receptor superfamily members binding to a common response element. *Nucleic Acids Res.* **21(7)** 1619-1627

Batterham, P., Lovett, J. A., Starmer, W. T. and Sullivan, D. T. (1983) Differential regulation of duplicate alcohol dehydrogenase genes in *Drosophila mojavensis*. *Dev. Biol.* **96(2)**: 346-54

Batterham, P., Chambers, G. K., Starmer, W. T. and Sullivan, D. T. (1984) Origin and expression of an alcohol dehydrogenase gene duplication in the genus *Drosophila*. *Evolution* **38**: 644-57

Bennetzen, J. L., Swanson J., Taylor, W. C. and Freeling M. (1984) DNA insertion in the first intron of maize *Adh1* affects message levels: cloning of progenitor and mutant *Adh1* alleles. *Proc. Natl. Acad. Sci. USA.* **81**: 4125-4128

Benyajati, C., Wang, N., Reddy, A., Weinberg, E. and Sofer, W. (1980) Alcohol dehydrogenase in *Drosophila*: isolation and characterisation of messenger RNA and cDNA clone. *Nucleic Acids Res.* **8:** 5649-5667

Benyajati, C., Place, A. R., Powers, D. A. and Sofer, W. (1981) Alcohol dehydrogenase gene of *Drosophila melanogaster*: Relationship of intervening sequences to functional domains in the protein. *Proc. Natl. Acad. Sci. USA* **78**: 2717-2721

Benyajati, C., Place, A. R., Wang, N., Pents, E. and Sofer, W. (1982) Deletions at intervening sequence splice sites in the alcohol dehydrogenase gene of *Drosophila*. *Nucleic Acids Res.* **10**: 7261

Benyajati, C., Spoerel, N., Haymerle, H. and Ashburner, M. (1983) The messenger RNA for alcohol dehydrogenase in *Drosophila melanogaster* differs in its 5' end in different developmental stages. *Cell* **33**: 125-133

Benyajati, C., Ayer, S., McKeon, J., Ewel, A. and Huang J. (1987) Roles of cis-acting elements and chromatin structure in *Drosophila* alcohol dehydrogenase gene expression. *Nucleic Acids Res.* **15**: 7903-7920

Benyajati, C., Ewel, A., McKeon, J., Chovav, M. and Juan, E. (1992) Characterisation and purification of *Adh* distal promoter factor 2, Adf-2, a cell-specific and promoter-specific repressor in *Drosophila. Nucleic Acids Res.* **20(17)**: 4481-4489

Berg, D. E. and Howe, M. M. (1989) *Mobile DNA* Am. Soc. Microbiol., Washington, DC.

Biemont, C. (1992) Population genetics of transposable DNA elements, *A* Drosophila *point of view. Genetica* **86**: 67-84

Bingham, P. M., Levis, R. and Rubin, G. M. (1981) Cloning of DNA sequences from the *white* locus of *D. melanogaster* by a novel and general method. *Cell* **25**: 693-704

Bingham, P. M. and Zachar, Z. (1989) Retrotransposons and the FB transposon from *Drosophila melanogaster.* In "*Mobile DNA*" (D. E. Berg and M. M. Howe, eds.), pp. 485-502. Am. Sco. Microbiol., Washington, D. C.

Birnstiel, M. L., Busslinger, M. and Strub, K. (1985) Transcription termination and 3' processing: the end is in site! *Cell* **41**: 349-359

Black, D. M., Jackson, M. S., Kidwell, M. G. and Dover, G. A. (1987) KP elements repress P-induced hybrid dysgenesis in *Drosophila melanogaster*. *EMBO J.* **6**: 4125-4135

Blackman, R. K., Grimaila, R., Koehler, M. M. D. and Gelbart, W. M. (1987) Mobilisation of hobo elements residing within the *decapentaplegic* gene complex: suggestion of a new hybrid dysgenesis system in *Drosophila melanogaster*. *Cell* **49**, 497-505

Blackman, R. K. and Gelbart, W. M. (1989) The transposable element hobo of *Drosophila melanogaster.* In "*Mobile DNA*" (D. E. Berg and M. M. Howe, eds.), pp. 523-529. Am. Sco. Microbiol., Washington, D. C.

Blackman, R. K., Koehler, M. M. D., Grimaila, R. and Gelbart, W. M. (1989) Identification of a fully-functional hobo transposable element and its use for germ-line transformation of *Drosophila. EMBO J.* **8**: 211-217

Boeke, J. D. (1989) Transposable elements in *Saccharomyces cerevisiae.* In "*Mobile DNA*" (D. E. Berg and M. M. Howe, eds.), pp. 335-374. Am. Sco. Microbiol., Washington, D. C.

Boussy, I. A., Healy, M. J., Oakeshott, J. G. and Kidwell, M. G. (1988) Molecular analysis of the P-M gonadal dysgenesis cline in eastern Australian *Drosophila melanogaster Genetics* **119**: 889-902

Brawerman, G. (1987) Determinants of Messenger RNA Stability. *Cell* **48**: 5-6

Bregliano, J. C. and Kidwell, M. G. (1983) Hybrid dysgenesis determinants. In: *Mobile Genetic Elements.* (ed: J. Shapiro). Academic Press, New York.

Brehme, K. S. and Demerec, M. (1942) A survey of Malpighian tubule colour in the eye colour mutants of *Drosophila melanogaster. Growth* **6**: 351-355

Buratowski, S., Hahn, S., Guarente, L. and Sharp, P. A. (1989) Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* **56**: 549-561

Buratowski, S. (1994) The basics of basal transcription by RNA polymerase II. *Cell* **77**: 1-3.

Chambers, G. K., Laver, W. G., Campbell, S. and Gibson, J. B. (1981) Structural analysis of an electrophoretically cryptic alcohol dehydrogenase variant from an Australian population of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **78(5)**: 3103-7

Chambers, G. K. (1991) Gene expression, adaptation and evolution in higher organisms: evidence from studies of *Drosophila* alcohol dehydrogenases. *Comp. Biochem. Physilo.* **99B**: 723-730

Charlesworth, B. and Langley, C. H. (1989) The population genetics of *Drosophila* transposable elements. *Annu. Rev. Genet.* **23**: 251-287

Charlesworth, B., Lapid, A. and Canada, D. (1992) The distribution of transposable elements within and between chromosomes in a population of *Drosophila melanogaster.* II. Inferences on the nature of selection against elements. *Genet. Res.* **60**: 115-30

Chia, W., Savakis, C., Karp, R., Pelham, H. and Ashburner, M (1985a) Mutation of the *Adh* gene of *Drosophila melanogaster* containing an internal tandem duplication. *J. Mol. Biol.* **186**: 679-688

Chia, W., Karp, R., McGill, S. and Ashburner M. (1985b) Molecular analysis of the *Adh* region of the genome of *Drosophila melanogaster. J. Mol. Biol.* **186**: 689-706

Chia, W. G., Howes, M., Martin, Y. B., Meng, K., Moses, K. and Tsubota, S. (1986) Molecular analysis of the *yellow* locus of *Drosophila. EMBO J.* **5**: 3597-3605

Chia, W., Savakis, C., Karp, R. and Ashburner, M. (1987) *Adh*^{n4} of *Drosophila melanogaster* is a nonsense mutation. *Nucleic Acids Res.* **15(9)**: 3931

Choudhary, M. and Laurie, C. C. (1991) Use of in *vitro* mutagenesis to analyse the molecular basis of the difference in Adh expression associated with the allozyme polymorphism in *Drosophila melanogaster.* **129**: 481-488

Clavi, B. R. and Gelbart, W. M. (1994) The basis for germline specificity of the hobo transposable element in *Drosophila melanogaster. EMBO J.* **13**: 1636-1644

Collet, C. (1988) Recent origin for a thermostable alcohol dehydrogenase allele of *Drosophila melanogaster. J. Mol. Evol.* **27**: 142-146

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**: 71-9

Corbin, V. and Maniatis, T. (1989a) Role of transcriptional interference in the *Drosophila melanogaster Adh* promoter switch. *Nature* **337**: 279-282

Corbin, V. and Maniatis, T. (1989b) The role of specific enhancer-promoter interactions in the *Drosophila Adh* promoter switch. *Genes Devl.* **3**: 2191-2200

Corbin, V. and Maniatis, T. (1990) Identification of *cis*-regulatory elements required for larval expression of the *Drosophila melanogaster* alcohol dehydrogenase gene. *Genetics* **124**: 637-646

Cross, S. R. H. and Birley, A. J. (1986) Restriction endonuclease map variation in the *Adh* region in populations of *Drosophila melanogaster. Biochemical Genetics* **24**: 415-433

Croston, G. E., Lira, L. M. and Kadonaga, J. T. (1991) A general method for purification of H1 histones that are active for repression of basal RNA polymerase II transcription. *Protein Expr. Purif.* **2**: 162-169

David, J. R., Bocquet, C., Arens, M. F. and Fouillet, P. (1976) Biological role of alcohol dehydrogenase in the tolerance of *Drosophila melanogaster* to aliphatic alcohols: utilisation of an ADH-null mutant. *Biochem. Genet.* **14(11-12)**: 989-97

Denhardt, D. T. (1966) A membrane filter technique for the detection of complementing DNA. *Biochem. Bioghys. Res. Comm.* **23**: 641-652

Dierks, P., Ooyen, A., Cochran, M. D., Dobkin, C., Reiser, J. and Weissmann, C. (1983) Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β-globin gene in mouse 3T6 cells. *Cell* **32**: 695-706

Di Franco, C., Galuppi, D. and Junakovic, N. (1992a) Genomic distribution of transposable elements among individuals of an inbred *Drosophila* line. *Genetica* **86**: 1-11

Di Franco, C., Pisano, C., Fourcade-Peronnet, F., Echalier, G. and Junakovic, N. (1992b) Evidence for *de novo* rearrangements of *Drosophila* transposable elements induced by the passage to the cell culture. *Genetica* **87**: 65-73

Dunn, G. R., Wilson, T. G. and Jacobson, K. B. (1969) Age-dependent changes in alcohol dehydrogenase in *Drosophila*. *J. Exp. Zool.* **171**: 185-189

Dunn, R. C. and Laurie, C. C. (1995) Effects of a transposable element insertion on alcohol dehydrogenase expression in *Drosophila melanogaster*. *Genetics*. **140**: 667-677

Eanes, W. F., Wesley, C., Hey, J. and Houle, D. (1988) The fitness consequences of P element insertion in *Drosophila melanogaster*. *Genet. Res.* **52**: 17-26

Eanes, W. F., Ajioka, J. W., Hey, J. and Wesley, C. (1989a) Restriction-map variation associated with the G6PD polymorphism in natural populations of *Drosophila melanogaster*. *Mol. Biol. Evol.* **6**:384-397

Eanes, W. F., Labate, J. and Ajioka, J. W. (1989b) Restriction-map variation with the yellow-achaete-scute region in five populations of *Drosophila melanogaster*. *Mol. Biol. Evol.* **6**:492-502

Eissenberg, J. C. and Elgin, S. C. R. (1987) *Hsp*^{28stl}: a P element insertion mutation that alters the expression of the heat shock gene in *Drosophila melanogaster*. *Genetics* **115**: 333-340

Eisses, K. T. (1989) On the oxidation of aldehydes by alcohol dehydrogenase of *Drosophila melanogaster*: evidence for the gem-diol as the reacting substrate. *Biorganic Chem.* **17**: 268-274

Engels, W. R. (1985) Guidelines for P-element transposon tagging. *Drosophila Inform. Serv.* **61**: 1

Engels, W. R. (1989) P element in *Drosophila melanogaster*. In "*Mobile DNA*" (D. E. Berg and M. M. Howe, eds.), pp. 437-484. Am. Sco. Microbiol., Washington, D. C.

Engels, W. R., Johnson-Schlitz, D. M., Eggleston, W. B. and Sved, J. (1990) High -frequency P element loss in *Drosophila* is homology dependent. *Cell* **62**, 515-525

England, B. P., Heberlein, U. and Tjian, R. (1990) Purified *Drosophila* transcription factor, *Adh* distal factor-1 (Adf-1), binds to sites in several *Drosophila* promoters and activates transcription. *J. Biol. Chem.* **265(9)**: 5086-5094

England, B. P., Admon, A. and Tjian, R. (1992) Cloning of *Drosophila* t transcription factor Adf-1 reveals homology to Myb oncoproteins. *Proc. Natl, Acad. Sci. USA* **89**: 683-687

Erilich, H. A. (1989) *PCR technology: principles and applications for DNA amplification.* Stockton Press

Ewel, A., Jackson, J. R. and Benyajati, C. (1990) Alternative DNA-protein interactions in variable-length internucleosomal regions associated with *Drosophila Adh* distal promoter expression. *Nucleic Acids Res.* **18(7)**: 1771-1781

Falb D. and Maniatis T. (1992a) A conserved regulatory unit implicated in tissue-specific gene expression in *Drosophila* and man. *Genes and Dev.* **6**: 454-465

Falb D. and Maniatis T. (1992b) *Drosophila* transcriptional repressor protein that binds specifically to negative control elements in fat body enhancers. *Mol. Cell. Biol.* **9:** 4093-4103

Feaver, W. J. and Pearlman, R. E. (1991) Chromatin structure in a region of a yeast transposable element regulating adjacent gene expression. *Biochem. Cell Biol.* **69**: 392-398

Felsenfeld, G. (1992) Chromatin as an essential part of the transcriptional mechanism. *Nature.* **355**(6357): 219-24

Finnegan, D. J. and Fawcett, D. H. (1986) Transposable elements in *Drosophila melanogaster. Oxf. Surv. Eukaryotic Genes* **3**: 1-62

Finnegan, D. J. (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet.* **5**: 103-107

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

Fossett, N. G., Arbour-Reily, P., Kilroy, G., McDaniel, M., Mahmoud, J., Tucker, A. B., Chang, S. H. and Lee, W. R. (1990) Analysis of ENU-induced mutations at the *Adh* locus in *Drosophila melanogaster. Mutat. Res.* **231(1)**: 73-85

Freeth, A. L. and Gibson, J. B. (1985) Alcohol dehydrogenase and *sn*-glycerol-3-phosphate dehydrogenase null activity alleles in natural populations of *Drosophila melanogaster*.. *Heredity* **55**: 369-374

Freeth, A. L., Gibson, J. B. and Wilks, A. V. (1990) Aberrant splicing of a naturally occurring alcohol dehydrogenase null activity allele in *Drosophila melanogaster*.. *Genome*. **33(6)**: 873-877

Freyermuth, C., Query, C. C. and Keene, J. D. (1990) Quantitative determination that one of two potential RNA-binding domains of the A protein component of the U1 small nuclear ribonucleoprotein complex binds with high affinity to stem-loop II of U1 RNA. *Proc. Natl. Acad. Sci. USA* **87**: 6393-6397

Garrell, J. and Modolell, J. (1990) The *Drosophila* extramacrochaetae locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop-helix protein. *Cell* **61**: 39-48

Georgel, P., Dretzen, G., Jagla, K., Bellard, F., Dubrovsky, E., Calco, V. and Bellard, M. (1993) GEBF-I activates the *Drosophila* Sgs3 gene enhancer by altering a positioned nucleosomal core particle. *J. Mol. Biol.* **234**: 319-330

Geyer, P. K., Spana, C. and Corces, V. G. (1986) On the molecular mechanism of gypsy-induced mutations at the *yellow* locus of *Drosophila melanogaster*. *EMBO J* **5**: 2657-2662

Geyer, P. K., Green, M. M. and Corces, V. G. (1988a) Reversion of a gypsy-induced mutation at the *yellow (y)* locus of *Drosophila melanogaster* is associated with the insertion of a newly defined transposable element. *Proc. Natl. Acad. Sci. USA* **85**: 3938-3942

Geyer, P. K., Green, M. M. and Corces, V. G. (1988b) Mutant gene phenotypes mediated by a *Drosophila melanogaster* retrotransposon require sequences homologous to mammalian enhancers. *Proc. Natl. Acad. Sci. USA* **85**: 8593-8597

Geyer, P. K., Richardson, K. L., Corces, V. G., and Green, M. M. (1988c) Genetic instability in *Drosophila melanogaster*: P-element mutagenesis by gene conversion. *Proc. Natl. Acad. Sci. USA* **85**: 6455-6459

Geyer, P. K., Green, M. M. and Corces, V. G. (1990) Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvertion in *Drosophila*. *EMBO J* **9**: 2247-2256

Geyer, P. K., Chien, A. J., Corces, V. G. and Green, M. M. (1991) Mutations in the *su(s)* gene affect RNA processing in *Drosophilae melanogaster*. *Proc. Natl. Acad. Sci. USA* **88**: 7116-7120

Gibson, J. B. and Oakeshott, J. G. (1982) Tests of the adaptive significance of the alcohol dehydrogenase polymorphism in *Drosophila melanogaster*: paths, pitfalls and prospects. In Barker, J. S. F. and Starmer, W. T. (eds.) *Ecological Genetics and Evolution. The Cactus-Yeast-Drosophila Model System*, Academic Press, New York, pp. 291-306

Gibson, J. B. and Wilks, A. V. (1989) Molecular structure of a naturally occurring alcohol dehydrogenase null activity allele in *Drosophila melanogaster. Biochem. Genet.* **27(11-12)**: 679-88

Gibson, J. B., Cao, A., Symonds, J. and Reed, D. (1991) Low activity *sn*-glycerol-3-phosphate dehydrogenase variants in natural populations of *Drosophila melanogaster. Heredity* **66**: 75-82

Gibson, J. B., Wilks, A. V. and Agrotis A. (1992) Molecular relationships between alcohol dehydrogenase null-activity alleles from natural populations of *Drosophila melanogaster. Mol. Biol. Evol.* **9(2)**:250-260

Gill, G., Pascal, E., Tseng, Z. H. and Tjian, R. (1994) A glutamine-rich hydrophobic patch in transcription factor Sp1 contacts the dTAF$_{II}$110 component of the *Drosophila* TFIID complex and mediates transcriptional activation. *Proc. Natl. Acad. Sci. USA* **91**: 192-196

Goldberg, D. A. (1980) Isolation and partial characterisation of the *Drosophila* alcohol dehydrogenase gene *Proc. Natl. Acad. Sci. USA* **77**: 5794-5798

Goldberg, D. A., Posakony, J. W. and Maniatis, T. (1983) Correct developmental expression of a cloned alcohol dehydrogenase gene transduced into the *Drosophila* germ line. *Cell* **34**: 59-73

Goodrich, J. A., Hoey, T., Thut, C. J., Admon, A. and Tjian, R. (1993) *Drosophila* TAFII40 interacts with both a VP16 activation domain and the basal transcription factor TFIIB. *Cell* **75**: 519-530

Goodrich, J. A. and Tjian, R. (1994) Transcription factors IIE and IIH and ATP hydrolysis direct promoter clearance by RNA polymerase II. *Cell* **77**: 145-156

Green, M. M. (1977) Genetic instability in *Drosophila melanogaster:* de novo induction of putative insertion mutations. *Proc, Natl. acad. Sci. USA* **74**: 3490-3493

Green, M. M. (1980) Transposable elements in *Drosophila* and other Diptera. *Annu. Rev. Genet.* **14**: 109-120

Grell, E. H., Jacobson, K. B., and Murphy, J. B. (1965) Alcohol dehydrogenase in *Drosophila melanogaster*: isozymes and genetic variants. *Science* **149**: 80-82

Grell, E. H., Jacobson, K. B. and Murphy, J. B. (1968) Alterations of genetic material for analysis of alcohol dehydrogenase isozymes of *Drosophila melanogaster*. *Ann. N. Y. Acad. Sci.* **151:** 441-55

Grosschedl, R. and Birnstiel, M. L. (1980) Spacer DNA sequences upstream of the T-A-T-A-A-A-T-A sequence are essential for promotion of H2A histone gene transcription in vivo. *Proc. Natl. Acad. Sci. U.S.A.* **77(12)**: 7102-7106

Hahn, S., Buratowski, S., Sharp, P. A. and Guarente, L. (1989) Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and non-consensus DNA sequences. *Proc. Natl. Acad. Sci. U.S.A.* **86**: 5718-5722

Hansen, S. K. and Tjian, R. (1995) TAFs and TFIIA mediate differential utilization of the tandem *Adh* promoters. *Cell* **82**: 565-575

Hatzopoullos, P., Monastirioti, M., Yannopoulos, G. and Louis, C. (1987) The instability of the TE-like mutation Dp(2;2) GYL of *Drosophila melanogaster* is intimately associated with the hobo element. *EMBO J.* **6**: 3091-3096

Hausner, W., Frey, G. and Thomm, M. (1991) Control regions of an archaeal gene a TATA box and an initiator element promote cell-free transcription of the $tRNA^{val}$ gene of *Methanococcus vannielii*. *J. Mol. Biol.* **222**: 495-508

Haymerle, H. (1983) The alcohol dehydrogenase gene of *Drosophila melanogaster*: DNA sequence, *in vivo* and *in vitro* transcripts. Thesis, University of Cambridge, Cambridge.

Heard, D. J. (1995) Analysis of sequences and factors involved in small RNA gene transcription in plants.

Heberlein, U., England, B. and Tjian, R. (1985) Characterization of *Drosophila* transcription factors that activate the tandem promoters of the alcohol dehydrogenase gene. *Cell* **41**: 965-977

Heinstra, P. W., Eisses, K. T., Scharloo, W. and Thorig, G. E. (1986) Metabolism of secondary alcohols in *Drosophila melanogaster*. Effects on alcohol dehydrogenase. *Comp. Biochem. Physiol. B.* **83:** 403-408

Heinstra, P. W., Geer, B. W., Seykens, D. and Langevin, M. (1989) The metabolism of ethanol-derived acetaldehyde by alcohol dehydrogenase (EC 1.1.1.1) and aldehyde dehydrogenase (EC 1.2.1.3) in *Drosophila melanogaster* larvae. *Biochem. J.* **259(3)**: 791-7

Henikoff, S. (1983) Cloning exons of mapping of transcription: characterisation of the *Drosophila melanogaster* alcohol dehydrogenase gene. *Nucleic Acids Res.* **11(14)**: 4735-52

Hipeau-Jacquotte, R., Brutlag, D. L. and Bregegere, F. (1989) Conversion and reciprocal exchange between tandem repeats in *Drosophila melanogaster Mol. Gen. Genet.* **220**: 140-146

Ho, Y. T., Weber, S. M. and Lim, J. K. (1993) Interacting hobo transposons in an inbred strain and interaction regulation in hybrids of *Drosophila melanogaster*. *Genetics* **134:** 895-908

Innis, M. a., Myambo, K. B., Gelfand, D. H. and Brow, M. A. D. (1988) DNA sequencing with *Thermus aquaticus* DNA polymerase and direct sequencing of polymerase chain reaction-amplified DNA. *Proc. Natl. Acad. Sci. USA* **85**: 9436-9440

Jackson, J. R. and Benyajati, C. (1992) *In vivo* stage- and tissue-specific DNA-protein interactions at the *D. melanogaster* alcohol dehydrogenase distal promoter and adult enhancer. *Nucleic Acids Res.* **20(20)**: 5413-5422

Jackson, J. R. and Benyajati, C. (1993) DNA-histone interactions are sufficient to position a single nucleosome juxtaposing *Drosophila Adh* adult enhancer and distal promoter. *Nucleic Acids Res.* **21(4):** 957-967

Jackson, R. J. and Standart, N. (1990) Do the poly(A) tail and 3' untranslated region control mRNA translation? *Cell* **62**:15-24

Jacobson, J. W., Medhora, M. M. and Hartl, D. L. (1986) Molecular structure of a somatically unstable transposable element in *Drosophila. Proc. Natl. Acad. Sci. USA* **83**: 8684-8688

Jiang, C. and Gibson, J. B. (1992a) The alcohol dehydrogenase polymorphism in natural populations of *Drosophila melanogaster*: restriction map variation in the region of the *Adh* locus in populations from two hemispheres *Heredity* **68**: 1-14

Jiang, C. and Gibson, J. B. (1992b) The alcohol dehydrogenase polymorphism in natural populations of *Drosophila melanogaster*: ADH activity variation restriction site polymorphism and the *Adh* cline. *Heredity* **68**: 337-344

Johnson, F. M. and Schaffer, H. E. (1973) Isozyme variability in species of the genus *Drosophila*. VII. Genotype-environment relationships in populations of *D. melanogaster* from the Eastern United States. *Biochem. Genet.* **10(2):** 149-63

Johnson, P. F. and McKnight, S. L. (1989) Eukaryotic transcriptional regulatory proteins. *Annu. Rev. Biochem.* **58**: 799-839

Karess, R. E. and Rubin, G. M. (1982) A small tandem duplication is responsible for the unstable *white-ivory* mutation in *Drosophila. Cell* **36**: 469-481

Karess, R. E. and Rubin, G. M. (1984) Analysis of P transposable element functions in *Drosophila*. *Cell* **38**: 135-146

Kelley, M. R., Mims, I. P., Farnet, C. M., Dicharry, S. A. and Lee, W. R. (1985) Molecular analysis of X-ray-induced alcohol dehydrogenase (ADH) null mutations in *Drosophila melanogaster*. *Genetics* **109**: 365-377

Kelley, M. R., Kidd, S., Berg, R. L. and Young, M. W. (1987) Restriction of P element insertions at the *Notch* locus of *Drosophila melanogaster*. *Mol. Cell Biol.* **7**: 1545-1548.

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

Kidwell, M. G. (1986) P-M mutagenesis pp59-82. *In: Drosophila: A practical Approach*, Ed: D. B. Roberts. IRL Press, Oxford.

Kingston, R. E., Bunker, C. A. and Imbalzano, A. N. (1996) Repression and activation by multiprotein complexes that alter chromatin structure. *Genes & Dev.* **10**: 905-920

Konforti, B. B., Koziolkiewicz, M. J. and Konarska, M. M. (1993) Disruption of base pairing between the 5' splice site and the 5' end of U1 snRNA is required for spliceosome assembly. *Cell* **75**: 863-73

Kornberg, R. D. (1996) RNA polymerase II transcription control. *Trend Biochem. Sci.* **21**: 325-326.

Kovacs, B. J. and Butterworth, H. W. (1986) The effect of changing the distance between the TATA-box and cap site by up to three base pairs on the selection of the transcriptional start site of a cloned eukaryotic gene *in vitro* and *in vivo*. *Nucleic Acids Res.* **14(6)**: 937-944

Kozak, M. (1992) Regulation of translation in eukaryotic systems. *Annu. Rev. Cell Biol.* **8:** 197-225

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

Kreitman, M. and Hudson, R. R. (1991) Inferring the evolotionary histories of the *Adh* and *Adh-dup* loci in *Drosophila melanogaster* from patterns of polymorphism and divergence. *Genetics* **127**: 565-582

Kubli, E., Schmidt, T., Martin, P. F. and Sofer, W. (1982) In vitro suppression of a nonsense mutant of *Drosophila melanogaster*. *Nucleic Acids Res.* **10:** 7145-7152

Kurkulos, M., Weinberg, J. M., Pepling, M. E. and Mount, S. M. (1991) Polyadenylylation in *copia* requires unusually distant upstream sequences *Proc. Natl. Acad. Sci. USA* **88**: 3038-3042

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

Langley, C. H. and Aquadro, C. F. (1987) Restriction-map variation in natural populations of *Drosophila melanogaster*: white-locus region. *Mol. Biol. Evol.* **4**: 651-663

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

Laurie-Ahlberg, C. C. (1985) Genetic variation affecting the expression of enzyme-coding genes in Drosophila: an evolutionary perspective. *Isozymes Curr. Top. Biol. Med. Res.* **12**: 33-38

Laurie-Ahlberg, C. C. and Stam, L. F. (1987) Use of P-element mediated transformation to identify the molecular basis of naturally occurring variants affecting *Adh* expression in *Drosophila melanogaster. Genetics* **115:** 129-140

Laurie, C. C. and Stam, L. F. (1988) Quantitative analysis of RNA produced by slow and fast alleles of *Adh* in *Drosophila melanogaster. Proc. Natl. Acad. Sci. USA* **85**: 5161-5165

Laurie, C. C., Heath, E. M., Jacobson, J. W. and Thomson, M. S. (1990) Genetic basis of the difference in alcohol dehydrogenase expression between *Drosophila melanogaster* and *Drosophila simulans. Proc. Natl. Acad. Sci. USA* **87:** 9674-9678

Laurie, C. C., Bridgham J. T. and Choudhary, M. (1991) Associations between DNA sequence variation and variation in expression of the *Adh* gene in natural populations of *Drosophila melanogaster. Genetics* **129**: 489-499

Le, L., Ayer, S., Place, A. R. and Benyajati, C. (1990) Analysis of formaldehyde-induced *Adh* mutations in *Drosophila* by RNA structure mapping and direct sequencing of PCR-amplified genomic DNA. *Biochem. Genet.* **28**(7-8): 367-87

Levis, R. and Rubin, G. M. (1982) The unstable *wDZL* mutation of *Drosophila* is caused by a 13 kilobase insertion that is imprecisely excised in phenotypic revertants. *Cell.* **30**: 543-50

Lewis, N. and Gibson, J. B. (1978). Variation in amount of enzyme protein in natural populations. *Biochem. Genet.* **16**: 159-170

Lim, J. K. (1979) Site-specific instability in *Drosophila melanogaster*: the origin of the mutation and cytogenetic evidence for site specificity. *Genetics* **93**: 681-701

Lim, J. K. (1981) Site-specific intrachromosomal rearrangements in *Drosophila melanogaster*: cytogenetic evidence for transposable elements. *Cold Spring Harbor Symp. Quant. Biol.* **45**: 553-560

Lim, J. K. (1988) Intrachromosomal rearrangements mediated by hobo transposons in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* **85**: 9153-9157

Lindsley, D. and Grell, E. H. (1968) *Genetic variations of Drosophila melanogaster* Carnegie Institution of Washington Publication **No. 627**

Lindsley, D. L. and Zimm, G. G. (1992) *The Genome of Drosophila melanogaster*. Academic Press, Inc.

Lockett, T. J. and Ashburner, M. (1989) Temporal and spatial utilisation of the alcohol dehydrogenase gene during the development of *Drosophila melanogaster*. *Dev. Biol.* **134:** 430-437

Luscher, B., Stauber, C., Schindler, R. and Schumperli, D. (1985) Faithful cell-cycle regulation of a recombinant mouse histone H4 gene is controlled by sequences in the 3'-terminal part of the gene. *Proc. Natl. Acad. Sci. USA* **82:** 4389-4393

Maniatis, T., Fritsch, E. F. and Sambrook, J. (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor Laboratory Press.

Maroni, G., Laurie-Ahlberg, C. C., Adams, D. A. and Wilton, A. N. (1982) Genetic variation in the expression of ADH in *Drosophila melanogaster*. *Genetics* **101**: 431-446

Maroni, G. and Stamey, S. C. (1983) Developmental profile and tissue distribution of alcohol dehydrogenase. *Drosophila Information Service* **59**: 77-79

Martin, P. F., Place, A. R., Pentz, E. and Sofer, W. (1985) UGA nonsense mutation in the alcohol dehydrogenase gene of *Drosophila melanogaster*. *J. Mol. Biol.* **184**:221-229

Mazo, A. M., Mizrokhi, L. J., Karavanov, A. A., Sedkov, Y. A. and Krichevskaja, A. A. (1989) Suppression in *Drosophila*: *su(Hw)* and *su(f)* gene products interact with a region of *gypsy* (mdg4) regulating its transcriptional activity. *EMBO J.* **8:** 903-911

McDonald, J. F. (1993) Evolution and consequences of transposable elements. *Current Opinion in Genetics and Development* **3**: 855-864

McGinnis, W., Shermoen, A. W. and Beckendorf, S. K. (1983) A transposable element inserted just 5' to a *Drosophila* glue protein gene alters gene expression and chromatin structure. *Cell* **34**, 75-84

McKnight, S. L., Gavis, E. R., Kingsbury, R. and Axel, R. (1981) Analysis of transcriptional regulatory signals of the HSV thymidine kinase gene: identification of an upstream control region. *Cell.* **25(2)**: 385-98

McPherson, M. J. Quirke, P. and Taylor, G. R. (1991) *PCR: a practical approach.* Oxford University Press

Montgomery, E. A., Charlesworth, B. and Langley, C. H. (1987) A test for the role of natural selection in the stabilization of transposable element copy number in a population of *Drosophila melanogaster*. *Genet. Res.* **49**: 31-41

Morris, T., Marashi, F., Weber, L., Hickey, E., Greenspan, D., Bonner, J., Stein, J. and Stein, G. (1986) Involvement of the 5'-leader sequence in coupling the stability of a human H3 histone mRNA with DNA replication. *Proc. Natl. Acad. Sci. USA.* **83(4)**: 981-985

Mozer, B., Marlor, R. Parkhurst, S. and Corces, V. (1985) Characterisation and developmental expression of a *Drosophila ras* oncogene. *Mol. Cell. Biol.* **5**: 885-889

Nikolov, D. B. and Burley, S. K. (1994) 2.1 A resolution refined structure of a TATA box-binding protein (TBP). *Nature Struct. Biol.* **1**: 621-637

Oakeshott, J. G., Chambers, G. K., East, P. D., Gibson, J. B. and Barker, J. S. F. (1982) Evidence for a genetic duplication involving alcohol dehydrogenase genes in *Drosophila buzzatii* and related species. *Aust. J. Biol. Sci.* **35**: 73-84

O'Donnell, J., Gerace, L., Leister, F. and Sofer, W. (1975) Chemical selection of mutants that affect alcohol dehydrogenase in *Drosophila:* 2. Use of 1-pentyn-3-ol. *Genetics* **79**: 73-83

O'Donnell, J., Mandel, H. C., Krauss, M., Sofer, W. (1977) Genetic and cytogenetic analysis of the *Adh* region in *Drosophila melanogaster*. *Genetics* **86**: 553-66

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

O'Hare, K., Murphy, C. Levis, R. and Rubin, G. M. (1984) DNA sequence of the *white* locus of *Drosophila melanogaster J. Mol. Biol.* **180**: 437-455

Parkhurst, S. M. and Corces, V. G. (1985) *Forked*, gypsys, and suppressors in *Drosophila*. *Cell* **41**: 429-37

Parkhurst, S. M. and Corces, V. G. (1986a) Interactions among the gypsy transposable element and the *yellow* and the *suppressor of Hairy wing* loci in *Drosophila melanogaster*. *Mol. cell. Biol.* **6**: 47-53

Parkhurst, S. M. and Corces, V. G. (1986b) Mutations at the suppressor of forked locus increase the accumulation of gypsy-encoded transcripts in *Drosophila melanogaster Mol. cell. Biol.* **6**: 2271-2274

Pascual, L. and Periquet, G. (1991) Distribution of hobo transposable elements in natural populations of *Drosophila melanogaster*. *Mol. Biol. Evol.* **8(3)**: 282-296

Peifer, M. and Bender, W. (1988) Sequences of the gypsy transposon of *Drosophila* necessary for its effects on adjacent genes. *Proc. Natl. Acad. Sci. USA* **85**:9650-9654

Perlmann, T. and Wrange, O. (1991) Inhibition of chromatin assembly in *Xenopus* oocytes correlates with derepression of the mouse mammary tumor virus promoter. *Mol. Cell. Biol.* **11**: 5259-5265

Piechaczyk, M., Yang, J. Q., Blanchard, J. M., Jeanteur, P. and Marcu, K. B. (1985) Posttranscriptional mechanisms are responsible for accumulation of truncated c-myc RNAs in murine plasma cell tumors. *Cell* **42**: 589-597

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

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

Place, A. R., Benyajati, C. and Sofer, W. (1987) Molecular consequences of two formaldehyde-induced mutations in the alcohol dehydrogenase gene of *Drosophila melanogaster*. *Biochem. Genet.* **25**: 621

Posakony, J. W., Fischer, J. A. and Maniatis, T. (1985) Identification of DNA sequences required for the regulation of *Drosophila* alcohol dehydrogenase gene expression. *Cold Spring Harbor Symp. Quant. Biol.* **50**: 515-520

Rabbitts, P. H., Forster, A., Stinson, M. A. and Rabbitts, T. H. (1985) Truncation of exon 1 from the c-myc gene results in prolonged c-myc mRNA stability. *EMBO J.* **4**: 3727-33

Rasmusson, K. E., Raymond, J. D. and Simmons, M. J. (1993) Repression of hybrid dysgenesis in *Drosophila melanogaster* by individual naturally occurring P elements. *Genetics* **133**: 605-622

Reed, D. S. (1993) Molecular structure and function relationships of low activity *sn*-Glycerol-3-phosphate dehydrogenase variants in *Drosophila melanogaster*. Thesis, The Australian National University

Reed, D. S. and Gibson, J. B. (1993) Defective P element insertions affect the expression of *sn*-glycerol-3-phosphate dehydrogenase alleles in natural populations of *Drosophila melanogaster*. *Proc. R. Soc. Lond. B* **251**: 39-45

Roberts, D. B. (1986) *Drosophila: a practical approach* IRL PRESS

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

Roeder, R. G. (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem. Sci.* **21**: 327-335

Roiha, H., Rubin, G. M. and O'Hare, K. (1988) P element insertions and re-arrangements at the *singed* locus of *Drosophila melanogaster*. *Genetics* **119**: 75-83

Rothberg, I., Hotaling, E. and Sofer, W. (1991) A *Drosophila Adh* gene can be activated in *trans* by an enhancer. *Nucleic Acids Res.* **19(20)**: 5713-5717

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

Rubin, G. M. (1983) Dispersed repetitive DNA sequences in *Drosophila*. In *Mobil Genetic Elements*. ed. J. A. Shapiro, pp. 329-361. Orlando Fla: Academic

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A. and Arnheim, N. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia Science **230**:1350-1354

Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988) Primer directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239**: 487-489

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning:* A laboratory manual (second edition) *Cold Spring Harbor Laboratory Press*

Savakis, C. and Ashburner, M. (1985) A simple gene with a complex pattern of transcription: the alcohol dehydrogenase gene of *Drosophila melanogaster*. *Cold Spring Harbor Symp. Quant. Biol.* **50**: 505-514

Savakis, C., Ashburner, M. and Wills, J. H. (1986) The expression of the gene coding for alcohol dehydrogenase during the development of *Drosophila melanogaster*. *Developmental Biology* **114**: 194-207

Schaeffer, S. W. and Aquadro, C. F. (1987) Nucleotide sequence of the *Adh* gene region of *Drosophila* pseudoobscura: evolutionary change and evidence for an ancient gene duplication. *Genetics.* **117(1)**: 61-73

Schott, D. R., East, P. D. and Paigen, K. (1988) Characterisation of the Adh^{SL} regulatory mutation in *Drosophila melanogaster Genetics* **119**: 631-637

Schwartz, M., O'Donnell, J. and Sofer, W. (1979) Origin of the multiple forms of alcohol dehydrogenase from *Drosophila melanogaster*. *Arch. Biochem. Biophys.* **194:** 365-378

Searles, L. L. and Voelker, R. A. (1986) Molecular characterisation of the *Drosophila vermillion* locus and its suppressible alleles. *Proc. Natl. acad. Sci. USA* **83**: 404-408

Sentenac, A., Riva, M., Thuriaux, P., Buhler, J., Treich, I., Carles, C., Wemer, M., Ruet, A., Huet, J., Mann, C., Chiannilkulchai, N., Stettler, S. and Mariotte, S. (1992) Yeast RNA polymerase subunits and genes. In: *Transcriptional Regulation*, 27-54. Edited by McKnight, S. L. and Yamamoto, K. R., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sheen, F., Lim, J. K. and Simmons, M. J. (1993) Genetic instability in *Drosophila melanogaster* mediated by *hobo* transposable elements. *Genetics.* **133**:315-334

Shen, N. L. L., Subrahmanyam G., Clark, W., Martin, P. and Sofer, W. (1989) Analysis of *Adh* gene regulation in *Drosophila*: studies using somatic transformation. *Devl. Genet.* **10**: 210-219

Shen, N. L. L., Hotaling, E. C., Subrahmanyam, G., Martin, P. F. and Sofer, W. (1991) Analysis of sequences regulation larval expression of the *Adh* gene of *Drosophila melanogaster*. *Genetics* **129**: 763-771

Smith, P. A. and Corces, V. G. (1991) *Drosophila* transposable elements: mechanisms of mutagenesis and interactions with the host genome. *Adv. Genet.* **29**: 229-300

Smith, D., Wohlgemuth, J., Calvi, B. R., Franklin, I. and Gelbart, W. M. (1993) *hobo* enhancer trapping mutagenesis in *Drosophila* reveals an insertion specificity different from P element. *Genetics* **135**: 1063-1076

Sofer, W. and Ursprung, H. (1968) *Drosophila* alcohol dehydrogenase: purification and partial characterization. *J. Biol. Chem.* **243**: 341-347

Sofer, W. and Hatkoff, M. A. (1972) Chemical selection of alcohol dehydrogenase-negative mutants in *Drosophila*. *Genetics* **72**: 545-49

Sofer, W. and Martin, P. F. (1987) Analysis of alcohol dehydrogenase gene expression in *Drosophila*. *Ann. Rev. Genet.* **21**:203-225

Spana, C. Harrison, D. A. and Corces, V. G. (1988) The *Drosophila melanogaster suppressor of Hairy-wing* protein binds to specific sequences of the gypsy retrotransposon. *Genes Dev* **2**: 1414-1423

Spana, C. and Corces, V. G. (1990) DNA bending is a determinant of binding specificity for a *Drosophila* zinc finger protein. *Genes Dev* **4**; 1505-1515

Spradling, A. C. and Rubin, G. M. (1983) The effect of chromosomal position on the expression of the *Drosophila* xanthine dehydrogenase gene. *Cell* **34**: 47-57

Strand, D. J. and McDonald J. F. (1989) Insertion of a *copia* element 5' to the *Drosophila melanogaster* alcohol dehydrogenase gene (*adh*) is associated with altered developmental and tissue-specific patterns of expression. *Genetics* **121**: 787-794

Streck, R. D., MacGaffey, J. E. and Beckendorf, S. K. (1986) The structure of hobo transposable elements and their insertion sites. *EMBO J.* **5**: 3615-3623

Sydney, M. P., Kimbrell, D., Hunkapiller, M., Hill, R., Fristrom, J. and Davidson, N. (1982) A transposable element that splits the promoter region inactivates a *Drosophila* cuticle protein gene. *Proc. Natl. Acad. Sci. USA* **79**: 7430-7434

Thatcher, D. R. and Camfield, R. (1977) Chemical basis of the electrophoretic variation between two naturally occurring alcohol dehydrogenase alloenzymes from *Drosophila melanogaster*. *Biochem. Soc. Trans.* **5(1)**: 271-2

Thatcher, D. (1980) The complete amino-acid sequence of three alcohol dehydrogenase alloenzymes (Adh n-11, Adh-S and Adh-uf) from the fruit-fly *Drosophila melanogaster*. *Biochem. J.* **187**: 875-876

Thomas, U., Jonsson, F., Speicher, S. A. and Knust, E. (1995) Phenotypic and molecular characterization of *Ser*^{D}, a dominant allele of the *Drosophila* gene *Serrate*. *Genetics* **139**: 203-213

Thompson, J. N., JR. and Kaiser, T. N. (1977) Selection action upon slow-migrating ADH alleles differing in enzyme activity. *Heredity* **38**: 191-195

Thompson, J. N., JR., Ashburner, M. and Woodruff, R. C. (1977) Presumptive control mutation for alcohol dehydrogenase in *Drosophila melanogaster*. *Nature* **270:** 363

Triglia, T., Peterson, M. G. and Kemp, D. J. (1988) A procedure for *in vitro* amplification of DNA segments that lie outside the boundaries of known sequences. *Nucl. Acids Res.* **16**: 8186

Truett, M. A., Jones, R. S. and Potter, S. S. (1981) Unusual structure of the FB family of transposable elements in *Drosophila. Cell* **24**: 753-763

Tsubota, S., Ashburner, M. and Schedl, P. (1985) P element induced control mutations at the *r* gene of *Drosophila melanogaster*. *Mol. Cell. Biol.* **5**: 2567-2574

Tsukiyama, T. and Wu, C. (1995) Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83**: 1011-1020

Ursprung, H., Sofer, W. and Burroughs, N. (1970) Ontogeny and tissue distribution of alcohol dehydrogenase in *D. melanogaster*. *Wilhelm Roux Arch. Entwichlungsmech.* **164:** 201-208

Van Delden, W. (1982) The alcohol dehydrogenase polymorphism in *Drolsophila melanogaster*. Selection at an enzyme locus. *Evol. Biol.* **15:** 187

Van Dyke, M. W., Roeder, R. G. and Sawadogo, M. (1988) Physical analysis of transcription preinitiation complex assembly on a class II gene promoter. *Science* **241**: 1335-1338

Verrijzer, C. P. and Tjian, R. (1996) TAFs mediate transcriptional activation and promoter selectivity. *Trend Biochem. Sci.* **21**: 338-341

Vigue, C. L. and Johnson, F. M. (1973) Isozyme variability in species of the genus *Drosophila*. VI. Frequency-property-environment relationships of allelic alcohol dehydrogenases in *D. melanogaster*. *Biochem. Genet.* **9(3)**: 213-27

Vigue, C. and Sofer, W. (1976) Chemical selection of mutants that affect alcohol dehydrogenase activity in *Drosophila*: 3. Effects of ethanol. *Biochem. Genet.* **14:** 127-35

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 natural populations of *Drosophila melanogaster*: Frequencies in a North Carolina population. *Proc. Natl. Acad. Sci. USA* **77**: 1091

Voelker, R. A., Graves, J., Gibson, W. and Eisenberg, M. (1990) Mobile element insertions causing mutations in the *Drosophila suppressor of sable* locus occur in DNase I hypersensitive subregions of 5'-transcribed nontranslated sequences. *Genetics* **126**: 1071-1082

White, R. J. and Jackson, S. P. (1992) The TATA-binding protein: a central role in transcription by RNA polymerases I, II and III. *Trends Genet.* **8**: 284-288

Wilks, A. V., Gibson, J. B., Oakshott, J. G. and Chambers, G. K. (1980) An electrophoretically cryptic alcohol dehydrogenase variant in *Drosophila melanogaster*. II. Post electrophoresis heat treatment screening of natural populations. *Aust. J. Biol. Sci.* **33,** 575-585

Willis, I. M. (1993) RNA polymerase III. *Eur. J. Biochem.* **212**: 1-11

Winberg, J.-O., Thatcher, D. R., and McKinley-McKee, J. S. (1982) Alcohol dehydrogenase from the fruit fly *Drosophila melanogaster*: Substrate specificity of the alleloenzymes ADH^S and ADH^{UF}. *Biochem. Biophys. Acta* **704**: 7-16

Winberg, J., Hovik, R. and McKinley-McKee, J. S. (1985) The alcohol dehydrogenase alleloenzymes Adh^S and Adh^F from the fruitfly *Drosophila melanogaster*: an enzymatic rate assay to determine the active-site concentration. *Biochem. Genet.* **23**: 205-216

Wolffe, A. P. (1994) Nucleosome positioning and modification: chromatin structures that potentiate transcription. *Trend Biochem. Sci.* **19**: 240-244

Woodruff, R. C. and Ashburner, M. (1979) The genetics of a small autosomal region of *Drosophila melanogaster* containing the structure gene for alcohol dehydrogenase EC 1.1.1.1. 1. Characterisation of deficiencies and mapping of *Adh* and visible mutations. *Genetics* **92**: 117-132

Woodruff, R. C., Blount, J. L. and Thompson, J. N. (1987) Hybrid dysgenesis is not a general release mechanism for DNA transpositions. *Science* **237**: 1206-1207

Yannopoulos, G., Stamatis, N., Zacharopoulou, A. and Pelecanos, M. (1983) Site specific breaks induced by the male recombination factor 23.5 MRF in *Drosophila melanogaster*. *Mutat. Res.* **108:** 185-202

Yannopoulos, G., Stamatis, N., Monastirioti, M., Hatzopoulos, P., and Louis, C. (1987) hobo in responsible for the induction of hybrid dysgenesis by strains of *Drosophila melanogaster* bearing the male recombination factor 23.5 MRF. *Cell* **49**: 487-495

Yannopoulos, G., Zabalou, S., Stamatis, N. and Tsamathis, G. (1994) Differential regulation of *P* and *hobo* mobile elements by two laboratory strains of *Drosophila melanogaster*. *Genet. Res.* **63:** 129-137

Young, M. W. and Schwartz, H. E. (1981) Nomadic gene families in *Drosophila*. *Cold spring Harb. Symp. Quant. Biol.* **45**: 629-640

Yukuhiro, K., Harada, K. and Mukai, T. (1985) Viability mutations induced by the P element in *Drosophila melanogaster*. *Jpn. J. Genet.* **60**:531-537

Zachar, Z. and Bingham, P. M. (1982) Regulation of white locus expression: the structure of mutant alleles at the white locus of *Drosophila melanogaster*. *Cell* **30**, 529-541

Zachar, Z., Davison, D., Garza, D. and Bingham, P. M. (1985) A detailed developmental and structural study of the transcriptional effects of insertion of the *copia* transposon into the *white* locus of *Drosophila melanogaster*. *Genetics* **111**: 495-515

Zapp, M. L. and Berget, S. M. (1989) Evidence for nuclear factors involved in recognition of 5' splice sites. *Nucleic Acids Res.* **17(7)**: 2655-74

Zawel, L. and Reinberg, D. (1993) Initiation of transcription by RNA polymerase II: a multi-step process. *Prog. Nucl. Acid Res. Mol. Biol.* **44**: 167-180

APPENDIX I

Comparison of the DNA sequence of the *Adh* gene in the three low activity alleles with the consensus *Adh* gene sequence.

The consensus sequence of *Adh* is shown from nucleotide -720 to +2000 (Numbering according to Kreitman (1983); Kreitman and Hudson, 1991; Benyajati *et al*, 1983). The *Adh* adult enhancer (AAE) is shown in magenta. The distal and proximal TATA boxes are shown in red. The 29 nucleotides of ∇1 polymorphism in the consensus sequence are indicated in bold. The *Adh* sequence from -63 to +1780 in *AD369* is shown in blue, and *AHA51* in green. The *Adh* sequence in *AAS44* is shown in light blue from +288 to +409 and +650 to +1697. The target site duplication in each allele is shown and underlined. "." refers to the consensus nucleotide at that position. The A to C polymorphism at position 1490 is indicated with a " * ".

```
-720                    -700
AGTGAGATTGCTAATGAGCT    GCTTTTAGGGGGCGTGTTGT
-680                    -660
GCTTGCTTTCCAACTTTTCT    AGATTGATTCTACGCTGCCT
-640                    -620
CCAGCAGCCACCCCTCCCAT    CCCCATCCCCATCACCATCC
-600                    -580
AGTCCCGTTGGCTCCCAGTC    ACAGTATTACACGTATGCAA
-560                    -540
ATTAAGCCGAAGTTCAATTG    CGACCGCAGCAACAACACGA
-520                    -500
TCTTTCTACACTTCTCCTTG    CTATGCTTGACATTCACAAG
-480                    -460
GTCAAAGCTCTTAATATTCT    GGCTCGTGGCCCTACACTGT
-440                    -420
AAGAAATTACTATAGAAATA    ACGGTACACGGAATAAGATA
-400                    -380
TTTTTTTTAGTCCATATGCT    TTTAACAAATGTGTTTTGAG
-360                    -340
TTTATGTTATATTATTGTTA    GAAAACCGGTGTTTTTTTTT
-320                    -300
AAATCGGTTAAAAAATTACT    ACGAGAGAAAAATACAAATT
-280                    -260
TTGTAAATAAGATTGACTCT    TTTTAGATTTTGGAATATTT
-240                    -220
TCATTCATTTTATGTTTTTA    CGTTTTCACTTATTTGTTTC
```

```
                -200                       -180
                TCAGTGCACTTTCTGGTGTT       CCATTTTCTATTGGGCTCTT
                -160                       -140
                TACCCCGCATTTGTTTGCAG       ATCACTTGCTTGCGCATTTT
                -120                       -100
                TATTGCATTTTACATATTAC       ACATTATTTGAACGCCGCTG
                -80                        -60
                CTGCTGCATCCGTCGACGTC       GACTGCACTCGCCCCCACGA
                                 ...       ....................
                                 ...       ....................

                -40                        -20                -1
Consensus       GAGAACAGTATTTAAGGAGC       TGCGAAGGTCCAAGTCACCG
AD369           ....................       .......GTCCAAGT..TGC
AHA51           ....................       .......GTCCAAGT..TGC
AAS44

                1                          21
                ATTATTGTCTCAGTGCAGTT       GTCAGTTGCAGTTCAGCAGA
                ....................       ....................
                ....................       ....................

                41                         61
                CGGGCTAACGAGTACTTGCA       TCTCTTCAAATTTACTTAAT
                ....................       ....................
                ....................       ....................

                81                         101
                TGATCAAGTAAGTAGCAAAA       GGGCACCCAATTAAAGGAAA
                ....................       ......A.....G.......
                ....................       ......A.....G.......

                121                        141
                TTCTTGTTTAATTGAATTTA       TTATGCAAGTGCGGAAATAA
                ....................       ....................
                ....................       ..G.................

                161                        181
                AATGACAGTATTAATTAGTA       AATATTTTGTAAAATCATAT
                ....................       ....................
                ........G...G.A.....       ....................

                201                        221
                ATAATCAAATTTATTCAATC       AGAACTAATTCAAGCTGTCA
                ....................       ....................
                ....................       ....................

                241                        261
                CAAGTAGTGCGAACTCAATT       AATTGGCATCGAATTAAAAT
                ....................       ....................
                ....................       ....................

                281                        301
                TTGGAGGCCTGTGCCGCATA       TTCGTCTTGGAAAATCACCT
                ............T.......       ...C................
                ....................       ....................
                        ....T.......       ...C................

                321                        341
                GTTAGTTAACTTCTAAAAAT       AGGAATTTTAACATAACTCG
                ....................       ....................
                ....................       ....................
                ....................       ....................

                361                        381
                TCCCTGTTAATCGGCGCCGT       GCCTTCGTTAGCTATCTCAA
                ....................       ....................
                ....................       ....................
                ....................       ....................
```

401	421
AAGCGAGCGCGTGCAGACGA	GCAGTAATTTTCCAAGCATC
....	..T.................
....	....
.....AGCG

441	461
AGGCA**TAGTTGGGCATAAAT**	**TATAAACATACAAACC**GAAT
.....TAATATACTAATACTAATACTAATACTAATATAA ....
....	....

481	501
ACTAATATAGAAAAAGCTTT	GCCGGTACAAAATCCCAAAC
....	....
....	....

521	541
AAAAACAAACCGTGTGTGCC	GAAAAATAAAAATAAACCAT
....	....
....	....

561	581
AAACTAGGCAGCGCTGCCGT	CGCCGGCTGAGCAGCCTGCG
....	.....A..............
............A..C.G..	....

601	621
TACATAGCCGAGATCGCGTA	ACGGTAGATAATGAAAAGCT
....	....
....	....

641	661
CTACGTAACCGAAGCTTCTG	CTGTACGGATCTTCCTATAA
....	....
....	....
....	....

681	701
ATACGGGGCCGACACGAACT	GGAAACCAACAACTAACGGA
....	....
......A.............	....
....	....

721	741
GCCCTCTTCCAATTGAAACA	GATCGAAAGAGCCTGCTAAA
....	....
....	....
....	....

761	781
GCAAAAAAGAAGTCACCATG	TCGTTTACTTTGACCAACAA
....	....
....	....
....	....

801	821
GAACGTGATTTTCGTTGCCG	GTCTGGGAGGCATTGGTCTG
....	....
....	....
................G...	....

841	861
GACACCAGCAAGGAGCTGCT	CAAGCGCGATCTGAAGGTAA
....	....
....	....
....	....

881	901
CTATGCGATGCCCACAGGCT	CCATGCAGCGATGGAGGTTA
....	....
....	....
....	....

```
921                      941
ATCTCGTGTATTCAATCCTA     GAACCTGGTGATCCTCGACC
....T...............     ....................
....................     ....................
....................     ....................

961                      981
GCATTGAGAACCCGGCTGCC     ATTGCCGAGCTGAAGGCAAT
....................     ....................
....................     ....................
....................     ....................

1001                     1021
CAATCCAAAGGTGACCGTCA     CCTTCTACCCCTATGATGTG
....................     ....................
....................     ....................
....................     ....................

1041                     1061
ACCGTGCCCATTGCCGAGAC     CACCAAGCTGCTGAAGACCA
....................     ....................
....................     ....................
....................     ....................

1081                     1101
TCTTCGCCCAGCTGAAGACC     GTCGATGTCCTGATCAACGG
....................     ....................
....................     ....................
....................     ....................

1121                     1141
AGCTGGTATCCTGGACGATC     ACCAGATCGAGCGCACCATT
....................     ....................
....................     ....................
....................     ....................

1161                     1181
GCCGTCAACTACACTGGCCT     GGTCAACACCACGACGGCCA
....................     ....................
....................     ....................
....................     ....................

1201                     1221
TTCTGGACTTCTGGGACAAG     CGCAAGGGCGGTCCCGGTGG
....................     ..............A.....
....................     ....................
....................     ....................

1241                     1261
TATCATCTGCAACATTGGAT     CCGTCACTGGATTCAATGCC
....................     ....................
....................     ....................
....................     ....................

1281                     1301
ATCTACCAGGTGCCCGTCTA     CTCCGGCACCAAGGCCGCCG
....................     ....................
....................     ....................
....................     ....................

1321                     1341
TGGTCAACTTCACCAGCTCC     CTGGCGGTAAGTTGATCAAA
....................     ....................
....................     ....................
....................     ....................

1361                     1381
GGAAACGCAAAGTTTTCAAG     AAAAAACAAAACTAATTTGA
....................     ....................
....................     ....................
....................     ....................
```

```
1401                    1421
TTTATAACACCTTTAGAAAC    TGGCCCCCATTACCGGCGTG
....................    ....................
....................    ....................
....................    ....................

1441                    1461
ACCGCTTACACCGTGAACCC    CGGCATCACCCGCACCACCC
..G........T........    ....................
....................    ....................
...........T........    ....................

1481      *             1501
TGGTGCACAAGTTCAACTCC    TGGTTGGATGTTGAGCCCCA
.........C..........    .................T..
....................    ....................
.........C..........    .................T..

1521                    1541
GGTTGCTGAGAAGCTCCTGG    CTCATCCCACCCAGCCATCG
......C.............    ................C...
....................    ....................
......C.............    ................C...

1561                    1581
TTGGCCTGCGCCGAGAACTT    CGTCAAGGCTATCGAGCTGA
....................    ....................
....................    ...............A....
....................    ....................

1601                    1621
ACCAGAACGGAGCCATCTGG    AAACTGGACTTGGGCACCCT
....................    ....................
....................    ....................
....................    ....................

1641                    1661
GGAGGCCATCCAGTGGACCA    AGCACTGGGACTCCGGCATC
....................    ....................
....................    ....................
....................    ....................

1681                    1701
TAAGAAGTGATAATCCCAAA    AAAAAAAACATAACATTAGT
............C.......    ....................
....................    ....................
............C....

1721                    1741
TCATAGGGTTCGCGAACCAC    AAGATATTCACGCAAGGCAA
....................    ....................
....................    ....................

1761                    1781
TTAAGGCTGATTCGATGCAC    ACTCACATTCTTCTCCTAAT
....................
....................
1801                    1821
ACGATAATAAAACTTTCCAT    GAAAAATATGGAAAAATATA
1841                    1861
TGAAAATTGAGAAATCCAAA    AAACTGATAAACGCTCTACT
1881                    1901
TAATTAAAATAGATAAATGG    GAGCGGCAGGAATGGCGGAG
1921                    1941
CATGGCCAAGTTCCTCCGCC    AATCAGTCGTAAAACAGAAG
1961                    1981
TCGTGGAAAGCGGATAGAAA    GAATGTTCGATTTGACGGGC
```

APPENDIX II

Oligonucleotide primers used in *Adh* sequence analyses

Oligonucleotides primers used for the polymerase chain reaction and sequencing the *Adh* gene are listed. The oligonucleotides are ranked from the beginning of the sequence. The nucleotide numbers and oligonucleotide sequences are based on the *Adh* sequence in Krietman (1983). The number in the name of the oligonucleotide represents the order in which they were synthesised. The melting temperature (Tm) was calculated using the equation Tm=4GC+2AT. All oligos were produced on a 380B Applied Biosystems DNA synthesiser.

Oligo name	Positions	Sequences	Tm	Size (mer)
JG97	-688~-670	5' CGTGTTGTGCTTGCTTTCC 3'	58	19
JG88	-458~-440	5' CTCGTGGCCCTACACTGTA 3'	60	19
JG87	-197~-179	5' GTGCACTTTCTGGTGTTCC 3'	58	19
JG95	-166~-147	5' GCTCTTTACCCCGCATTTG 3'	58	19
JG16b	-67~-48	5' CGACGTCGACTGCACTCGCC 3'	68	20
JG98*	-11~-30	5' GACCTTCGCAGCTCCTTAAA 3'	60	20
JG96*	67~49	5' GAAGAGATGCAAGTACTCG 3'	56	19
JG82	84~104	5' CAAGTAAGTAGCAAAAGGGC 3'	58	20
JG83*	104~ 84	5' GCCCTTTTGCTACTTACTT 3'	58	20
JG13	288~307	5' CCTGTTCCGCATATTCCTCT 3'	60	20
JG22*	307~288	5' AGAGGAATATGCGGAACAGG 3'	60	20
JG102*	454~437	5' GCCCAACTATGCCTGATG3'	56	18
JG14	647~666	5' AACCGAAGCTTCTGCTGTAC 3'	60	20

JG17*	666~647	5' GTACAGCAGAAGCTTCGGTT 3'	60	20
JG15	987~1006	5' GAGCTGAAGGCAATCAATCC 3'	60	20
JG21	1006~ 987	5' GGATTGATTGCCTTCAGCTC 3'	60	20
JG84	1141~1160	5' ACCAGATCGAGCGCACCATT 3'	62	20
JG5	1320~1339	5' GTGGTCAACTTCACCAGCTC 3'	62	20
JG18*	1339~1320	5' GAGCTGGTGAAGTTGACCAC 3'	62	20
JG1	1546~1564	5' CCCACCCAGCCCTCGTTGG 3'	66	19
JG6*	1687~1668	5' CTTCTTAGATGCCGGAGTCC 3'	62	20
JG48*	1781~1764	5' TGTGCATCGAATCAGCCT 3'	54	18
JG85*	1910~1891	5' CCTGCCGCTCCCATTTATCT 3'	62	20
JG86*	1974~1955	5' TCCGCTTTCCACGACTTCTG 3'	62	20

* primers complementary to the coding strand.

APPENDIX III

Oligonucleotide primers used in KP and hobo element sequence analyses

Oligonucleotide primers used for the polymerase chain reaction and sequencing the KP element and hobo element are listed. The oligonucleotides are ranked from the beginning of the sequence. The position numbers and oligonucleotide sequences are based on the KP element sequence in Black *et al* (1987) and the hobo element sequence in Streck *et al* (1986).

Oligo name	Element name	Position	Sequences	Tm	Size (mer)
JG68	KP	36~58	5' CCGAAGCTTACCGAAGTATAC3'	62	21
JG73*	KP	103~ 84	5' GCACACAACCTTTCCTCTCA3'	60	20
JG94	KP	249~257	5' TTGGGTGCAGCCTTGGTG3'	58	18
JG93*	KP	354~337	5' GTCGCCTCCTTTTAAATG3'	52	18
JG70	KP	483~502	5' AACGTGACTGTGCGTTAGGT3'	60	20
JG71*	KP	573~554	5' ACGACGCATTTCGTACTCCA3'	60	20
JG92	KP	757~776	5' GGATTTCCTTTGCCCAGTCG3'	62	20
JG91*	KP	831~812	5' CTTGTTTATCAACATCGACG3'	56	20
JG72	KP	992~1011	5' CAAACCCCACGGATATGCTA3'	60	20
JG67*	KP	1011~ 992	5' TAGCATATCCGTGGGGTTTG3'	60	20
JG100*	hobo	231~214	5' CCCACAAGTGTTTTTCTC3'	52	18
JG101	hobo	2772~2790	5' AGAACTGCAGCCCGCATAC3'	60	19

APPENDIX IV

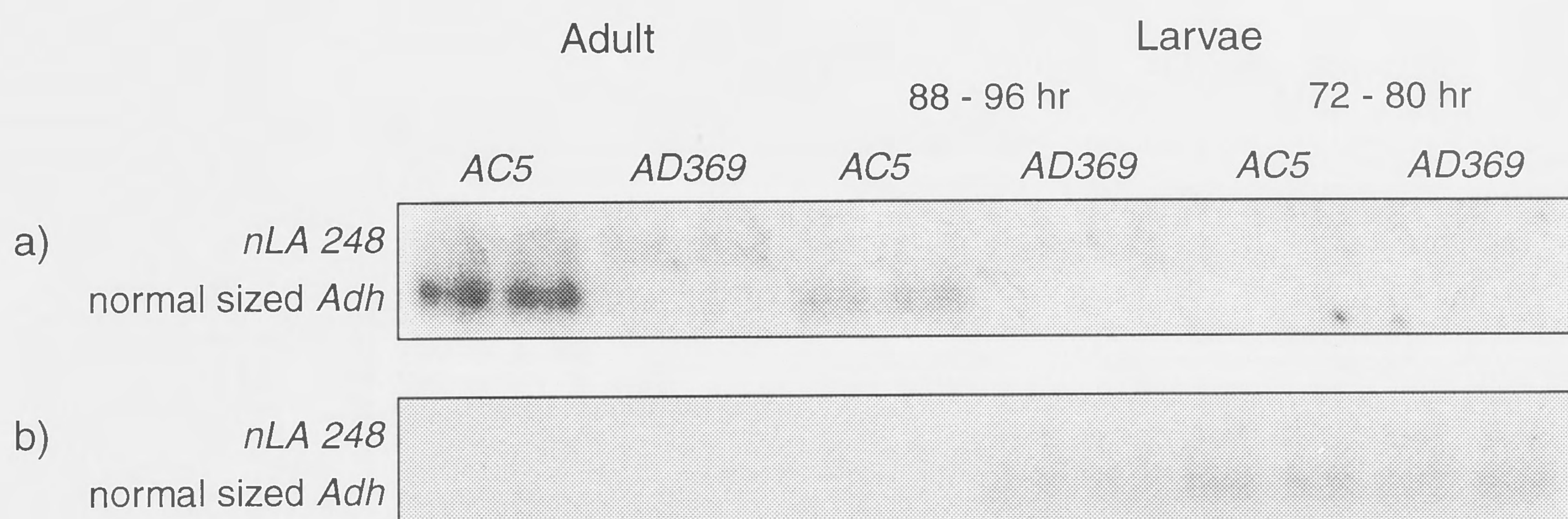
Comparison of the *Tirant* element sequences in insertions at the *Adh* locus in *AAS44* and at the *emc* locus in the *Ach* mutant (Garrell and Modolell, 1990)

Sequence of the ends of the *Tirant* element at the *emc* locus is shown in black, with the 2-bp inverted terminal repeat sequences in magenta and polyadenylation signal in red. The 5' termini of the *Tirant* element sequence is marked as +1. The sequence of the 5' end of the *Tirant* element at the *Adh* locus is shown in light blue and aligned with the *Tirant* sequence in *Ach*. The 4-bp duplication at the insertion site in each locus is underlined.

```
    +1                     +17
CCCGAGTTAAGTCTGTGATC       GAGGGTGGAGCCTTTTGTGA
    ::::::::: ::::::       ::::::::::::::: ::::
AGCGAGTTAAGTCCGTGATC       GAGGGTGGAGCCTTT-GTGA

+37                        +57
AGATATGATGTGTAAATGTA       TGAAGGTTTTTAAGTTTCAG
::::::::::::::::::::       ::::::::::::::
AGATATGATGTGTAAATGTA       TGAAGGTTTT-AAG

+77                        +97
CAGATAGTTGTAGTGTGAAG       ATTTCCAATAAAGAATACGA
+107
CTGACTT.............       ....................

....GGACAATGTTGCAGTG       CGACTTTTGTGTAATGTCAG

CATTCGGCCGGAATGTCAGC       ATTCGGCCGGATGCGGTGTT

CCAGTTGCCGTCTTCCCGGT       TGAGTGTCCGATGATTTGTT

TAGAGGCGTGGGGGTTTAGG       GGGTGGGGTGGTAACTCCCG
```

APPENDIX V

Northern analysis of *Adh* transcripts in adults and larvae of *AD369/nLA248* and *AC5/nLA248* heterozygotes



Samples were loaded as indicated. Filter a) was probed with the distal promoter specific primer JG96 (Appendix II). Filter b) was probed with the primer JG113 (729-710, 5'GAAGAGGGCTCCGTTAGTTG3') — a proximal promoter specific probe. On each filter, the upper band represents the *nLA248* specific hybridisation signal, the lower band represents the normal *Adh* specific signal.

On filter a), the distal promoter specific probe JG96 detected adult transcript from both *AC5* and *AD369* adults and from *AC5* late third instar larvae (88-96 hours), but not from *AD369* late third instar larvae.

On filter b), the proximal promoter specific probe JG113 detected larval transcript from both *AC5* and *AD369* early third instar (72-80 hours) larvae, and also from late third instar larvae of *AD369*.

These results indicate that the distal promoter in *AD369* does not transcribe until the adult stage. The proximal promoter keeps transcribing from early through late third instar larvae in *AD369*. In the control allele *AC5*, *Adh* transcription switches from the proximal promoter to the distal promoter in late third instar larvae (Corbin and Maniatis, 1989).