

**GEOGRAPHIC VARIATION IN TRANSPOSABLE ELEMENTS AND ISOZYMES
IN SOUTHERN AFRICAN POPULATIONS OF *Drosophila melanogaster***

Chonai Greer Louise Getz

**A thesis submitted to the Faculty of Science,
University of the Witwatersrand, Johannesburg
for the Degree of Doctor of Philosophy.**

Johannesburg 1990

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to the following people for their help and contribution towards this study:

My two supervisors, Professor Nancy van Schaik and Dr Stewart Grant for their constant help, advice, guidance and friendship.

Gwen MacLachlan for her expertise, efficiency and patience in the typing of this manuscript.

My family Maish, Matty and Ally for their love, support, encouragement and good humour.

All those people who so assiduously and willingly collected flies throughout the country for me: Florrie Baron, Margot Berger, Paul Bibby, Professor and Mrs Peter Geerthsen, Maish Getz, Dee Gillette, Dr Stewart Grant, Helen Hirsch, Barney Israels, Stephanie Le Clerq, Isla Locke, Rosemary Maisels, Martine Margoles, Martha Mashila, Pauline Schwartzman, Dr Mary Seely, Louis van Heerden and Professor Nancy van Schaik.

Jocelyn Webster and Susan Hearn for their supervision, guidance and assistance in the molecular laboratory.

Professor Paul Fatti for his statistical advice.

Mary Lynn van der Watt for aid in the use of the scanning electron microscope.

Professor Barry Fabian for allowing me the use of the Nikon microscope and camera equipment in the Zoology Department of the University of the Witwatersrand and Clem Penny for showing me how to use them.

Professor William R Engels for supplying the strains $sn^W(M)$ and π_2 and the plasmid $p\pi 25.1$.

The Institute of Toxicology Schwerzenbach for supplying the strains mwh and $f1r^3/Ser$.

Professor Nancy van Schaik for constructing and supplying the strain πmwh .

Dr Ulrich Graf for obtaining papers that were unavailable in South Africa.

Dr Heinz Juon and Professor F.E. Würgler for permission to use their electron micrographs of wing hairs.

Professor Jean R. David, Professor Margaret G. Kidwell and Professor Ron C. Woodruff for permission to quote from their articles in Drosophila Information Service.

Dr Stewart Grant for taking the photographs of the starch gels.

Dimitri Vratsanos and Philip Tshabalala for developing some of the photographs.

Sike Margoles and Linda Steenkamp for computer assistance with the diagrams and Kay Styer for all the photocopying.

DECLARATION

I declare that this thesis is my own, unaided work except where otherwise stated. It is being submitted for the degree of Doctor of Philosophy in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination in any other University.

Johnat Gitz

2nd day of May 1990

DEDICATION

To my family

Maish, Matty and Ally

ABSTRACT

This thesis reports on the investigation of two genetic aspects of Southern African populations of *Drosophila melanogaster*: the family of transposable elements and the allelic variation present in several enzyme systems.

In the study of transposable elements, 206 isofemale lines from 21 different populations were surveyed for their ability both to produce and resist gonadal dysgenesis. Both P factor activity and regulatory ability exhibited clinal variation, being strong in the northeast in Harare, and becoming extremely weak in the southwestern part of the country in Cape Town.

Three further aspects of P elements were investigated in this study: Firstly P factor activity was evaluated using the level of *single-weak* mutability as a measure. Secondly the presence of P elements in M strains, noted on a worldwide basis, was confirmed for the Southern African strains using molecular techniques. Finally a low level of P element activity was shown to be present in wing imaginal disc tissue of some female flies.

In the study of allozyme heterogeneity, 17 different populations comprising 301 isofemale lines were investigated. Starch gel electrophoretic techniques were used to analyse variation at eight gene loci in these populations. A significant degree of allele frequency heterogeneity was present at several hierarchical levels of geographic separation.

A gene diversity analysis indicated that 87.8% of the total genetic variation was due to variation among individuals within populations. The remaining 12.2% was due to geographic subdivision at various hierarchical levels of structure.

A gene flow analysis suggested that there was a large amount of gene flow among Southern African populations. These populations appeared to be closely related to Central and West African populations and together probably form a single evolutionary unit.

The frequencies of the most common allozymes of Alcohol dehydrogenase and Isocitrate dehydrogenase were found to vary with increasing rainfall and minimum temperature respectively.

The possible effects of P element activity on the presence and distribution of the allozymes under consideration was investigated. However there was no correlation between P element geographic variation and allele frequency variation.

TABLE OF CONTENTS

	Page
Acknowledgements...	i
Declaration ...	iii
Dedication. ...	iv
Abstract...	v
Table of Contents...	vii
List of Tables. ...	xiii
List of Figures ...	xv
CHAPTER 1. BACKGROUND AND MOTIVATION...	1
1.1 Elements...	1
1.1.1 Introduction ...	1
1.1.1.1 Mobile genetic elements in <i>Drosophila</i> ...	4
1.1.1.2 Systems of hybrid dysgenesis ...	6
1.1.2 P-M hybrid dysgenesis - historical commentary...	10
1.1.2.1 Mutations...	10
1.1.2.2 Insertion mutants and hybrid dysgenesis. ...	13
1.1.3 Physiological and molecular characteristics of the P-M system ...	16
1.1.3.1 Gonadal dysgenesis...	17
1.1.3.2 Molecular characteristics...	18
1.1.4 Motivation for the project...	21
1.2 Allozymes...	24
1.2.1 Introduction ...	24
1.2.2 Background for the allozyme study...	25
CHAPTER 2. GEOGRAPHICAL DISTRIBUTION OF P ELEMENTS IN SOUTHERN AFRICA...	31
2.1 Introduction...	31
2.1.1 Classification of wild <i>Drosophila melanogaster</i> ...	32
2.2 Materials and methods...	35
2.2.1 Collection of wild <i>D. melanogaster</i> and establishment of isofemale lines...	35
2.2.2 Geographical distribution of samples ...	36
2.2.3 <i>Drosophila</i> stocks used for P-M testing...	38

	<u>Page</u>
2.2.4 Description of crosses made and methods used for GD sterility testing.. . . .	38
2.2.5 Confidence intervals	40
2.3 Results.	44
2.3.1 Classification scheme.	44
2.3.1.1 Analysis of Southern African wild strains.	45
2.3.1.2 Indefinite strains & intermediate strains.	67
2.3.2 Polymorphisms and trends	68
2.3.3 Intrastrain and reciprocal crosses.	70
2.3.3.1 Reciprocal crosses.	70
2.3.3.2 Intrastrain crosses.	71
2.3.4 Geographical distribution.	75
2.3.4.1 Clines.	75
2.4 Discussion.	79
2.4.1 Origins of P elements.	80
2.4.2 Dynamics of P elements.	87
2.4.3 Regulatory ability.	92
2.4.3.1 Autonomous P elements involved in regulation	93
2.4.3.2 KP and other deletion derivatives.	95
2.4.4 Movement of P elements.	98
2.4.4.1 Initial conditions and P element dynamics.	98
2.4.4.2 Temperature.	100
2.4.4.3 Stochastic effects.	100
2.4.4.4 Temporal trends.	101
 CHAPTER 3. sn^W MUTABILITY.	 104
3.1 Introduction	104
3.1.1 sn^W Destabilization as an assay system for P factor activity.	111
3.1.2 The relationship between GD sterility and sn^W mutability	113
3.2 Materials and methods.	114
3.3 Results.	115
3.4 Discussion.	118
3.4.1 M strains and Q strains.	120
3.4.2 Exceptional cases.	122

	<u>Page</u>
CHAPTER 4. MOLECULAR ANALYSIS OF P ELEMENTS	124
4.1 Introduction	124
4.1.1 Cloning of P elements... ..	124
4.1.2 Objectives... ..	127
4.2. Materials and methods... ..	127
4.2.1 Extraction of <i>Drosophila</i> DNA	127
4.2.2 Preparation of the probe	128
4.2.2.1 Plasmid isolation... ..	130
4.2.2.2 Digestion of p π 25.1.	132
4.2.2.3 Radioactive labelling of DNA	134
4.2.3 Agarose gel electrophoresis.	135
4.2.4 Hybridization... ..	136
4.2.4.1 Binding of DNA to nitrocellulose	136
4.2.4.2 Prehybridization	136
4.2.4.3 Hybridization... ..	137
4.2.4.4 Posthybridization... ..	137
4.2.4.5 Autoradiography.	137
4.3 Results.	137
4.3.1 Digestion of plasmid p π 25.1.	137
4.3.2 Dot blot hybridization..	141
4.4 Discussion..	150
4.4.1 Wild M strains possess P elements... ..	150
4.4.2 Intensity of hybridization signal... ..	151
4.4.3 Susceptibility to P activity	155
4.4.4 Q strains... ..	158
 CHAPTER 5. P ELEMENT SOMATIC ACTIVITY..	 160
5.1 Introduction	160
5.1.1 Motivation and background... ..	161
5.2. Materials and methods... ..	163
5.2.1 <i>Drosophila</i> stocks... ..	163
5.2.2 Testing of stocks... ..	164

	<u>Page</u>
5.2.3 Testing for somatic P element activity in dysgenic flies	165
5.2.3.1 Clone formation.	166
5.2.3.2 Preparation and microscopic analysis of wings...	166
5.2.3.3 Scoring of wings	170
5.2.3.4 Statistical procedures.. ...	175
5.3 Results.	175
5.3.1 GD sterility	175
5.3.2 Mutant clones... ..	177
5.3.3 Proportion of different spot sizes for single spots.	180
5.4 Discussion.. ...	181
5.4.1 Developmental temperature and sex of affected progeny... ..	183
5.4.2 Frequency of spots.. ...	185
5.4.3 Conclusion.. ...	187
 CHAPTER 6. ALLOZYMES.	 189
6.1. Introduction... ..	189
6.1.1 Selection and neutralism	189
6.1.1.1 Gene flow... ..	191
6.1.1.1.1 Effectiveness of gene flow.. ...	191
6.1.1.1.2 Measurement of gene flow	192
6.2. Materials and methods... ..	195
6.2.1 Geographic distribution of samples.. ...	195
6.2.2. Electrophoretic analysis	195
6.2.2.1 Sample preparation.. ...	198
6.2.2.2 Gel preparation, loading and running	198
6.2.2.3 Cutting and staining of gels	199
6.2.2.4 Genetic interpretation of banding patterns.. ...	199
6.2.2.5 Isozyme/allozyme nomenclature and homology.. ...	201
6.2.3 Statistical analysis	202
6.2.3.1 Hierarchical population model... ..	202
6.2.3.2 Allele frequency differences	204
6.2.3.3 Gene diversity analysis.	206
6.3 Results.	209
6.3.1 Biochemical genetic variation... ..	209

	<u>Page</u>
6.3.1.1 Alcohol dehydrogenase (<i>Adh</i>)	210
6.3.1.2 Dipeptidase-A (<i>Dp-A</i>)	217
6.3.1.3 sn-Glycerol-3-phosphate dehydrogenase (<i>Gpdh</i>)	217
6.3.1.4 Hexokinase (<i>Hex-1</i> , <i>Hex-2</i>)	218
6.3.1.5 Isocitrate dehydrogenase-NADP (<i>Idh-NADP</i>)	221
6.3.1.6 6-Phosphogluconate dehydrogenase (<i>Pgd</i>)	222
6.3.1.7 Phosphoglucosmutase (<i>Pgm</i>)	222
6.3.2 Allele frequencies	223
6.3.2.1 Allele frequency ranges and clinal patterns	226
6.3.3 Nested contingency table analysis (G-tests)	226
6.3.3.1 Heterogeneity between collection sites within districts	238
6.3.3.2 Heterogeneity between districts within areas	238
6.3.3.3 Heterogeneity between areas within regions	238
6.3.3.4 Heterogeneity between regions	239
6.3.3.5 Comparison of heterogeneity between levels	239
6.3.4 Gene diversity analysis	243
6.3.4.1 Between collection sites within districts	243
6.3.4.2 Between districts within areas	246
6.3.4.3 Between areas within regions	246
6.3.4.4 Between regions	247
6.3.4.5 Comparison of gene diversity among population levels	248
6.3.5 Estimation of gene flow	250
6.3.6 Comparison of Southern African distributions with those in other parts of the world	254
6.3.6.1 Alcohol dehydrogenase (<i>Adh</i>)	254
6.3.6.2 6-Phosphogluconate dehydrogenase (<i>Pgd</i>)	258
6.3.6.3 sn-Glycerol-3-phosphate dehydrogenase (<i>Gpdh</i>)	260
6.3.6.4 Other enzymes	261
6.4 Discussion	262
6.4.1 Microgeographic variation	262
6.4.2 Macrogeographic variation	264
6.4.2.1 Gene flow measurements	265
6.4.2.2 Sampling techniques	265
6.4.2.3 Modelling considerations	267
6.4.3 Origin of Southern African populations of <i>D. melanogaster</i>	269

	<u>Page</u>
CHAPTER 7. CLINAL PATTERNS AND OTHER CORRELATIONS..	273
7.1 Introduction	273
7.1.1 Alcohol dehydrogenase...	274
7.1.1.1 Ethanol tolerance...	274
7.1.1.2 Other alcohols...	277
7.1.1.3 Latitudinal clines and temperatures...	280
7.1.1.4 Rainfall ...	282
7.1.1.5 Non-environmental factors...	283
7.2 Materials and methods...	284
7.3 Results...	285
7.4 Discussion...	289
7.4.1 Dipeptidase-A...	290
7.4.2 Isocitrate dehydrogenase-NADP...	290
7.4.3 Alcohol dehydrogenase...	291
CHAPTER 8. SUMMARY AND GENERAL CONCLUSIONS.	295
8.1 P elements...	296
8.1.1 GD sterility and geographic distribution	296
8.1.2 s^w mutability...	297
8.1.3 Molecular analysis...	297
8.1.4 Somatic activity	298
8.2 Allozymes...	299
8.2.1 Allozyme heterogeneity...	299
8.2.2 Clines and correlations...	301
8.3 P elements and allozymes	302
Appendix A.	307
Appendix B.	312
Appendix C.	317
Appendix D.	319
Appendix E.	322
Chemicals.	324
Photography	326
Literature Cited.	327

LIST OF TABLES

		<u>Page</u>
Table 2.1	Collection sites, dates and number of isofemale lines tested for P-M gonadal dysgenesis... ..	37
Table 2.2	P-Q-M classification scheme.. ..	46
Table 2.3	The numbers of P, Q, M and intermediate strains obtained at each collection site. ...	69
Table 2.4	Strains showing intrastrain sterility greater than zero	73
Table 2.5	Mean GD sterilities obtained in crosses A and A* for 21 collection sites in Southern Africa	77
Table 2.6	A Species of the subgenus <i>Sophophora</i> containing P homologous sequences... ..	82
	B Species of the subgenus <i>Sophophora</i> which did not show positive hybridization with P element probes... ..	83
Table 2.7	Comparisons between 1987 and 1988 proportions of P, Q and M type strains... ..	102
Table 3.1	Southern African strains which showed levels of <i>an^w</i> mutability greater than zero and which tested as M or M-Q in GD sterility tests. ...	119
Table 4.1	The expected and estimated fragment sizes obtained from the <i>Acc I</i> digestion of p _{25.1} ...	143
Table 4.2	Molecular characteristics of 31 wild South African strains.. ..	149
Table 5.1	A Results of GD sterility tests performed on the stocks <i>f³/Ser</i> , <i>msh</i> and <i>nmsh</i>	176
	B Verification of the (dysgenic or nondysgenic) state of the hybrid flies	176

Table 5.2	Number of wings of flies reared at 25°C having 0, 1, 2 or ≥3 spots per wing.. ...	178
Table 5.3	Number of wings of flies reared at 29°C having 0, 1, 2 or ≥3 spots per wing.. ...	179
Table 6.1	Locations of samples, collection dates and numbers of isofemale lines of <i>D. melanogaster</i> for each location	197
Table 6.2	Enzymes electrophoretically examined, composition of staining solutions and buffer systems used.. ...	200
Table 6.3	Alleles observed at all loci, their composite relative frequencies, structures and genetic map positions	216
Table 6.4	Number of alleles and number of samples polymorphic for each of eight loci examined in Southern African populations.. ...	224
Table 6.5	Number of loci polymorphic in each of the Southern African samples.	225
Table 6.6	Nested contingency-table analysis of allele frequencies in Southern African populations of <i>D. melanogaster</i>	237
Table 6.7	Approximate F ratios.	241
Table 6.8	Estimates of heterozygosity and gene differentiation coefficients for Southern African populations of <i>D. melanogaster</i>	244
Table 6.9	Absolute gene diversity estimates at the region, area, district and collection site levels... ..	245
Table 6.10	Comparison between G_{ST} values obtained in Southern African populations of <i>D. melanogaster</i> and F_{ST} values obtained for <i>D. melanogaster</i> populations in other parts of the world	255

LIST OF FIGURES

		Page
Fig. 2.1	A Scanning electron microscopy photograph of a normal ovary from an adult raised at 29°C. ...	41
	B Pairs of normal ovaries from adults raised at 29°C. ...	42
	C A pair of bilaterally dysgenic ovaries from an adult raised at 29°C. ...	43
Fig. 2.2	Illustration of the P-Q-M classification system using joint confidence probabilities...	47
Fig. 2.3	The distributions of GD sterility frequencies:	49
	A: Swakopmund. ...	50
	B: Gobabeb. ...	51
	C: Ai Ais. ...	52
	D: Cape Town, Gordons Bay. ...	53
	E: Mettenberg Bay. ...	54
	F: Port Elizabeth. ...	55
	G: Maclear. ...	56
	H: Durban. ...	57
	I: Johannesburg. ...	58
	J: Pretoria. ...	59
	K: Hammanskraal. ...	60
	L: Nelspruit. ...	61
	M: Pietersburg. ...	62
	N: Agatha. ...	63
	O: Thohoyandou. ...	64
	P: Harare. ...	65
Fig. 2.4	Distribution of <i>D. melanogaster</i> populations sampled in Southern Africa ...	76
Fig. 3.1	A Scanning electron microscopy photographs of thoracic bristles	
	(i) sn^+ male ...	105
	(ii) sn^W male ...	106
	(iii) sn^e male ...	107
	B (i) Lateral view of sn^+ males. ...	108
	(ii) Lateral view of sn^W males. ...	109
	(iii) Lateral view of sn^e males. ...	110
Fig. 3.2	Relationship between cross A GD sterility and sn^W mutability ...	117

	<u>Page</u>
Fig. 4.1	Examples of <i>Drosophila</i> DNA extractions ... 129
Fig. 4.2	Estimation of the concentration of plasmid DNA using ethidium bromide fluorescent quantitation methods... 133
Fig. 4.3	Partial restriction map of the plasmid p π 25.1. 138
Fig. 4.4	The plasmid p π 25.1 digested with <i>Acc</i> 1 on a 0.65% low melting point agarose gel... 140
Fig. 4.5	Standard curve of log molecular weight versus distance migrated for the λ DNA molecular weight marker III. ... 142
Fig. 4.6	Responses to hybridization of genomic <i>Drosophila</i> DNA with the 2.4 kb <i>Acc</i> 1 fragment of p π 25.1 ... 144
Fig. 4.7	Structure of the 2.9 kb P element in p π 25.1... 153
Fig. 4.8	Relationship between intensity of hybridization signal and cross A* percentage GD sterility for M strains ... 157
Fig. 5.1	Genetic development which could lead to formation of mutant spots. ... 167
Fig. 5.2	Half mesothorax showing the regions of the wing surface scored for spots. ... 171
Fig. 5.3	Scanning electron micrographs of wing blade trichomes
	A: Normal trichome. ... 172
	B: <i>msh</i> trichome ... 173
	C: <i>flr</i> ³ trichome... 173
	D: Multi-celled <i>msh</i> spot against a normal background.. 174
	E: Multi-celled <i>flr</i> ³ spot against a normal background.. 174
Fig. 5.4	Distribution of clones obtained at 29°C according to size classes. ... 182
Fig. 6.1	Map of Southern Africa showing locations of <i>D. melanogaster</i> collection sites... 196
Fig. 6.2	Hierarchical arrangement of <i>Drosophila</i> collection sites.. 203
Fig. 6.3	Hierarchical grouping as used in the gene diversity analysis ... 207

Fig. 6.4 A	Banding patterns obtained for the allele of <i>Adh</i> ...	211
B	(i) Electrophoretic variation of <i>Adh</i> ...	213
	(ii) Electrophoretic variation of <i>Gpdh</i> ...	214
	(iii) Electrophoretic variation of <i>Idh-NADP</i> ...	215
Fig. 6.5	Allele frequencies and approximate 95% confidence intervals for the most common allele of each of eight enzyme loci...	227
	A: <i>Adh</i> ¹⁰⁰ ...	228
	B: <i>Dip-A</i> ¹⁰⁰ ...	229
	C: <i>Gpdh</i> ¹⁰⁰ ...	230
	D: <i>Hes-1</i> ¹⁰⁰ ...	231
	E: <i>Hes-2</i> ¹⁰⁰ ...	232
	F: <i>Idh-NADP</i> ¹⁰⁰ ...	233
	G: <i>Pgd</i> ¹⁰⁰ ...	234
	H: <i>Pgm</i> ¹⁰⁰ ...	235
Fig. 6.6	Illustration of the proportional contribution to total gene diversity due to population subdivision...	249
Fig. 6.7	Conditional average frequency plotted against the occupancy rate...	252
Fig. 7.1	Variation in the frequency of <i>Dip-A</i> ¹⁰⁰ with longitude...	286
Fig. 7.2	Variation in the frequency of <i>Idh-NADP</i> ¹⁰⁰ with extreme minimum temperature...	287
Fig. 7.3	Variation in the frequency of <i>Adh</i> ¹⁰⁰ with average monthly rainfall...	288

CHAPTER ONE

1. Background and Motivation

1.1 Elements

1.1.1 Introduction

Transposable genetic elements or transposons, are segments of DNA which are capable of moving from one position to another within the genome of an organism. Most transposons have the ability to replicate themselves while moving to a new location and, as a result, one copy of the transposon remains behind in its original position. The copy number is thus increased. The rate of transposition and replication is apparently self-regulated and, in general, decreases as the copy number increases.

Transposable elements were first identified in plants during the 1940s and early 1950s by the geneticist Barbara McClintock (1951) who challenged the hitherto widely held concept of static genomic organisation. She demonstrated the existence of elements in the maize genome which were capable of jumping from one genetic locus to another and, in so doing, these elements were able either to disrupt the function of the target gene, to alter the expression of an adjacent gene or to cause chromosome breakage. Subsequently transposable elements have been identified in plants other than maize and, in particular, their presence in *Antirrhinum majus* (snapdragon) has been demonstrated (for reviews on maize and other plants see McClintock 1965; Fedoroff 1983, 1984; Nevers *et al.* 1986).

In the 1960s and 1970s parallels to the maize transposable elements were recognized in the insertion sequences of *Escherichia coli* and the bacteriophage Mu. Subsequently a wealth of information on prokaryotic transposons such as Tn 3 and Tn 10, as well as on eukaryotic transposons such as those that occur in yeast, maize and *Drosophila* has accumulated. Certain common features are present in both prokaryotic and eukaryotic transposable elements. Firstly, the ends of transposons, which are believed to serve as recognition sequences for the transposition enzymes, are highly conserved. They are usually characterized by the presence of repeat structures. These may be inverted repeats or direct repeats.

Among the bacterial transposons, inverted repeats occur in the insertion sequence (IS) and composite transposons as well as in the Tn 3 family of transposons. They range from 8 base pairs (bp) to 40 bp in length and may show perfect homology (38/38 for Tn3) or imperfect homology (12/18 for IS 50) (Grindley and Reed 1985). Inverted repeat sequences have also been found in eukaryotes such as in *Zea mays* (maize) in *Spm-18* as well as in members of the *Ac-Ds* system, in *Glycine max* (soybean) in *Tgm-1*, and in *A. majus* in *Tam-1*, *Tam-2* and *Tam-3* (Coen *et al.* 1986; Nevers *et al.* 1986). The lengths vary from 11 bp to 14 bp. Inverted repeats are also present in the P elements of *Drosophila* and also in the foldback (FB) elements of this genus (Rubin 1983).

Retrovirus like transposable elements which are found both in eukaryotes and prokaryotes contain long terminal direct repeats which are several hundred base pairs in length and usually contain short inverted terminal repeats within them. Examples of these are: Tn 9 in bacteria, Ty 1 in yeast, *copia*-like elements of *Drosophila* and the *ein-1* elements of *Zea mays* (Varmus 1983; Baltimore 1985; Nevers *et al.* 1986).

Whereas the terminal sequences of transposable elements appear to be highly conserved, the internal sequences often manifest a great deal of variation in both size and function. For example ISs generally contain a single transposase gene which ranges in size from 750 bp to 1600 bp (Grindley and Reed 1985). Other bacterial transposons may contain many genes, for example Tn 3 contains genes for transposase (*Tnp A*), cointegrate resolvase (*Tnp R*) and ampicillin resistance (β -lactamase) and is 4.9 kb in length. A eukaryotic transposon which closely resembles Tn 3 is the *Ac* element in maize in that it contains a long and a short open reading frame (ORF), oriented so that they are transcribed in opposite directions just like the two Tn 3 genes and there is evidence that the larger of the two *Ac* genes encodes a protein analogous to the transposase of Tn 3 (see Fedoroff 1984).

Finally, a characteristic property of the insertion process for nearly all transposons is the duplication of a short base sequence of target host DNA. This is thought to be brought about by two staggered single stranded nicks which are produced

at either end of the target sequence by an enzyme encoded by the transposon. The subsequent insertion of the transposon and filling in of the single stranded gaps results in the formation of two copies of the target sequence flanking the transposon. The length of the target sequence is variable. For example it is 3 bp long for the *Tgm-1* element from soybean, the *Tom-1* and *Tom-2* element from *A. majus* and the *Spm-18* receptor element from maize (Nevers *et al.* 1986); it is 5 bp long for *Tn 3*, bacteriophage Mu (Grindley and Reed 1985) the Ty elements in yeast (Roeder and Fink 1983), and the *copia* element in *Drosophila* (Varmus 1983) and 8 bp long for the P elements in *Drosophila* and the *Ac* and *De* elements in maize (Fedoroff 1984).

Transposition may be either replicative, as in the case of the *Tn 3* family (Grindley and Reed 1985) or conservative, as is hypothesized to be the case for *Tn 10* (Bender and Kleckner 1986). The retrovirus-like transposons are believed to transpose via the formation of an RNA intermediate and reverse transcription (see review by Baltimore 1985). However for most eukaryotic elements the transposition mechanisms are still largely unknown.

1.1.1.1 Mobile genetic elements in *Drosophila*

Mobile genetic elements can produce a remarkable range of phenotypic variations in both plants and animals, and were first noted and studied because of this. Later molecular studies of dispersed moderately repetitive DNA sequences were initiated by cloning such elements from *D. melanogaster*

(Finnegan *et al.* 1978; Ilyin *et al.* 1978). The mobility of these sequences became apparent when, as a result of *in situ* hybridization techniques, these sequences were found to lie on different regions of chromosomes; this was particularly true when different "stocks" of *D. melanogaster* were compared (Ilyin *et al.* 1978; Strobel *et al.* 1979). Many different families were subsequently investigated, both within *D. melanogaster* and within some of its sibling species (Young 1979; Potter *et al.* 1980; Rubin *et al.* 1981, 1982; Tchurikov *et al.* 1980; Truett *et al.* 1981; Young and Schwartz 1981; Bingham *et al.* 1982; Dowsett 1983; Martin *et al.* 1983; Bucheton *et al.* 1984).

About 12% of the *Drosophila* genome consists of moderately repetitive DNA sequences most of which are dispersed throughout the genome (for a review see Spradling and Rubin 1981). Many of these DNA sequences have properties consistent with their being mobile elements. Rubin (1983) grouped moderately repetitive sequences into four classes based primarily on their internal structures. Firstly the most variable class is that consisting of ribosomal gene insertions (namely the "type I and type II insertions found in the 28S rDNA genes), and scrambled clustered arrays of moderately repetitive DNA sequences. These are heterogeneous in structure and their mobility is not certain. The second class consists of *copia*-like elements. These range in size from 5-10 kb. Elements from each family carry long direct terminal repeats and are represented by 10-100 copies, scattered throughout the chromosomes. Elements

in this family are the mobile dispersed genetic elements (*mdg* elements), 412, 297, B104 and *gypsy*. The third class consists of foldback (FB) elements. These are long inverted repeats which may or may not flank a central non-repeating region. They are heterogeneous in size and display a certain amount of homology to one another. Finally Rubin (1983) described the P elements. These are about 2.9 kb long and possess short inverted 31 bp terminal repeats which are important for transposition. These will be fully described later.

1.1.1.2. Systems of hybrid dysgenesis

Possibly Rubin's (1983) class of P elements could have included I elements and *hobo* since all three of these are capable of causing hybrid dysgenesis. Nonreciprocal dysgenic traits were noted in 1975 when Kidwell and Kidwell (1975a,b) reported large differences in male crossing over and F_1 sterility between reciprocal crosses of strains derived from natural populations and commonly used laboratory stocks. This they attributed to some kind of cytoplasm chromosome interaction. They found that F_1 fertility was drastically reduced and male crossing over was significantly increased when wild males and laboratory stock females were used as parents but not in the reciprocal cross. The term "hybrid dysgenesis" (Sved 1976; Kidwell and Kidwell 1976; Kidwell *et al.* 1977b) is used to describe these phenomena and is defined as "a syndrome of correlated genetic traits that is spontaneously induced in hybrids between certain mutually interacting strains, usually in one direction only" (Kidwell and Kidwell 1976).

Initially two distinct systems of hybrid dysgenesis were recognized, the P-M system and the I-R system (reviewed by Bregliano and Kidwell (1983). In the P-M system strains are partitioned into two types designated P and M according to whether they produced dysgenesis when used as either the paternal (P) or maternal (M) parent. Dysgenesis is brought on when an M strain female is crossed with a P strain male. The reciprocal cross yields normal progeny. For any particular dysgenic trait this dichotomous classification system is absolute in that hybrids from all P x P and M x M crosses do not show this trait (Kidwell *et al.* 1977b). A third type of strain, designated Q (Kidwell 1982), is virtually indistinguishable from the P strain in all properties except one - the progeny of Q males crossed to M females are fully fertile.

In a series of experiments in which hybrids were tested for sterility Engels (1979b) interpreted the cause of hybrid dysgenesis as the interaction of two components, a paternally derived polygenic chromosomal factor or factors and a maternal component designated "cytotype" i.e. a property of the entire cell. Thus a combination of P element chromosomes and M cytotype produces dysgenesis whereas these P chromosomes plus P cytotype results in a stable condition. A detailed genetic analysis of cytotype determination was provided by both Engels (1979b) and Kidwell (1981).

The type of sterility which is manifested in dysgenic hybrids in the P-M system is due to a gonadal dysgenesis (GD sterility) in both males and females (Engels and Preston 1979; Schaefer *et al.* 1979; Kidwell and Novy 1979). No eggs are laid by the sterile females, and this GD sterility is dependent on a high temperature (24°-29°) during embryonic and larval stages of development. The I-R system, on the other hand, is responsible for an altogether different type of sterility called "stérilité femelle" (SF) sterility (Picard 1971). Here males are not affected and eggs are laid by the (SF) female progeny of R (Reactive) strain females and I (Inducer) strain males but these eggs show a varying degree of reduced hatchability and the embryos die during late cleavage. Further, there is a period during oogenesis, one or two days before oviposition, when increasing the temperature has a curative effect on SF sterility (Bucheton 1978).

Associated dysgenic traits in the I-R system include X-linked visible mutation and non-disjunction (Picard *et al.* 1978). The I-R system also produces a neutral (N) strain, analogous to the Q strain. Further, the stable condition of the P cytotype is echoed in the stability (in the I-R system) of the inducer state for I elements (Bucheton and Picard 1978). Despite these similarities, the two systems are independently distributed (Kidwell 1979). Thus strains can be doubly classified as IP, RM etc. Cloning and sequencing of the I factor revealed a 5.4 kb sequence with highly conserved ends but lacking repeats and having no sequence homology with the P factor (Bucheton *et al.* 1984; Fawcett *et al.* 1986).

An element called *hobo*, having several features in common with the P element but lacking homology to it, was characterized by McGinnies *et al.* (1983) and Strack *et al.* (1986). The presumed complete *hobo* element is 3016 bp long and has a 12 bp inverted terminal repeat. The *hobo* element like the P element, produces an 8 bp target site duplication sequence upon insertion. Further, there are two types of strains differing at the molecular level. The H strains contain both complete and deleted *hobo* elements and the E strains are devoid of *hobo* elements. Genetic instabilities such as a temperature independent GD sterility and some somatic activity produced by the male recombination factor, 23.5 MRF (Stamatis *et al.* 1981; Yannopoulos *et al.* 1983), are thought to be caused by the *hobo* elements (Yannopoulos *et al.* 1987). Also, the *hobo* element is responsible for high rates of chromosomal rearrangements (Blackman *et al.* 1987; Lim 1988).

The *hobo* system appears to be, both molecularly and genetically, more similar to the P system than to the I system, in the size of the elements, in the existence of terminal repeats and in the 8 bp duplication of target sequence. Matings of P strains to M strains and H strains to E strains produce a GD sterility in both males and females.

1.1.2 P-M Hybrid Dysgenesis - Historical Commentary

The following discussion which took place at the Cold Spring Harbor Symposium in 1941, is an indication of the degree of change that has taken place over the last 50 years:

BANTA: It seems remarkable that Dr Demerec found in *D. virilis* four unstable genes in a relatively short time and that such have not been found elsewhere. Have you looked for them in other species of *Drosophila*?

DEMEREK: I have looked for many years but have found none.
(M Demerec 1941).

1.1.2.1 Mutations

Unstable genes have the capacity of mutating in several directions, but usually mutate from a mutant recessive to a dominant wild-type. In 1928, Demerec (1928) reported the presence of an unstable gene, *reddish-alpha*, in *D. virilis*. Three other unstable genes, *miniature wings-8*, *magenta* eye colour and *miniature wings-5* (discussed in Demerec 1941) were also found in *D. virilis*. At the 1936 summer meeting of the Genetics Society of America, Demerec (1937) gave an account of an unusually high frequency of spontaneous lethals in an inbred Florida stock of *D. melanogaster*. At this same meeting, Plough and Holthausen (1937) reported cases of high mutability in the progeny of interstrain crosses using the

Florida stock. Subsequently Sturtevant (1939), Neel (1942) and Ives (1945, 1950) reported high mutation rates in some of the progeny of certain interstrain hybrids, in *D. pseudoobscura* and *D. melanogaster* respectively.

The maize geneticists were also aware of the potential for change in certain genes. In 1914, Emerson (1914) suggested that genetic instability could be responsible for pericarp variation in maize. Then, in the 1930s, Rhoades found that genetic instability in maize could be conditional on the presence of a particular gene, the *Dotted* (*Dt*) locus (Rhoades 1938, 1941).

It was McClintock who, in 1947, while studying the properties of broken chromosomes in maize, first realized that a genetic element can transpose (McClintock 1947). The next decade saw the beginning of an extensive investigation into the controlling element families in the maize genome, the *Ac-Ds*, *Spm*, and *Dt* families (McClintock 1950, 1951, 1953, 1955; Brink and Milan 1952; Peterson 1953; Nuffer 1954).

Several researchers visualized the impact of transposable elements in *Drosophila*. Green (1967, 1969a) studied the high mutation rates of the gene *white-crimson* (*w^c*) and concluded that this high mutation rate was due to a "controlling element" which was integrated at a particular site in the *white* gene.

Further, Green (1969b) showed that this "controlling element" was capable of spontaneous transposition from its position on the X chromosome to a new location on an autosome where it integrated. A series of papers by Minamori and co-workers described high mutational events in *D. melanogaster* and ascribed them to an extrachromosomal element called delta (see for example Minamori and Ito 1971). The integration of controlling elements at several loci as a possible cause of high rates of mutation was suggested by Kidwell *et al.* (1973).

Crow (1988) considers "the first unequivocal observation of hybrid dysgenesis due to P elements" to have been made by Hiraizumi (1971) who reported the phenomenon of spontaneous male crossing over in *D. melanogaster*. This occurred in the germ line of the hybrid progeny of males from a wild Texas population and females from laboratory stocks. Soon after, worldwide reports of male recombination began to appear. Areas where this took place included Northwestern Ohio (Waddle and Oster 1974), Australia (Sved 1974), Jacksonville, Florida (Voelker 1974), Southern Texas (Matthews and Hiraizumi 1976), Yugoslavia, Taiwan and Japan (Yamaguchi 1976), Greece (Yannopoulos and Pelecanos 1977) and Cambridge, England and Oklahoma, USA (Woodruff and Thompson 1977). Chromosome breakage was often found to be associated with male recombination (Voelker 1974; Yamaguchi and Mukai 1974; Yamaguchi 1976) and a few authors suggested that male recombination was caused by episomes

or viruses (Waddle and Oster 1974; Voelker 1974). Later indirect evidence that male recombination was due to transposable elements gradually accumulated. Slatko (1978) reported that new male recombination activity occurred on third chromosomes of a marker strain which were originally free of the ability to induce male recombination after being associated with a second chromosome exhibiting a high male recombination potential.

Matthews *et al.* (1978) demonstrated the presence of new male recombination (MR) elements in certain chromosomal regions after homologous association with a region known to carry MR elements. Yannopoulos (1979) provided evidence for successive transposition of MR factors from one autosome to another. The data suggested that the sites of integration were not completely random.

1.1.2.2 Insertion mutants and hybrid dysgenesis

The years 1976 to 1977 were watershed years when the possibility of a common source for the phenomena of mutation, male recombination, chromosome breakage and sterility was envisaged. Occurrences, such as the reciprocal cross effects associated with male recombination and F_1 female sterility (Kidwell and Kidwell 1975a; Sved 1976; Kidwell *et al.* 1977a; Woodruff and Thompson 1977) and the correlation between high spontaneous frequencies of recessive sex linked lethals and female sterility (Kidwell *et al.* 1977a), led to the unifying concept of "hybrid dysgenesis" (Sved 1976; Kidwell and Kidwell 1976; Kidwell *et al.* 1977b) as described in section 1.1.1.2. Subsequently many studies of these phenomena were made.

A case for insertion mutants was made by Golubovsky *et al.* (1977) when they examined a series of 11 independent mutants at the sex-linked *singed* (*sn*) bristle locus of *D. melanogaster*. The *sn* mutants all stemmed from wild males collected at Tashkent (1973) and Krasnodar (1974), in the U.S.S.R. On the basis of the diagnostic genetic criterion of an inordinately high mutation rate Golubovsky *et al.* (1977) presumed the mutations to be due to insertions and inferred that "foreign" DNA - something analogous to IS DNA or phage DNA such as Mu was inserted.

Then Green (1977a) isolated a male recombination (MR) chromosome from wild flies trapped on Mount Carmel, Haifa, Israel and used these chromosomes to generate mutations on the X chromosome. Some of these male flies induced the putative *sn* insertion mutation. Further, MR significantly increased the frequency of mutation not only at the *sn* locus but at the *rasberry* (*rxa*) and *yellow* (*y*) loci (two other sex-linked mutable loci). Green suggested that the majority of MR generated mutations were insertion mutations. Insertions were directly identified by Simmons and Lim (1980) using cytological methods in a screen for mutations in the *zeate-white* region of the X chromosome of P-M dysgenic hybrids.

Engels (1979a, 1981b) investigated an unstable allele *sn^W* (weakly *singed*) at the *singed* bristle locus and demonstrated that the instability was under the same extrachromosomal control as hybrid dysgenesis i.e. the mutability of *sn^W* was

reversibly suppressed when placed in the background cytotype known to prevent sterility (i.e. P cytotype). Engels¹⁷ (1979a) suggested that agents causing hybrid dysgenesis be regarded as "a family of mobile genes ("P genes") with similar DNA sequences scattered throughout the genomes of the P strains". He further suggested that "These genes may become activated and perhaps mobilized in certain cells, particularly germ cells, that possess the M cytotype. The *sn^M* allele might then represent one such gene, which, by virtue of its position, yields various observable effects on bristles when it undergoes its usual extrachromosomally controlled changes of state".

Indirect evidence of the transposition of P factors was provided by Raymond and Simmons (1981) who found a statistical coincidence in P-M dysgenic hybrids between the phenomena of *sn^M* instability when in the M cytotype and X-linked lethal mutation induction. They interpreted this observation by postulating that "mutations at the *singed* locus sometimes release a genetic element capable of reinserting elsewhere in the chromosome" - resulting in lethality. Evidence that P element insertion was not entirely arbitrary, but took place preferentially at certain chromosomal "hot-spots" also gradually accumulated (Berg *et al.* 1980; Engels and Preston 1981).

1.1.3 Physiological and Molecular Characteristics of the P-M System

Investigations into the nature of GD sterility, male recombination and mutational events have revealed many of the processes and mechanisms responsible for these phenomena. There is general agreement as to the nature of the cell line involved, the necessary cytoplasmic background as well as the timing of these occurrences. Apart from this study (Chapter 5) and a few others (see references in Chapter 5) most investigations have shown hybrid dysgenesis to be restricted to the germ line. Further, dysgenic symptoms require an M cytotype background before they can occur and are suppressed in the P cytotype (Engels 1979b). Finally as regards timing, the occurrence of clustering indicates the pre-meiotic nature of hybrid dysgenesis. For example clustering shows that male recombination generally occurs in the mitotically dividing premeiotic germ cell lines (Hiraizumi *et al.* 1973; Kidwell and Kidwell 1976; Woodruff and Thompson 1977; Hiraizumi 1979; Isackson *et al.* 1981; Sinclair and Brigliatti 1985). Clustering of mutants also indicated the premeiotic occurrence of sn^W mutability (Engels 1979a).

Temperature experiments have also been used to ascertain the timing of dysgenic events. The temperature sensitive period of GD sterility begins in the late embryonic stages and continues through the first and second instars thus indicating that GD sterility is being brought about by the failure in the early development of germ cells which may be taking place at the onset of mitosis (Engels and Preston 1979; Kidwell

and Novy 1979). Another method that was later used to investigate the time of GD sterility was microscopy. Niki and Chigusa (1986) analysed GD sterility using light and electron microscopes. They found that the first appearance of an abnormality at 29°C occurred during migration of the primordial germ cells to the gonadal mesoderm at five to six hours. Further necrosis of these germ cells was prominent in the gonads during the first instar when these cells undergo mitotic proliferation in normal larvae.

1.1.3.1 Gonadal dysgenesis

Female sterility is often selected as an indication of hybrid dysgenesis, being more amenable to study in that it can be observed directly in the dysgenic hybrids themselves rather than in their offspring and it affects a larger proportion of hybrids than any other trait. Although sterility occurs in hybrids of both sexes, the frequency observed in females usually exceeds that of the males (Kidwell and Kidwell 1975a). Developmental temperature tends to be a critical factor in the induction of sterility. Both male and female sterility reach a maximum at a temperature of 29°C with the temperature sensitive phase occurring in a period extending from late embryonic development to the second larval instar (see above). GD female sterility was shown to be due to the inability of ovaries to develop normally — the blockage occurring at an early stage of ovarian development and resulting in either unilateral or, more commonly bilateral rudimentary ovaries

(Schaefer *et al.* 1979). Thus the term "gonadal dysgenesis" was used to describe the type of female sterility associated with the P-M system (Kidwell and Novy 1979).

1.1.3.2 Molecular characteristics

Molecular cloning of P elements has been achieved and the transposon hypothesis confirmed by Rubin *et al.* (1982). They showed that mutant *white* alleles contained additional DNA sequences. Since these mutations were induced by P-M hybrid dysgenesis, the additional DNA sequences were called P elements.

P elements were found by molecular hybridisation studies to be present in 30-50 copies per haploid genome in most P strains (Bingham *et al.* 1982). Further, Bingham *et al.* (1982) showed that members of the P family transposed frequently in P-M dysgenic hybrids and that chromosomes descended from these hybrids often acquired P elements. These authors proposed that the heterogeneity observed in the size and structure of P elements corresponded to a heterogeneity in their genetic potential in that, some copies of the P element retained, while other copies lacked, certain properties such as the capacity to express functions necessary for transposition and the capacity to induce sterility. This concept, they used to explain the workings of Q strains by postulating that Q strains lacked copies of P elements which possessed the functional genetic factors responsible for hybrid sterility,

but retained other defective P elements. They also used this to explain M' strains, i.e. strains which have M cytotype but contain P elements. Their approach was far sighted in view of recent research on deletion derivatives of P elements.

Several members of the P family of transposable elements were isolated and characterized by O'Hare and Rubin (1983). who showed that this family of transposable elements ranged in size from less than 500 bp up to a 2.9 kb autonomous P element. All naturally occurring P elements were shown to have 31 bp perfect inverted terminal repeats and to duplicate an 8 bp host sequence upon insertion. O'Hare and Rubin (1983) suggested that the small elements had arisen by deletion of internal sequences of the 2.9 kb elements, while retaining the 31 bp inverted terminal repeats. These authors showed further that P elements can excise precisely during hybrid dysgenesis in that the P element plus one copy of the 8 bp duplication is lost in revertants and thus the original DNA sequence is restored. They suggested that this type of excision is accomplished by reversing the first steps of the insertion mechanism as follows: staggered cuts are made at both copies of the 8 bp sequence bounding the P element. The excised P element and the original insertion site are now both left with 8 bp "sticky ends" that could be ligated to restore the original sequence of the insertion site and to form a circular P element containing one copy of the 8 bp host DNA. Such circular structures could be the extrachromosomal elements which Engels (1981a) has suggested may play a role in the determination of cytotype (see section 2.4.3.1).

Excision and transposition of P elements is effected by a function (transposase activity) supplied by the 2.9 kb P element (Spradling and Rubin 1982) which can act in trans (Engels 1984) and is thus able to mobilize smaller defective non-autonomous P elements (Rubin and Spradling 1982). P element mobilization may also be aided by a host-encoded protein (Rio and Rubin 1988). The 2.9 kb P element contains four ORFs (O'Hare and Rubin 1983), all four of which are necessary for transposase activity (Karess and Rubin 1984).

Finally, the high mobility of P elements ensured their usefulness as vectors for gene transfer, and P elements were the first eukaryotic transposable elements to be used as such. This occurred in 1982. A plasmid containing the 3 kb P element was microinjected into *Drosophila* embryos and the P elements were found to transpose frequently from plasmid sequences into the chromosomes of developing germ line cells (Spradling and Rubin 1982). Subsequently, a chromosomal DNA fragment containing the *rosy* gene was inserted into a P transposable element and the resulting transposon was used to transform germ line *Drosophila* cells (Rubin and Spradling 1982). P transposons have subsequently been used to analyse many P element functions in *Drosophila* (Karess and Rubin 1984; Daniels *et al.* 1987a) and P element mediated transformation (fully described in Spradling 1986) has obvious enormous potential in the field of genetic engineering. The molecular characteristics of P elements are further discussed more fully in Chapters 3, 4 and 5.

1.1.4 Motivation for the Project

The evolutionary significance of mobile genetic elements is not yet fully understood. Their actions as mutagens by inserting into or in juxtaposition with a gene and thus altering its expression must have important genetic consequences. Indirect evidence showing that most of these mutations are deleterious to the organism (Golding *et al.* 1986) suggests that selection against individuals harbouring these elements would take place. On the other hand, theories suggesting that transposable elements are selected for, either at the population level due to their ability to promote genetic diversity, or at the level of the transposon itself (the "selfish gene" theory) have been propounded (see Syvanen (1984) for a review, Engels (1986) for an overview). It is thought that some transposable elements (such as P elements in *D. melanogaster*) may be directly responsible for causing reproductive isolation and hence be instrumental in speciation (Bingham *et al.* 1982; Ginzburg *et al.* 1984). Finally, the rapid spread of transposable elements within a population has led to several theories regarding the population dynamics of these elements (see Brookfield (1986) for an overview).

The empirical testing of these theories, together with clearer insight into the biological role played by transposons, has been aided by an enormous body of investigative practical

work which has taken place in the past decade. In particular, the study of P elements in the fruit fly *D. melanogaster* has provided and continues to provide a rich source of genetic information.

Studies of *D. melanogaster* populations have taken place in many parts of the world. Lines collected before the 1950s in America and before the 1960s in Europe, Asia and Africa and kept in laboratories since then, have nearly all proved to be true M type, completely lacking P elements (Anxolabéhère *et al.* 1988b). P and Q type strains began to appear in flies collected after 1950 in America and then subsequently in Europe and elsewhere in the world (Anxolabéhère *et al.* 1982a, 1984, 1985b, 1988b; Kidwell *et al.* 1983; Yamamoto *et al.* 1984; Kidwell and Novy 1985; Boussy and Kidwell 1987; Woodruff *et al.* 1989). Today there appears to be a worldwide clinal pattern in the distribution of P, Q, M and M' strains. This distribution pattern has aided in both forming and testing hypotheses regarding the origin and the population dynamics of P elements (see, for example, Engels 1981a; Kidwell 1979, 1983; Anxolabéhère *et al.* 1988b). Southern Africa is one of the few large single land areas which has been omitted from any intensive studies of P elements in *D. melanogaster* up to now. However, the distribution of P strains has been characterized in both North and Central Africa (Anxolabéhère *et al.* 1982a, 1984; Woodruff *et al.* 1989). Thus in order to complete the survey of the African continent I have investigated the geographical distribution of P element strains

collected from 21 different sites in Southern Africa, mainly during the summers of 1987 and 1988. I have used female GD sterility as a means of strain classification.

The measure of sn^W mutability has proved to be a sensitive assay for P factor activity (Engels 1979a, 1984) and also to correlate well with a strain's ability to induce GD sterility (Engels 1984). I have measured the sn^W mutation rate through the male germline in 78 out of the 206 strains collected in order to assay for P factor activity and to test for a correlation between sn^W mutability and GD sterility levels.

The use of molecular techniques has revealed the presence of defective P elements in all recently captured wild populations which tested phenotypically as M. This is a worldwide phenomenon. In order to ascertain whether this is the case in the Southern Africa populations, I have used the *Acc* I fragment of the plasmid p π 25.1 (O'Hare and Rubin 1983) kindly supplied by W.R. Engels to probe for P elements in a subset of the 1987 collection.

Finally, most evidence to date points to P-M hybrid dysgenesis being restricted to the germ line. On the molecular level, Laski *et al.* (1986) showed that this germ line specificity of P element transposition was due to differential mRNA splicing in germ line and somatic tissue. However, there are certain indications that small amounts of P induced somatic activity may be taking place. For example, Engels (1979a) obtained 11 somatic sn^W mosaics out of a total of 8027 flies tested.

I have consequently made use of a P-line carrying a wing-hair marker constructed by one of my supervisors (N van Schaik) to test for P element somatic activity by examining the frequency of mutant clones on the wings of dysgenic hybrids.

1.2. Allozymes

1.2.1 Introduction

The study of the distribution of P elements in Southern African populations is aimed at examining the variation that is present in a non-Mendelian system. In order to complement this investigation a parallel study of genetic variability in allozyme systems was undertaken. Any connection between the presence or absence of P elements in populations and the degree of variability within these allozyme systems could be investigated.

Theoretical and technological advances in the fields of population genetics and molecular biology have resolved many of the problems regarding the origin, amount and role of the genetic diversity inherent in all species. The Muller/Dobzhansky controversy regarding the presence or absence of genetic variation within natural species was apparently solved with the advent of electrophoretic techniques which made it possible for scientists to study polymorphic variation in the primary structure of proteins in both natural and laboratory populations of most organisms.

Pioneering work with *Drosophila* originated in the 1960s with studies of protein variation in *D. melanogaster* (Hubby 1963), *D. pseudoobscura* (Hubby and Lewontin 1966; Lewontin and Hubby 1966) and *D. ananassae* (Johnson *et al.* 1966). These empirical studies revealed the presence of a substantial amount of genetic variation in the majority of natural populations (see, for example, books by Lewontin 1974 and Nei 1975). Finally, the theoretical works of the population geneticists such as Haldane, Fisher and Wright, established the importance of the forces of migration, mutation, selection and drift in determining the content and subsequent evolutionary path of natural populations. Thus the combination of theoretical considerations and practical work on species such as *D. melanogaster* (which lend themselves to laboratory studies) together with the unity of hereditary mechanisms has led to a greater understanding of evolutionary forces in many other species.

1.2.2 Background for the Allozyme Study

Populations of *D. melanogaster* occur worldwide and have been found in a diversity of habitats. Since populations are very large and since there appear to be few barriers to gene flow because of the commensal nature of this species, populations might be expected to be genetically homogeneous over large areas. Studies of *D. melanogaster* populations have revealed that this is not always the case. On the contrary it appears that for particular enzymes a great deal of genetic heterogeneity exists among populations. This lack of homogeneity has been reported for groups of populations located in different regions,

primarily in temperate regions in the northern hemisphere including North America, Japan and Europe. Additional studies have been conducted in tropical and subtropical areas of Asia, in North, West and Central Africa and in Australia.

There are several examples of allozyme heterogeneity on a mesogeographic scale of 100 to 2000 kilometres. In the Eastern United States, considerable mesogeographic variation was observed by O'Brien and MacIntyre (1969) for the 6-Phosphogluconate dehydrogenase (*Pgd*) and larval Alkaline phosphatase (*Aph*) loci and weak variability for several other enzymes. Also in the Eastern United States, Johnson and Schaffer (1973) found clinal patterns exhibited in several enzyme systems including Esterase-C (*Est-C*), Esterase-6 (*Est-6*) and Octanol dehydrogenase (*Odh*), and Vigue and Johnson (1973) found that Alcohol dehydrogenase (*Adh*) also exhibited a clinal pattern in this area. In the Western United States, Voelker *et al.* (1977) found significant allele-frequency differences among populations for *Adh*, sn-Glycerol-3-phosphate dehydrogenase (*Gpdh*), Hexokinase-C (*Hex-C*), and Malate dehydrogenase (*Mdh*). In Japan, Watada *et al.* (1986) investigated allele-frequency differences for ten loci from the main island and from southern island populations, including Ogasawara and Ryukyu Islands. They found apparent geographical differentiation among *D. melanogaster* populations, not only between the main island and Ogasawara, but also among the 14 main-island populations. In France, Charles-Palabost *et al.* (1985) found considerable variation among fourteen populations stretching from Tostes

near Louviers in the north and Tautavel near Perpignan in the south. The enzymes showing a significant difference in allele frequencies were *Adh*, *Est-C*, *Est-6*, *Gpdh*, and Phosphoglucosmutase (*Pgm*).

Considerable genetic heterogeneity has also been observed on a macrogeographic scale of several thousand kilometres across continents. An investigation of six Afrotropical populations from the People's Republic of Bénin in the east to Burundi in the west, showed that there was a great degree of heterogeneity among these populations (David 1982). Investigations of the pattern of allele-frequencies in three zoogeographic zones : Australasia, North America and Europe/Asia revealed latitudinal variation in five enzyme systems (*Adh*, *Adh*, *Gpdh*, *Pgd* and *Est-6*) as well as geographic variation in Acid phosphatase (*Acp*) and *G6pd* (Oakeshott *et al.* 1981, 1982a, 1983a, 1983b). Finally, the investigations of Singh *et al.* (1982) and those of Singh and Rhomberg (1987a,b) who examined 26 and 117 loci, respectively over several continents, showed that, although there was considerable genetic variation among areas for some loci, the overall degree of differentiation among these widely-separated populations was still typical of differentiation among populations of the same species.

In order to comprehend fully the reasons for genetic differentiation it is necessary to obtain as much knowledge as possible about its nature by investigating natural populations of *D. melanogaster* in as many varied environments as possible. Up to the present, no large scale study has been conducted on Southern African populations of *D. melanogaster*. This was noted by authors Singh, Hickey and David as long ago as 1982 (Singh *et al.* 1982) at which time (to my knowledge) the first study of Afrotropical populations took place (David 1982). In an attempt to fill this gap I have investigated the patterns of gene frequency and their implications in 17 Southern African populations of *D. melanogaster* extending over 5 000 kilometres from Northern Namibia, through South Africa to Zimbabwe. A hierarchical structure consisting of four different levels of population division has been superimposed on these 17 subpopulations. The four levels are, in ascending order of spatial separation:

- Level 1 : sites within districts
- Level 2 : districts within areas
- Level 3 : areas within regions
- Level 4 : regions within Southern Africa.

I evaluated the amount of genetic differentiation between the subpopulations within each hierarchical level and between the levels as follows:

(1) The G statistic of Sokal and Rohlf (1981) was used in order to assess any significant differences between subpopulations within levels. Where significant heterogeneity precluded pooling, approximate F ratios were computed in order to compare the nested levels with each other.

(2) A gene diversity analysis was used to quantify the amount of heterozygosity that is lost due to population subdivision.

This study also made use of indirect methods to assess the amount of gene flow that has taken place in the past between Southern African populations of *D. melanogaster*. The population structure revealed in (1) and (2) above was analysed in the light of the gene flow results.

The environmental geographic and climatic conditions in Southern Africa are not repeated anywhere else in the world. Consequently this study of the distribution of allozyme-frequencies has afforded a unique opportunity to investigate clinal patterns that may shed some light on possible allozyme adaptations. I have thus examined the correlation between the frequency of the most common allele in five of the enzymes studied, with latitude and longitude. I have also looked at the correlation with the two climatic variables — rainfall and extremes of temperature.

Finally I have made use of this study to try to shed some light on the origins of the Southern African populations of *D. melanogaster*. *D. melanogaster* is one of eight species of the *melanogaster* subgroup of which six are endemic to the Afrotropical region (Lemeunier *et al.* 1986). This fact, among others, led to the hypothesis (now widely accepted) that *D. melanogaster* is of Afrotropical origin (Tsacas *et al.* 1981; Lachaise *et al.* 1987). It is not certain how long *D. melanogaster* has been present in Southern Africa. These populations could have originated due to a southward migration of the ancestral populations in Central and West Africa during the late Pleistocene or Holocene periods (Lachaise *et al.* 1987) or even earlier. Alternatively present day populations may have originated at a later time by dispersal from West and Central Africa or by colonization from Europe and Asia through human immigration (see David and Capy 1988). In this study I compared patterns of allele-frequency distribution among Southern African populations with those in other parts of the world particularly Africa, in order to obtain some insight into the origin of these populations.

CHAPTER TWO

2. Geographical Distribution of P elements in Southern Africa

2.1 Introduction

The P-M system of hybrid dysgenesis (Kidwell and Kidwell 1976; Sved 1976; Kidwell *et al.* 1977b) is activated when males of a P strain (paternally contributing) are mated with females of an M strain (maternally contributing) in that, the progeny of these matings manifest various dysgenic symptoms (discussed in Chapter 1). These symptoms are not apparent in the progeny of the reciprocal cross ($P \sigma \times M \sigma$), nor do they appear in the progeny of either $P \sigma \times P \sigma$ or $M \sigma \times M \sigma$ crosses. The P-M system of hybrid dysgenesis is interpreted as being due to the interaction of two components. Firstly, P strains possess many copies of a chromosomal genetic factor, called a P factor, which leads to hybrid dysgenesis and which is found on all major chromosomes (Engels 1979b; Engels and Preston 1980). Secondly a maternally derived component is responsible for the non-reciprocity of the dysgenic syndrome. The term 'cytotype' is used to designate the ability to transmit this maternal component; cytotype (which may be of type M or P) is determined by chromosomal factors but with limited cytoplasmic transmission (Engels 1979b; Kidwell 1981). Thus hybrid dysgenesis may result from the interaction of P factors and M cytotype being present together in the same zygote.

2.1.1 Classification of Wild *Drosophila melanogaster*

Studies of *Drosophila melanogaster* populations have taken place in many parts of the world. Over the years a variety of techniques have been developed for investigating the P-M composition of wild strains, these include GD sterility studies, *on^w* mutability tests and molecular hybridization experiments.

Unknown strains have been divided into four different categories, P, Q, M' and M. P strains (Kidwell *et al.* 1977b) possess the P cytotype (Engels 1979b) and have the potential to produce P factor activity in a suitable background cytotype. This P factor activity is thought to be closely related to transposase production. Q strains have P cytotype but a low level of P factor activity (Kidwell 1979). Although they may carry complete 2.9 kb P elements, they do not cause GD sterility when crossed to M strain females. M strains (Kidwell *et al.* 1977b) can be divided into two types, pseudo M and true M. M' (pseudo M) strains carry a variable number of P elements (Bingham *et al.* 1982; Kidwell 1985) but they do not induce hybrid dysgenesis when crossed to M strain females. These strains, however, have been shown to reduce levels of hybrid dysgenesis when crossed to P strain males (Kidwell 1985; Simmons and Bucholz 1985). This partial repression is distinct from P cytotype in that it is inherited biparentally (Kidwell 1985; Black *et al.* 1987) and it is thought to be brought about by deletion derivatives in

general and in particular by the KP element - a P factor internal deletion derivative cloned out of the Russian strain Krasnodar (Black *et al.* 1987). M (true M) strains possess M cytotype (Engels 1983) and have no homology to P element probes (Bingham *et al.* 1982).

Worldwide findings have revealed the following:

Lines collected before the 1950s in America and Japan and before the 1960s in Europe, Africa and the Middle East and kept in a laboratory since then were nearly all (true) M type (Kidwell 1983; Kidwell *et al.* 1983; Yamamoto *et al.* 1984; Anxolabéhère *et al.* 1988b). P and Q type strains began to appear in strains collected after 1950 in America and then subsequently in Europe.

The geographical distribution of strains collected after 1970 is as follows:

(a) P mainly, but sometimes Q strains are found in North and South America where the M cytotype is either absent or present only at very low frequencies (Engels and Preston 1980; Kidwell 1983; Kidwell *et al.* 1983; Kidwell and Novy 1985; Anxolabéhère *et al.* 1988b).

(b) In Europe the Q populations appear to dominate (particularly in France) although there are the occasional occurrences of P populations in places such as England, Belgium, Italy and France. P elements appear to decrease in frequency as one moves eastwards towards Asia, so that populations in Middle Asia are

all M cytotype but in Central Europe gonadal dysgenesis studies suggest a cytotype polymorphism (i.e. P/M cytotype) showing a mixing of the strains (Anxolabéhère *et al.* 1982a, 1985b; Kidwell *et al.* 1983). In Japan and the Far East, P or Q populations predominate, although some M strains are present (Ohishi *et al.* 1982; Kidwell 1983; Kidwell *et al.* 1983; Yamamoto *et al.* 1984).

(c) In Australia a north to south clinal pattern has been observed along the eastern seaboard. The northern populations from 16,9° S to 33,4° S are phenotypically P, populations from 34,4° S to 37,1° S are Q, while populations from 37,5° S to 43,2° S are M (Boussy and Kidwell 1987; Boussy 1987).

(d) On the African continent mainly the northern and central parts have been studied to date. In North Africa, M strains predominate, whereas Central sub-Saharan Africa is populated by P and Q strains (Anxolabéhère *et al.* 1982a, 1984; Woodruff *et al.* 1989).

In South Africa, six M type and one Q type strain were obtained from collections made prior to 1970 (Kidwell *et al.* 1983). Also two Southern African strains obtained in the 1970s tested as pure M lines (Daniels *et al.* 1987b). Thus, not much work appears to have been done on the geographic distribution of P elements in Southern African populations of *D. melanogaster*.

This study constitutes the first in depth survey of P-M hybrid dysgenesis in Southern Africa. Populations, apart from one line, were collected during the summers of 1987 and 1988 from



Fig. 2.1 C A pair of bilaterally dysgenic ovaries from an adult raised at 29°C. Magnification 35 x.

within the joint 95% confidence rectangle. For this reason, I have introduced four intermediate categories: P-Q, Q-P, M-Q and Q-M and three indefinite categories M?, Q? and P? in addition to the previous three categories P, Q and M. Table 2.2 and Fig. 2.2 illustrate this classification. The axes (Fig. 2.2) represent relative frequencies of GD sterility in crosses A and A* (Yamamoto *et al.* 1984). Each square on the diagram represents a joint confidence region. Thus the point with coordinates (P_A , P_{A^*}) - the estimated GD sterilities in crosses A and A* respectively - is the centre of the square which has sides of length $2S_A$ ($=2S_{A^*}$). This illustration in Fig. 2.2 is purely diagrammatic - it assumes the joint confidence rectangle is a square i.e. $S_A = S_{A^*}$ and also that these values are ≈ 0.02 . In practice $S_A \neq S_{A^*}$ and the values are closer to 0.1. The strains which were classified as P?, Q? or M? are likely to have been so classified due to a high level of background sterility. Intermediate strains i.e. those which were classified as P-Q, Q-P, M-Q or Q-M may be unstable strains which were undergoing a transition from one state to another or they may be stable strains which were maintained in this intermediate state due to the balancing forces of P factor activity and repression. A strain was unclassifiable if the values P_A and P_{A^*} were both >0.1 .

2.3.1.1 Analysis of Southern African wild strains

The classification of the Southern African wild strains together with their 95% confidence regions is given in Appendix A.

21 different localities ranging from Northern Namibia, through South Africa up to Harare in central Zimbabwe. These populations were surveyed for their ability both to produce and resist gonadal dysgenesis using the standard reference crosses, A and A*. (The A cross is between a male of the unknown strain and a female of a standard laboratory M strain and the A* cross is between a female of the unknown strain and a male of a standard laboratory P strain (Kidwell *et al.* 1977b; Engels and Preston 1980)). In addition the reciprocal crosses, B and B*, were performed on all strains collected in 1987 and intrastain crosses were performed on the majority of strains collected in 1988. The results of this chapter indicate the presence of P, Q and M strains in Southern Africa. Further a clinal pattern from northeast to southwest was observed in the regulatory ability and P factor activity of these collections.

2.2 Materials and Methods

2.2.1 Collection of Wild *D. melanogaster* and Establishment of Isotemale Lines

Empty plastic bottles such as 2 litre Coca Cola bottles with a small window cut into the side were used as traps for wild flies. A ripe banana, split lengthways was placed at the bottom of the bottle on top of a wad of cottonwool. The cottonwool was used to absorb the liquid produced by the ripe bananas and thus prevented flies from drowning, particularly during

transportation of the traps. Generally, a few traps were placed at each collection site for anything from 3 days to 2 weeks (the bananas were replaced with fresh ones when necessary). Traps were placed under fruit-trees near garbage bins or even indoors occasionally. These traps were nearly always effective but in addition to *D. melanogaster*, this procedure also trapped individuals of *D. simulans*, a sibling species of *D. melanogaster* which closely resembles it and is found in similar circumstances. Thus, after capture, individual females were placed in separate vials and the genitalia of their male progeny were examined under the dissecting microscope -- this being the simplest way to differentiate between the two species.† The isofemale lines of *D. melanogaster* were thus established.

2.2.2 Geographic Distribution of Samples

Samples of wild type *D. melanogaster* were collected from 17 different sites in Southern Africa. An effort was made to cover as variable a terrain and as large an area as possible and stocks were obtained from Namibia and Zimbabwe as well as from South Africa. Table 2.1. shows the site locations, the date the flies were caught and the numbers of isofemale lines tested for P-M gonadal dysgenesis.

In the males of both *D. melanogaster* and *D. simulans* a discoid process arises from the posterior margin of the genital arch. This process is large and protuberant in *simulans* (resembling a small bivalve shell when viewed together with its partner on the opposite side of the body), whereas in *melanogaster* the corresponding process is much smaller and is discernible only under extremely high magnification. Thus separation of males of the two species in this manner is unequivocal.

Table 2.1 Collection sites, dates and number of isofemale lines tested for P-M gonadal dysgenesis

Collection Site	Date	Number of isofemale lines
Namibia		
1. Swakopmund	Oct 1988	10
2. Gobabeb	Oct 1988	18
3. Ai Ais	Oct 1988	15
South Africa		
4. Cape Town	Oct 1988	14
5. Gordons Bay	Dec 1987	4
6. Plettenberg Bay	Dec 1987	4
Port Elizabeth		
7. Summerstrand	Oct 1988	19
8. Walmer	Oct/Nov 1988	6
9. Maclear	Dec 1987	4
Durban		
10. Berea	Dec 1988	16
11. Addington Beach	Oct 1988	16
Johannesburg		
12. Northern suburbs 1988	Dec 1988	8
13. Northern suburbs 1984, 1987	Dec 1987/1984	5
14. Southern suburbs	Dec 1987	4
15. Pretoria	Dec 1987	4
16. Hammanskraal	Dec 1987	4
17. Nelspruit	Nov 1988	16
18. Pietersburg	Dec 1987	4
19. Agatha	Dec 1987	4
20. Thohoyandou	Oct 1988	12
Zimbabwe		
21. Harare	Oct/Nov 1988	19
Total		<u>206</u>

2.2.3 Drosophila Stocks Used for P-M Testing.

All genetic symbols not otherwise explained are in Lindsley and Grell (1968).

Two hundred and six (206) isofemale wild lines derived from 21 different localities in South Africa, Namibia and Zimbabwe.

*₂: A strong P strain obtained from W.R. Engels. This strain has many effective transposase and GD sterility producing P factors throughout the genome.

y *en^W(M)/y⁺*; bw; st: Obtained from W.R. Engels. The cytotype is M and there are no P factors other than those inserted at or near the *singed* locus. So this strain is strictly speaking an M' strain. However, it is a stable strain (Engels 1984) and it has been found by Anxolabehère *et al.* (1985b) to react as a more sensitive M cytotype than Canton-S (a standard M tester strain). This strain will hereafter be referred to as *en^W(M)*.

2.2.4 Description of Crosses Made and Methods Used for GD Sterility Testing

In order to determine the type of an unknown (U) strain, the following crosses were made:

cross A: *en^W(M)* ♀ x U ♂

cross A*: U ♀ x *en^W(M)* ♂

Female progeny were dissected for gonadal dysgenesis.

Cross A determines the level of activity of a strain's functional P elements when they are in a derepressed state, i.e. it measures P activity potential.

Cross A* determines the cytotype of an unknown strain, i.e. it measures the ability of an unknown female to resist the action of a known P strain male in causing gonadal dysgenesis.

In addition 37 of the wild strains (those collected in 1987 and the one collected in 1984) were tested for background sterility by performing the following crosses (and dissecting the female progeny for GD sterility):

Cross B (reciprocal of cross A): $U \text{♀} \times \pi^W(M) \text{♂}$

Cross B* (reciprocal of cross A*): $\pi_2 \text{♀} \times U \text{♂}$

Cross B measures the background sterility in females genotypically identical to the female progeny from cross A. This sterility may arise from a low level of P element activity in lines with P cytotype or from other causes. Consequently any sterility in cross B hybrids could serve as a measure against which the sterility of cross A hybrids can be judged and in this sense cross B serves as a control. Cross B* usually does not lead to sterility regardless of the P-M properties of the unknown males and females. So cross B* measures background levels of GD sterility in non-dysgenic crosses involving the unknown lines.

Finally 134 of the strains collected in 1988 were tested for intrastrain GD sterility.

Virgin females were held for 3-7 days at 20°C after eclosion before crossing them. For each cross, 15-20 pairs of flies were mated *en masse* in half-pint bottles containing the standard yeasted medium as prepared by the Department of Genetics, University of the Witwatersrand, and immediately placed at 29°C. About one to two days after eclosion, flies of both sexes were transferred to fresh bottles and allowed to mature at 20°C for three to four days before the females were dissected. Fig. 2.1 shows the difference between the normal and dysgenic ovaries. An ovary was scored as nondysgenic if even one ovariole was developed (Schaefer *et al.* 1979). The frequency of gonadal sterility was calculated by dividing the number of dysgenic gonads by the total number of gonads scored.

2.2.5 Confidence Intervals

In order to account for sampling errors in crosses A and A* a 95% joint confidence probability was employed (Louw and van Aarde 1989). Thus, if P_A and P_{A^*} are the GD sterility frequencies and n_A and n_{A^*} the sample sizes in crosses A and A* respectively, then a 95% joint confidence region for the proportions of dysgenic ovaries obtained in crosses A and A* is given by:

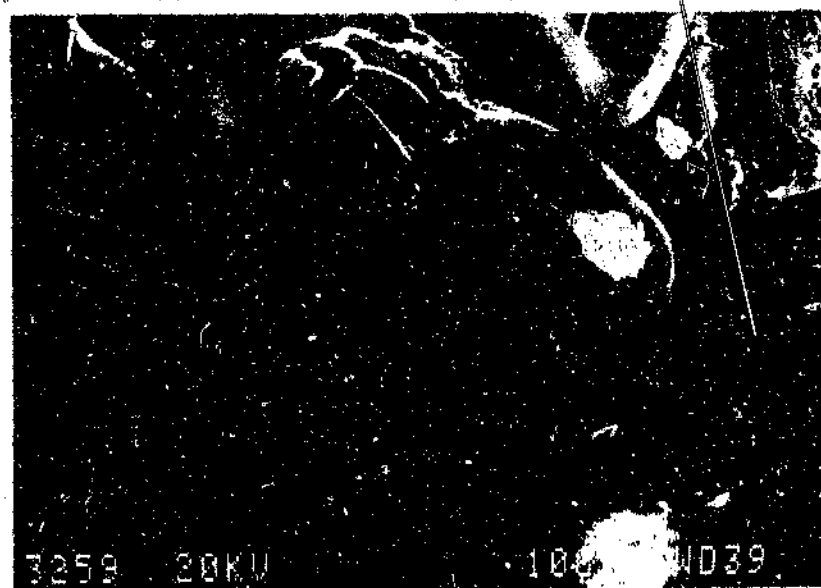
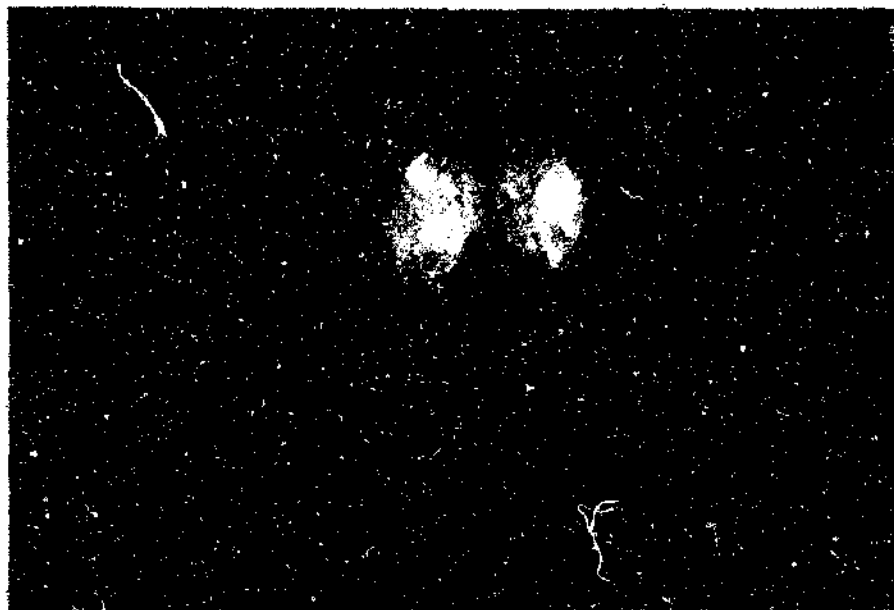


Fig. 2.1 A Scanning electron microscopy photograph of a normal tarry from an adult raised at 29°C.

(i)



(ii)

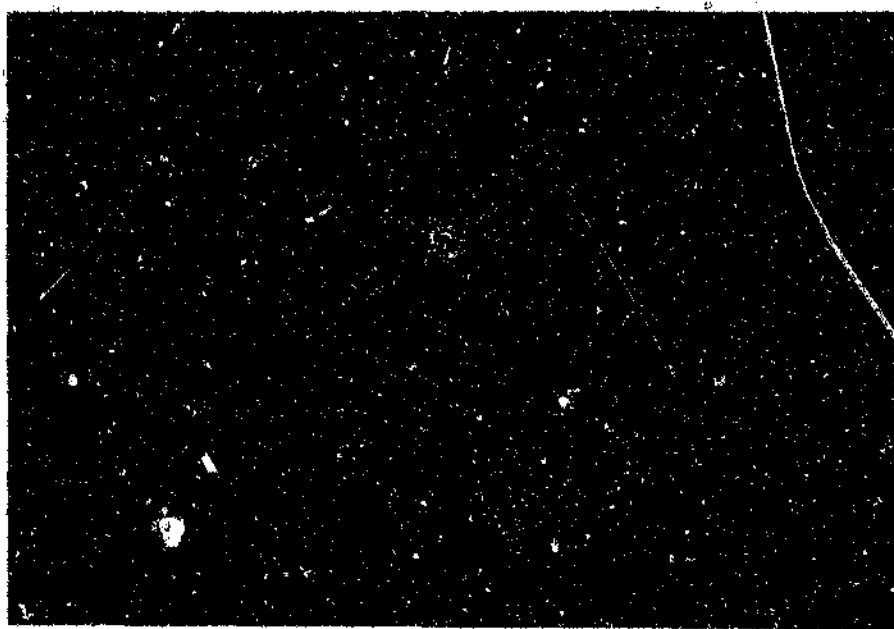


Fig. 2.1 Pairs of normal ovaries from adults raised at 29°C.
Magnification: (i) 35 x
(ii) 96 x

$$P_A \pm Z_A (P_A(1-P_A)/n_A)^{1/2} \text{ and } P_{A^*} \pm Z_{A^*} (P_{A^*}(1-P_{A^*})/n_{A^*})^{1/2}$$

where $p(1-p)/n$ is the variance ($p = P_A$ or P_{A^*} ; $n = n_A$ or n_{A^*}) and Z ($Z = Z_A$ or Z_{A^*}) is the standard normal deviate. If the value of Z corresponds to a confidence level (for each individual cross) of α , then, since an overall 95% confidence level is required, α is obtained from the equality: $(1-\alpha)^2 = 0.95$, whence $\alpha = 0.0254$. Thus $Z = Z_A = Z_{A^*} = 2.235$.

2.3 Results

2.3.1 Classification Scheme

The estimated GD sterilities (P_A and P_{A^*}) obtained in crosses A and A^* , together with 95% confidence regions ($P_A \pm S_A$ and $P_{A^*} \pm S_{A^*}$) are given in Appendix A. ($S_A = 2.235(P_A(1-P_A)/n_A)^{1/2}$; $S_{A^*} = 2.235(P_{A^*}(1-P_{A^*})/n_{A^*})^{1/2}$. n_A and n_{A^*} are the numbers of ovaries dissected in crosses A and A^* respectively). It is useful to be able to type the wild lines as M, Q or P in order to obtain a picture of the geographic distribution. Following Kidwell (1983) and Kidwell *et al.* (1983), the 10% GD sterility level was used as the demarcation line in that a strain was designated M if it produced <10% dysgenic females in cross A and >10% in cross A^* , Q if it gave <10% dysgenesis in both reference crosses, and P if it gave >10% dysgenesis in cross A and <10% in cross A^* . Remarkably most wild strains fell clearly into one or other of these three categories. However, when the 95% confidence regions (section 2.2.5) were evaluated it was found that a certain amount of overlapping between P, Q and M areas sometimes occurred

Table 2.2

P-Q-M type classification scheme used for Southern African populations of *Drosophila melanogaster* on the basis of 95% confidence regions for GD sterilities in crosses A and A*

Cross A ¹			Cross A* ¹			Type
PA-SA ³	PA ²	PA+SA ³	PA*-SA* ³	PA* ²	PA*+SA* ³	
		≤0,1			≤0,1	Q
	≤0,1	>0,1			≤0,1	Q-P
≤0,1	>0,1				≤0,1	P-Q
>0,1					≤0,1	P
		≤0,1		≤0,1	>0,1	Q-M
	≤0,1	>0,1		≤0,1	>0,1	Q?
≤0,1	>0,1			≤0,1	≤0,1	P-Q
>0,1				≤0,1	≤0,1	P?
		≤0,1	≤0,1	>0,1		M-Q
	≤0,1	>0,1	≤0,1	>0,1		M-Q
≤0,1	>0,1		≤0,1	>0,1		?
>0,1			≤0,1	>0,1		?
		≤0,1	>0,1			M
	≤0,1	>0,1	>0,1			M?
≤0,1	>0,1		>0,1			?
>0,1			>0,1			?

¹Cross A: $sn^W(M) \text{♀} \times U \text{♂}$

Cross A*: $U \text{♀} \times \pi_2 \text{♂}$

U is the unknown strain; $sn^W(M)$ and π_2 are described in Section 2.2.3.

²PA, PA* are estimated GD sterilities obtained in crosses A and A* respectively.

³SA = $Z_A (PA(1-PA)/n_A)^{1/2}$, SA* = $Z_{A*}(PA*(1-PA*)/n_{A*})^{1/2}$ where n_A , n_{A*} are the sample sizes and Z_A , Z_{A*} are the standard normal deviates for crosses A and A* respectively. For a joint 95% confidence region $Z_A = Z_{A*} = 2,235$.

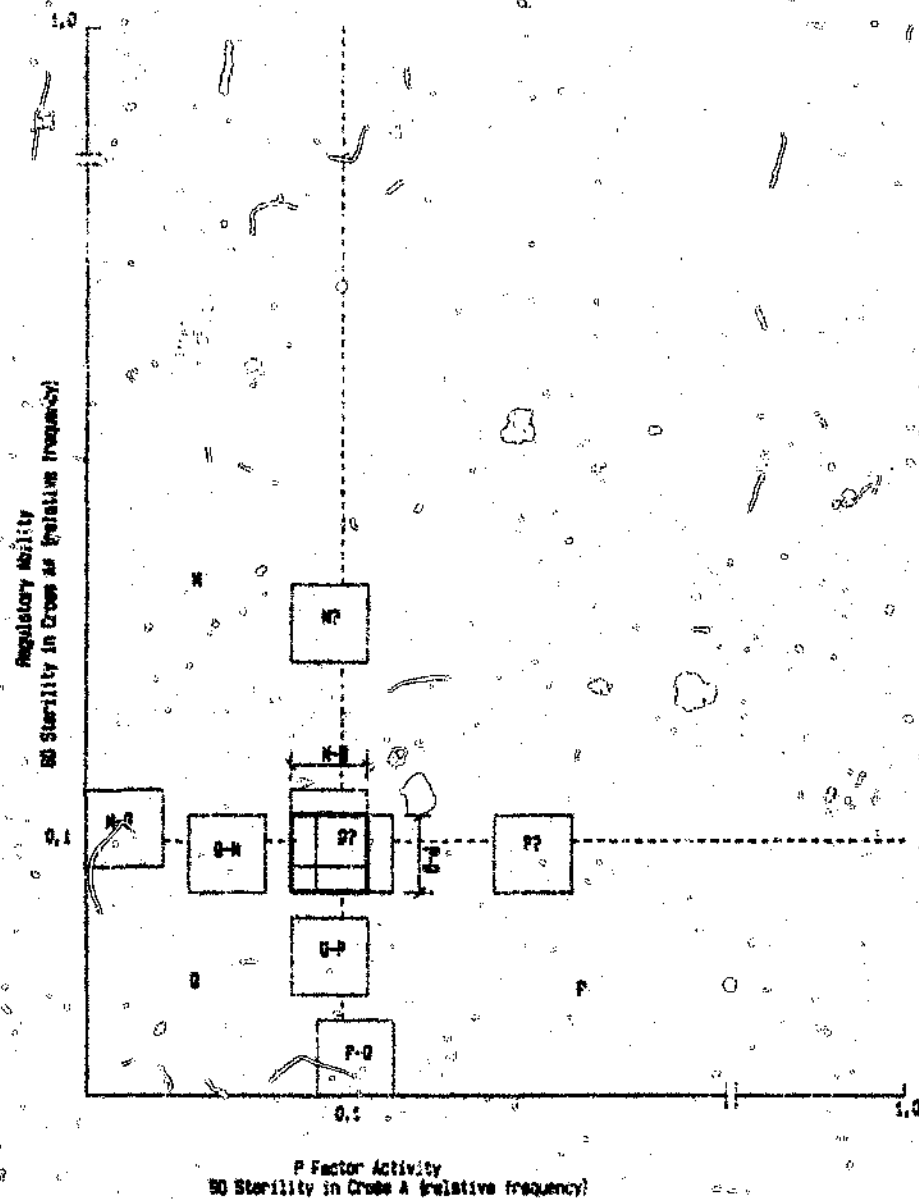


Fig. 2.2 Illustration of the P-Q-M classification system using joint confidence probabilities.

Each square represents a joint confidence region.

Kidwell (1986b) introduced a refinement in the classification of M and P strains by dividing M strains into true M and pseudo M (M') strains (Kidwell 1985) and P strains into moderate and strong P as follows:

Strain type	% GD sterility	
	Cross A	Cross A*
M (true)	0	100
M' (pseudo M)	0-?	0-100
Q (weak P)	0-10	0-10
P (moderate)	11-80	0-10
P (strong)	81-100	0-10

According to this classification, all wild M strains in this study except Ais 4 and Wal 3 were of type pseudo M. Ais 4 and Wal 3 both had GD sterility relative frequencies of 0 and 1 in crosses A and A* respectively (Appendix A). Thus according to the above classification these strains could be either true M or pseudo M. No strong P strains were obtained. Figs 2.3 (A-P) show GD sterility frequencies for each individual collection site as points on a graph with axes representing cross A and cross A* percentage GD sterilities (Yamamoto *et al.* 1984). Only two strains were completely unclassifiable. These were Ais 14 and Rut 2 (Figs 2.3 C and I, Appendix A). Intrastrain

[†](Kidwell's (1986b) Table 4 also included references to *an*^w mutability which have been omitted here).

Fig. 2.3 The distributions of GD sterilities presented as points on graphs of cross A vs. cross A* (Yamamoto et al. 1964)

- A: Senegamb
- B: Gambia
- C: Al Ais
- D: Cape Town; Gordons Bay
- E: Plattenberg Bay
- F: Port Elizabeth
- G: Muclear
- H: Durban
- I: Johannesburg
- J: Victoria
- K: Mammanskraal
- L: Nelspruit
- M: Pietersburg
- N: Agatha
- O: Thohoyandou
- P: Harare

Numbers on the graphs refer to the number of strains having these coordinates.

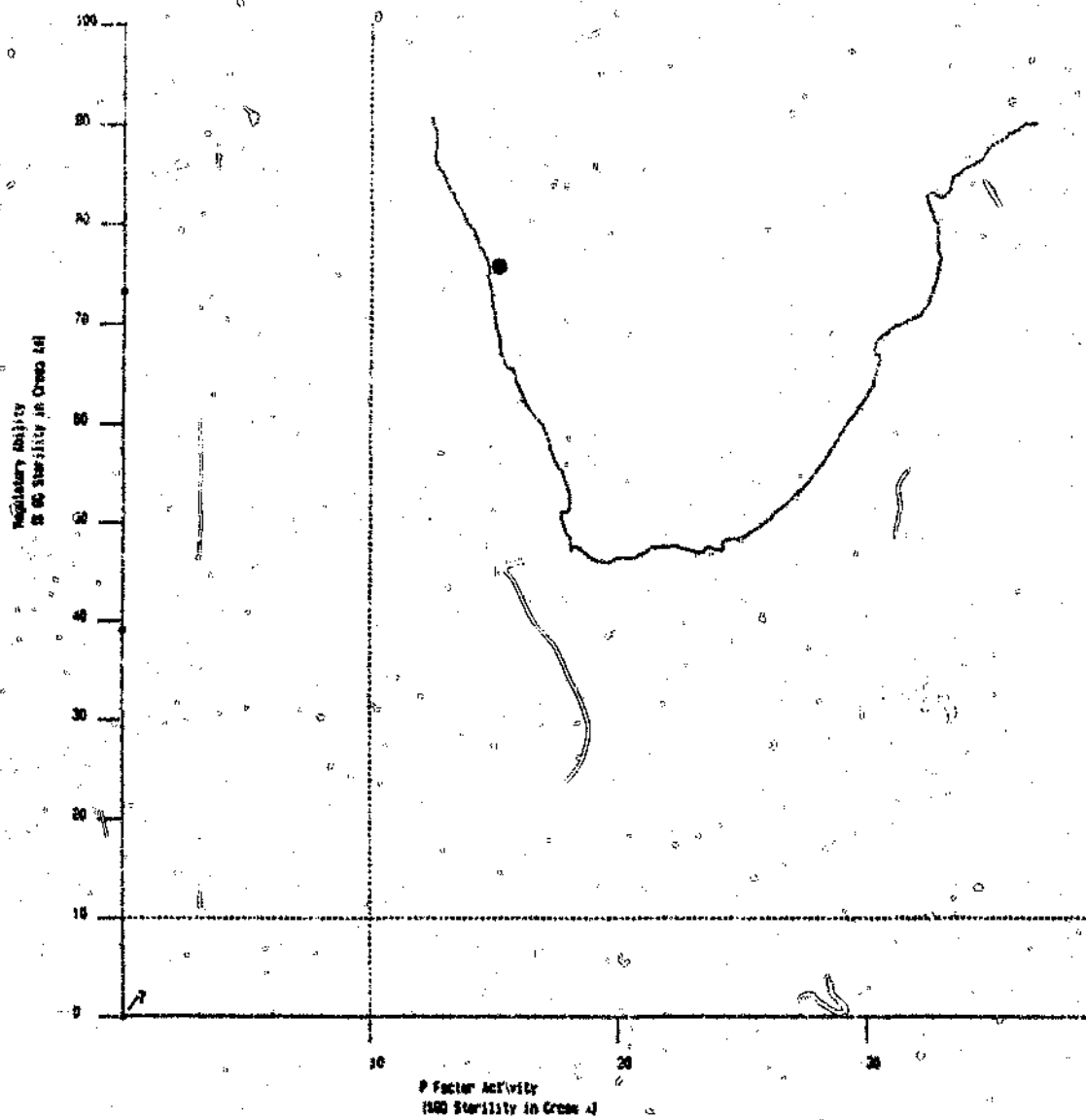


Fig. 2.3 A: Sankopand

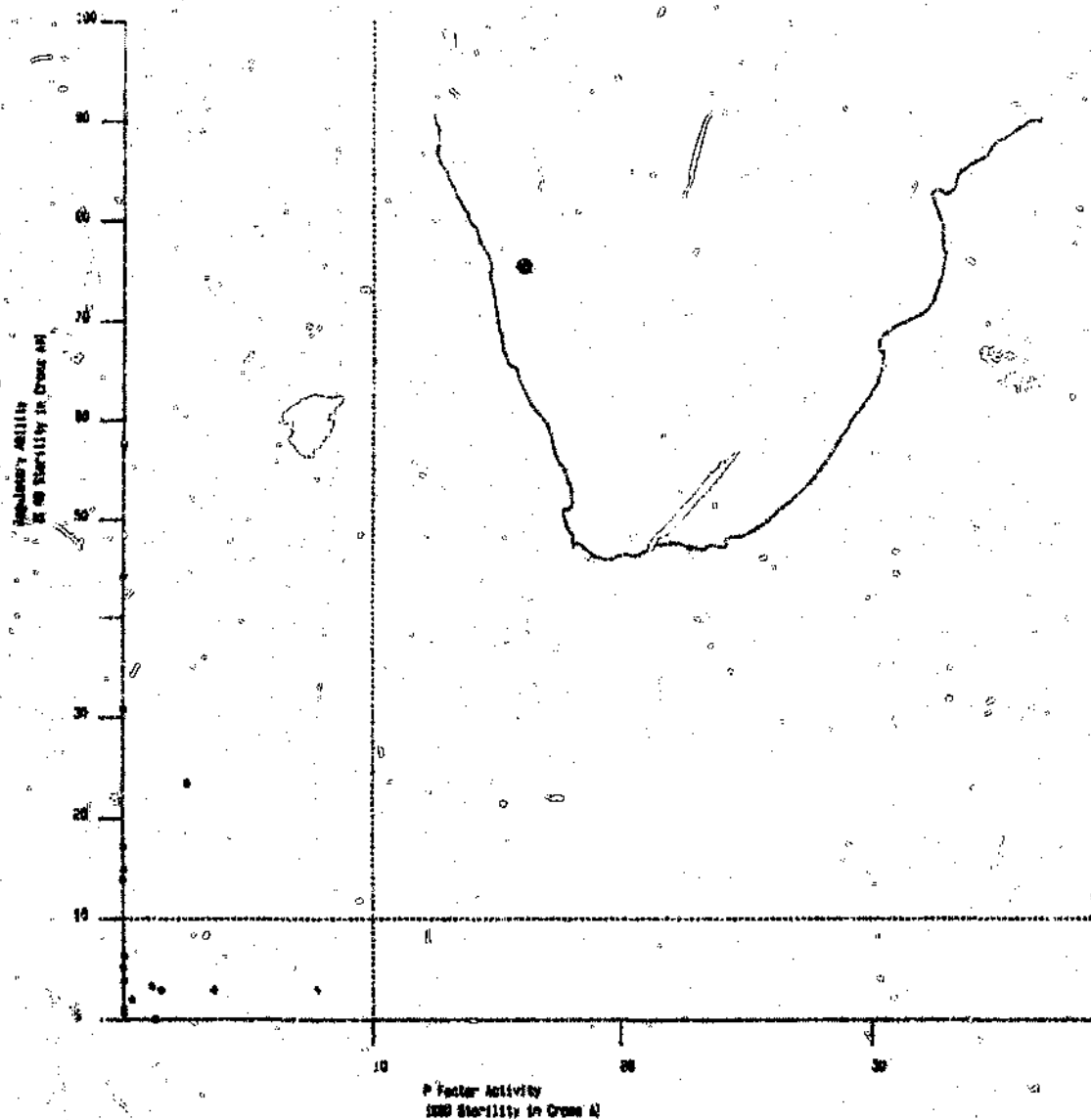


Fig. 2.3 B: Goshok

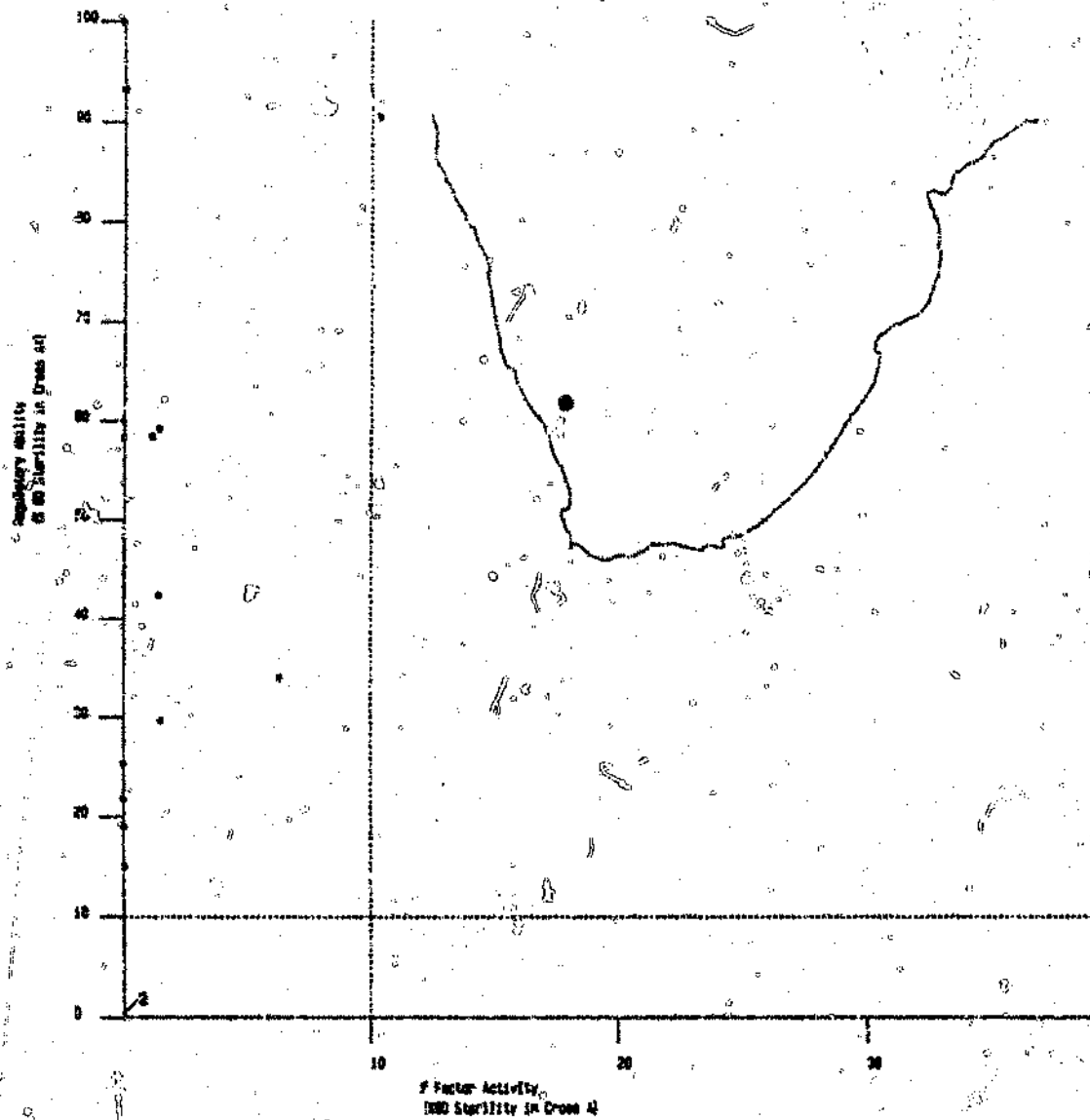


Fig. 2.3 C: At Air

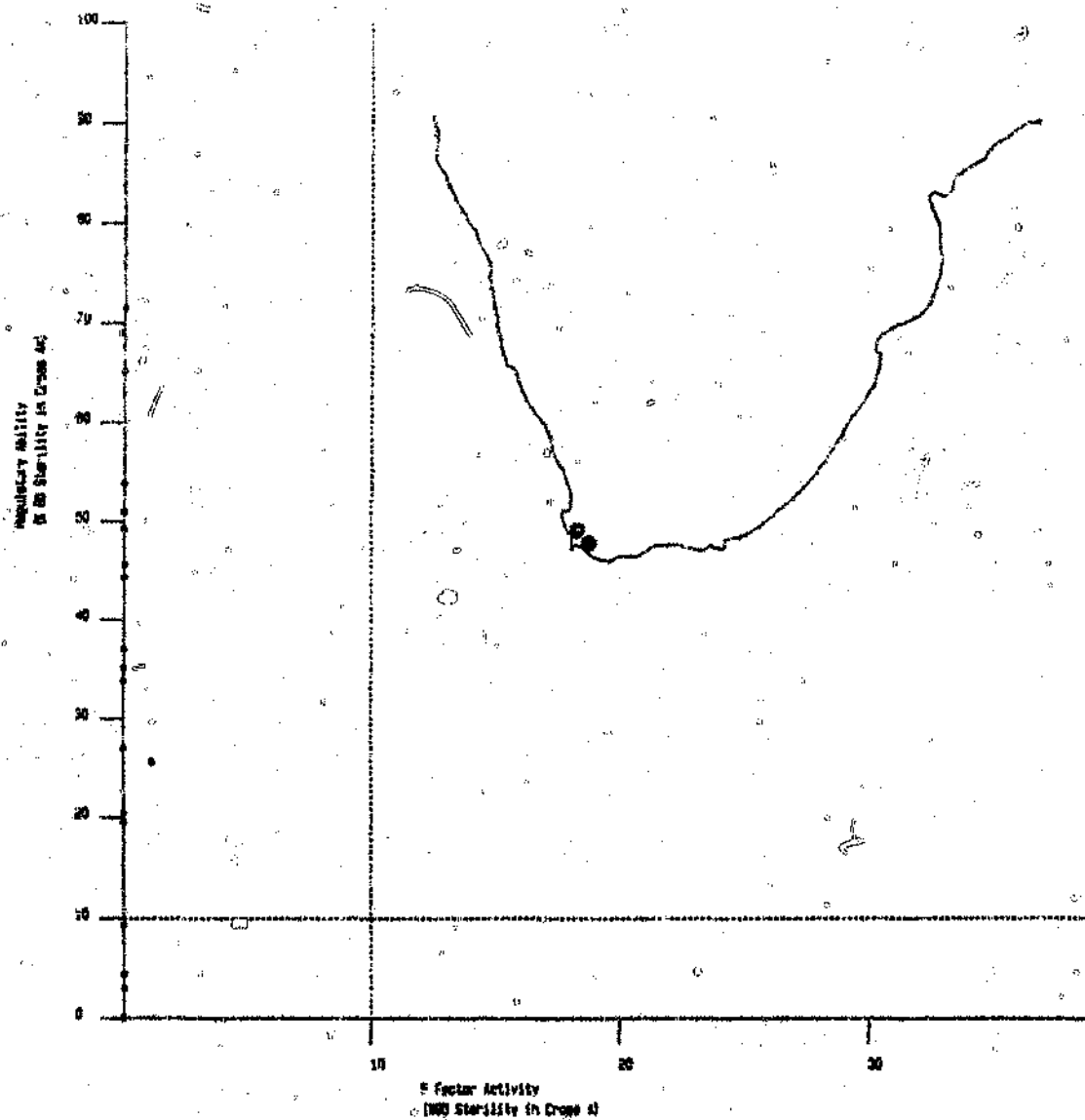


Fig. 2.2 B: Cape Town: Gordons Bay

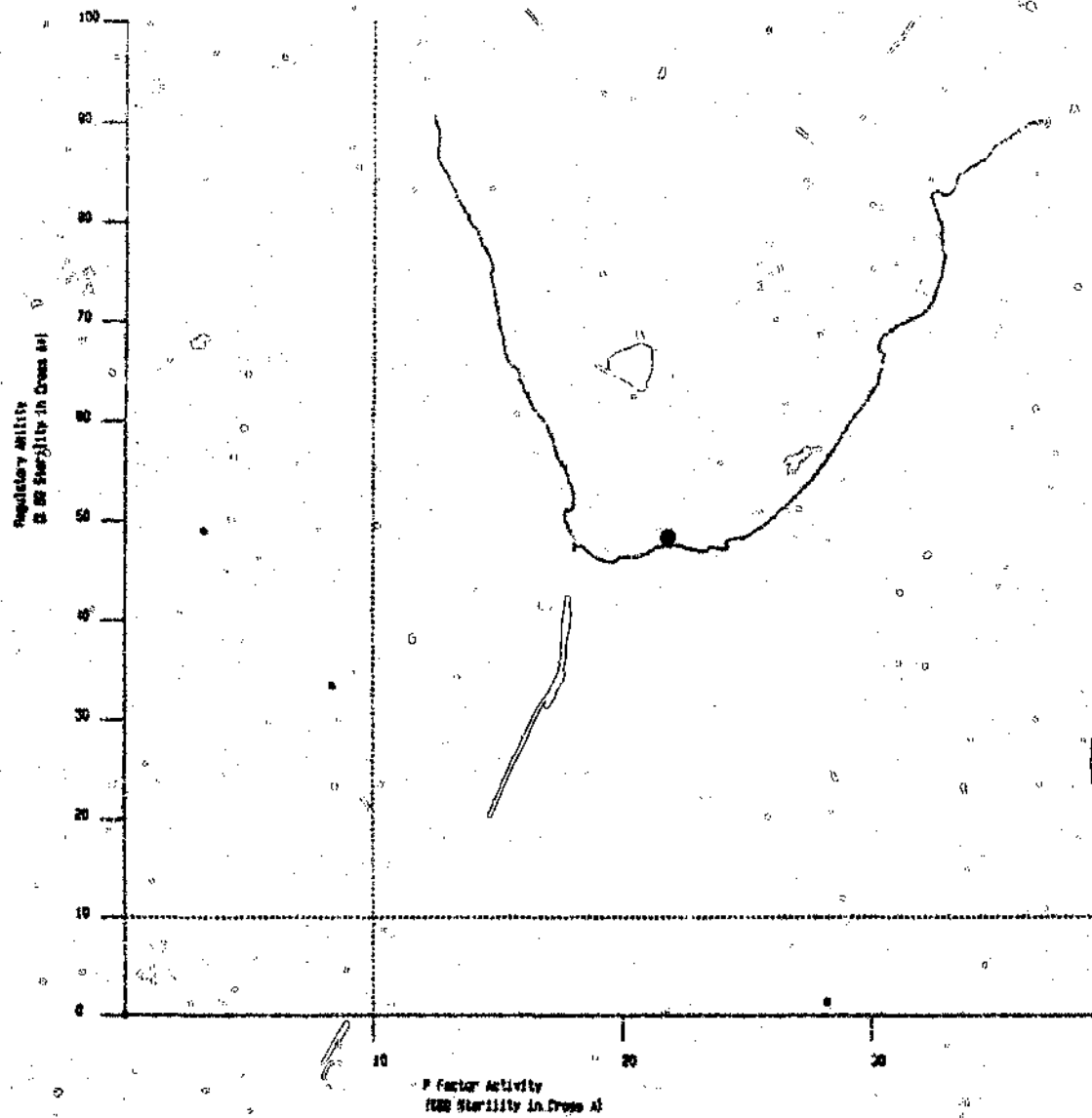


Fig. 2.3 E: Plettenberg Bay

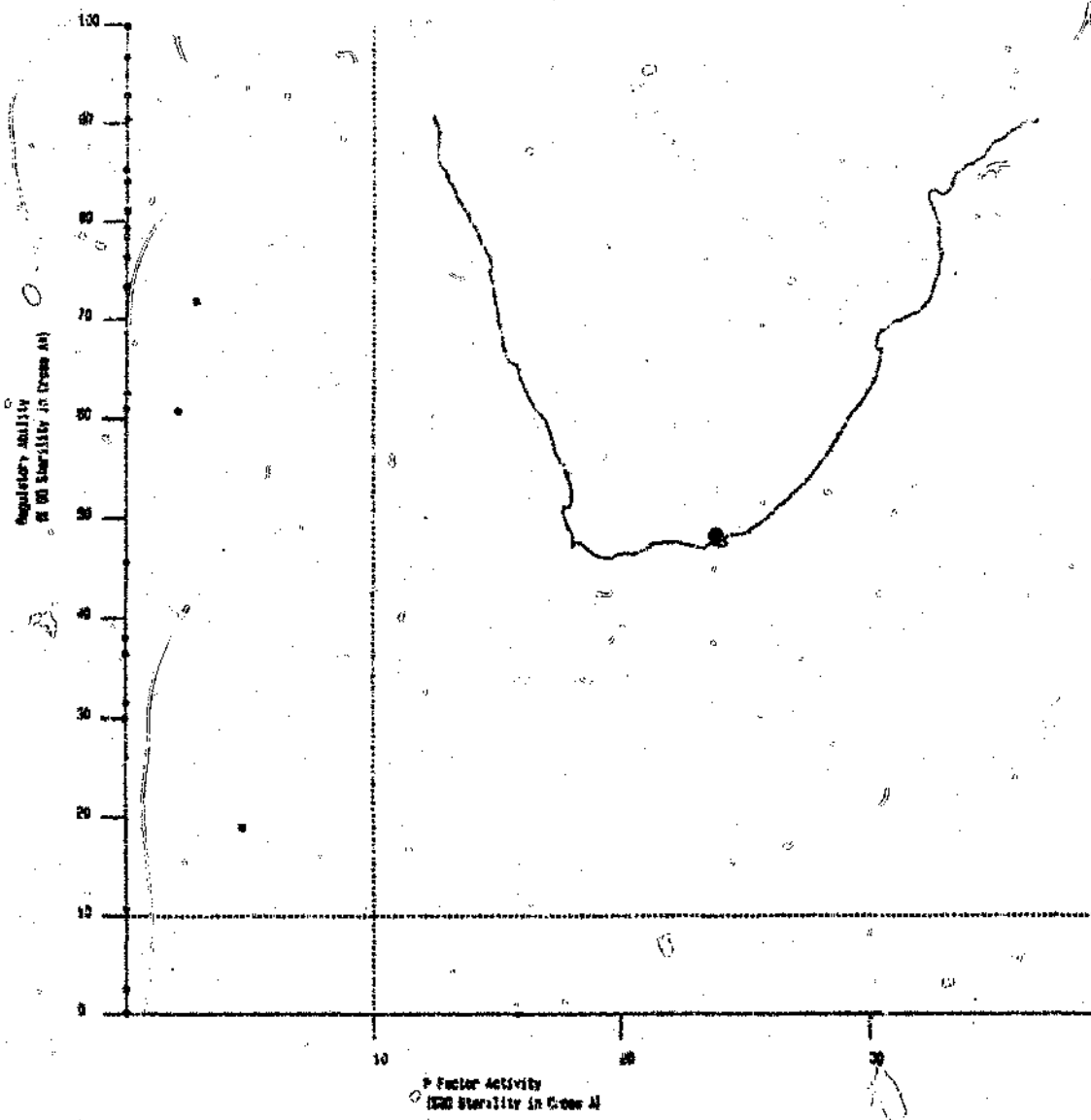


Fig. 2.3 F: Port Elizabeth; Summerstrand, Malmor

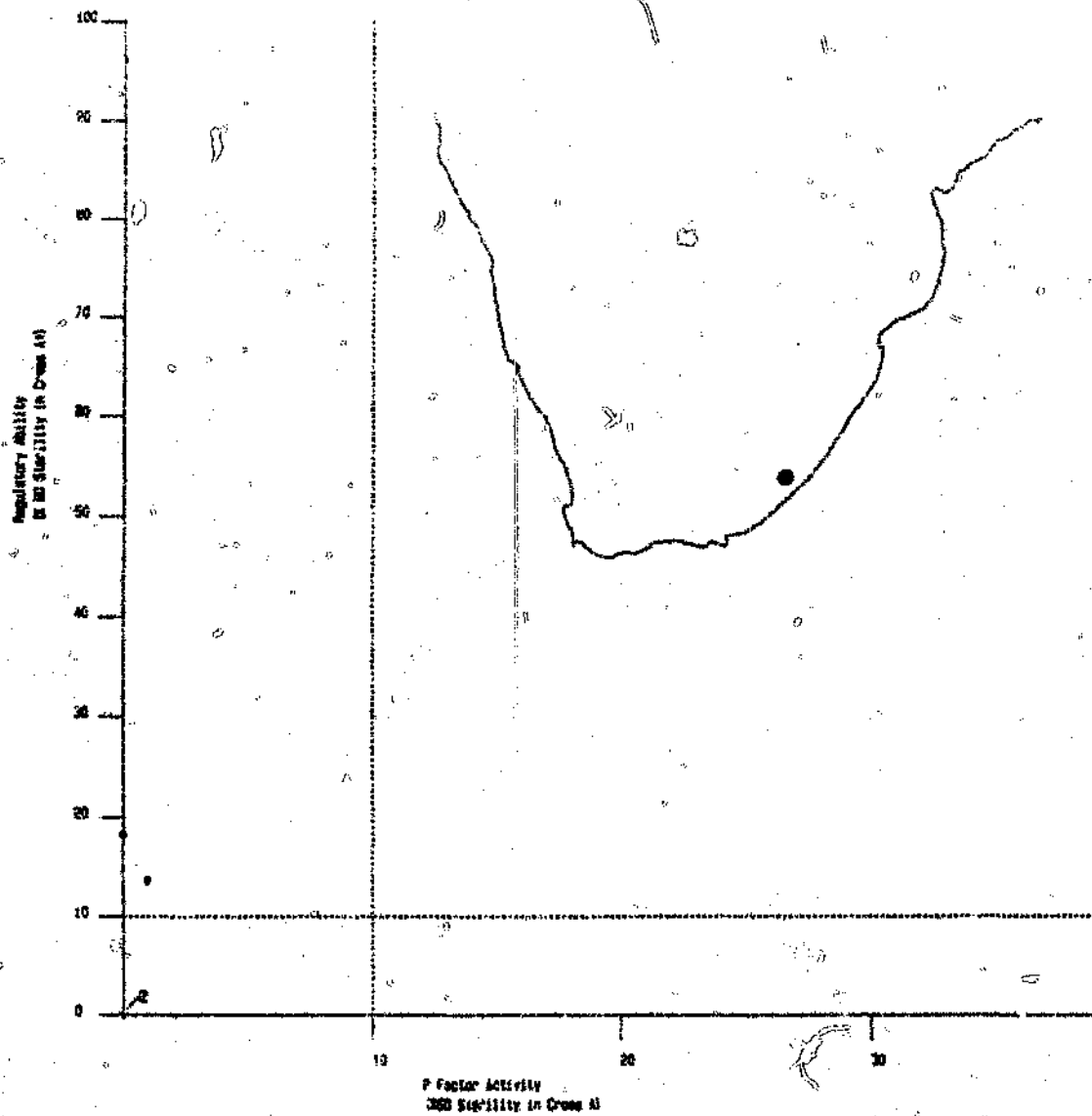


Fig. 2.3 G: Nuclear

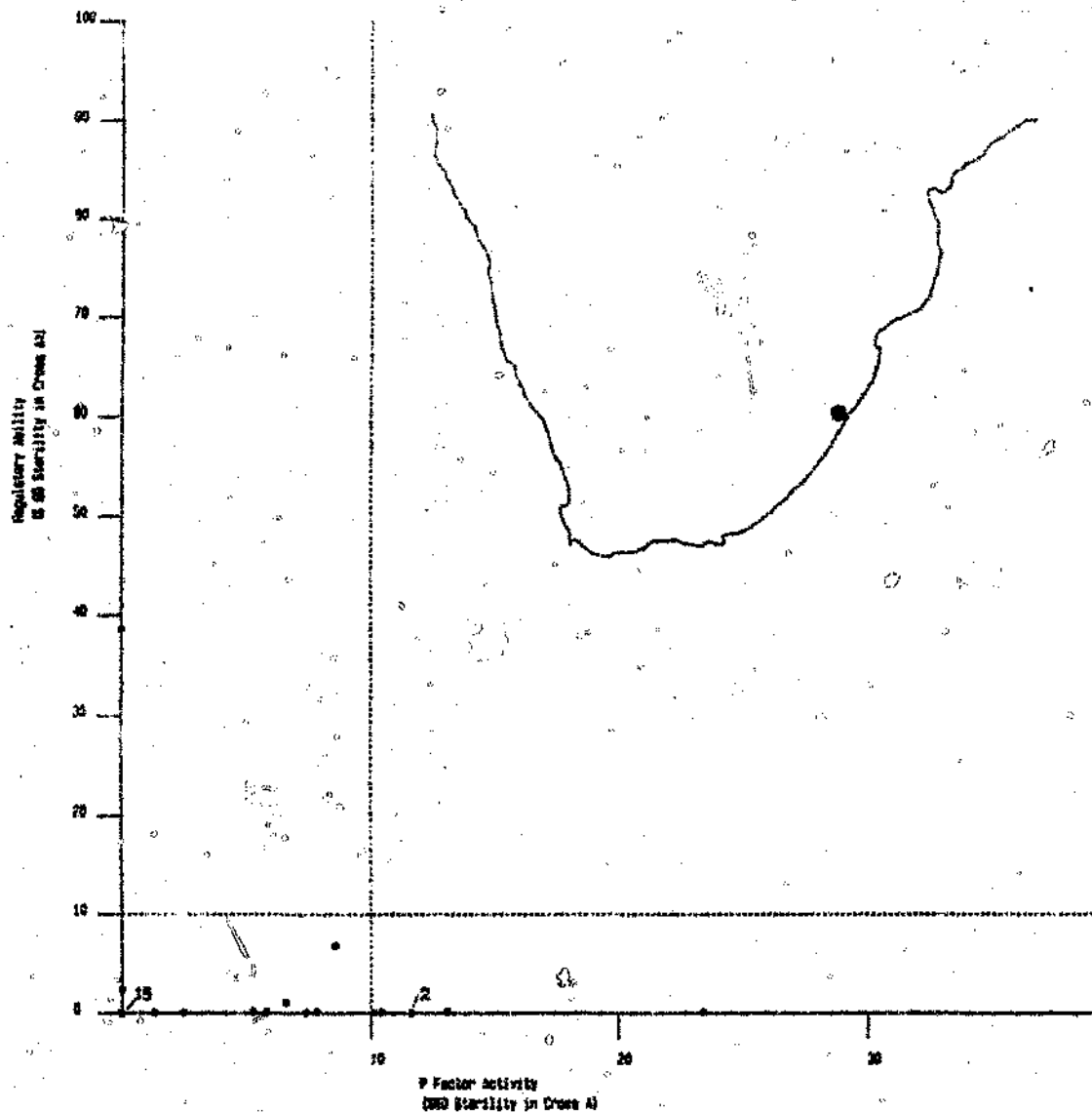


Fig. 2.2 H: Durban; Addington Peach, Barea.

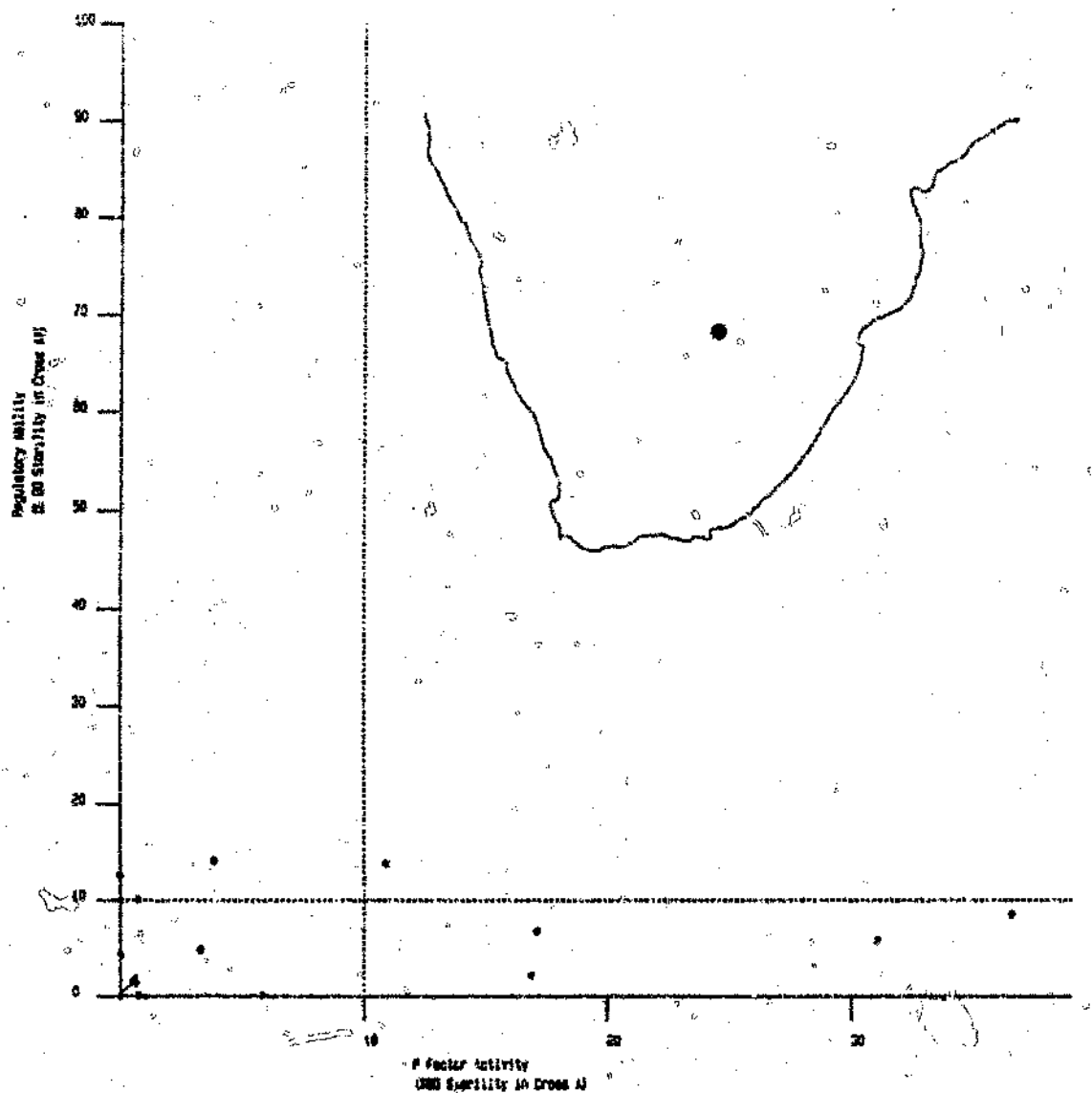


Fig. 2.3 I: Johannesburg:

Northern Suburbs 1984/1987,
 Northern Suburbs 1988
 Southern Suburbs.

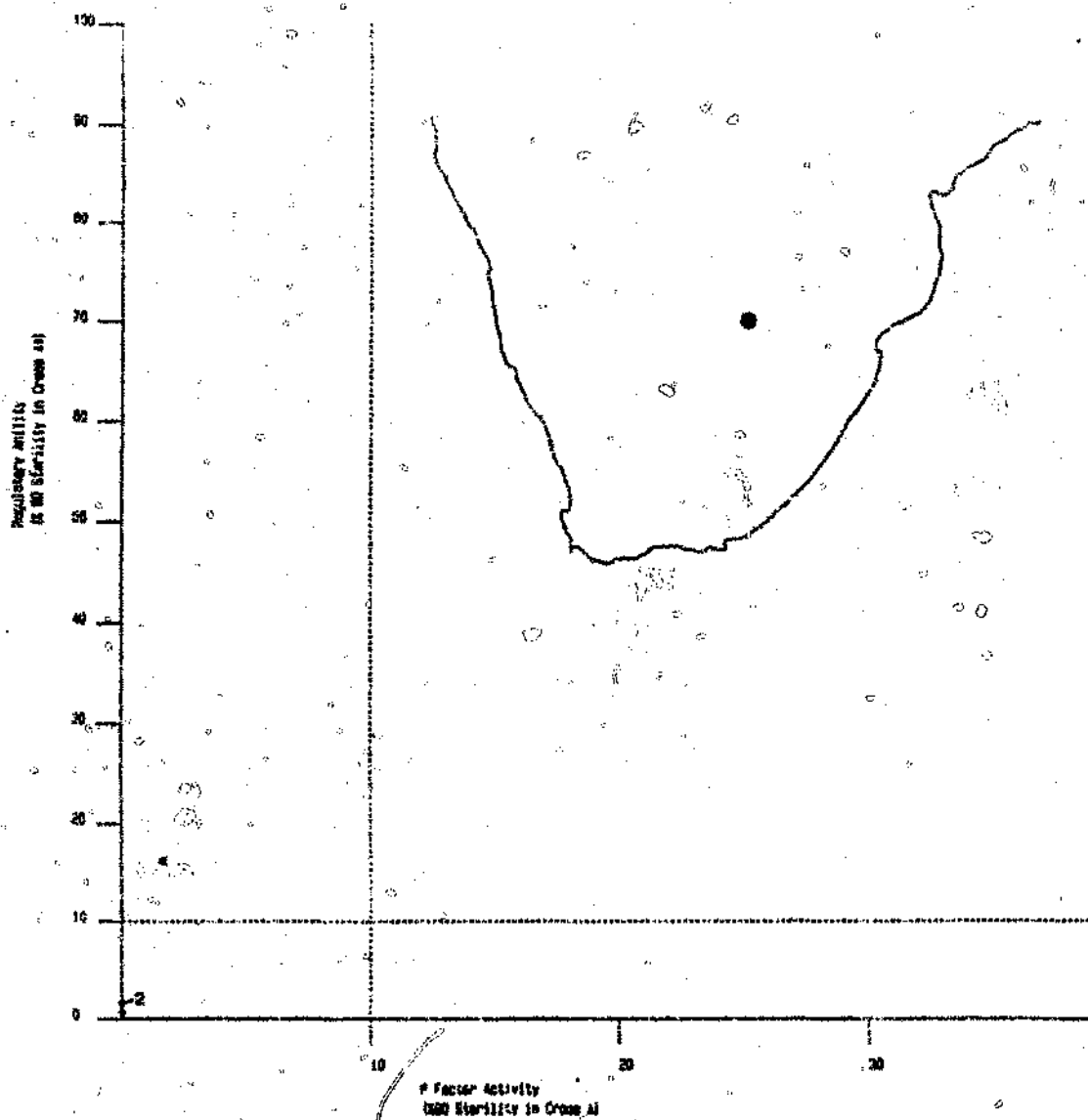


Fig. 2.3 J: Proteoria

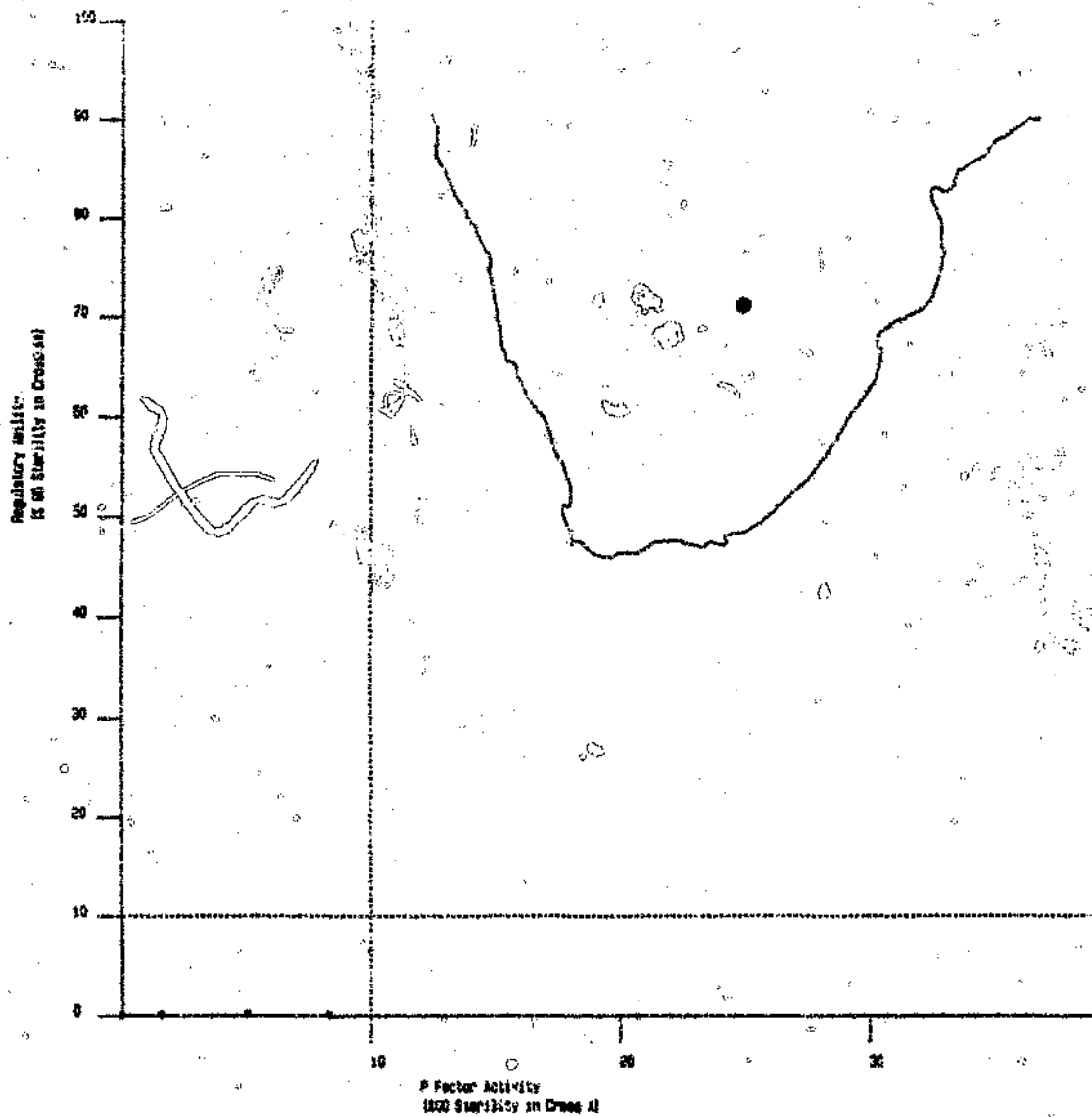


Fig. 2.3 K: Hammonskraal

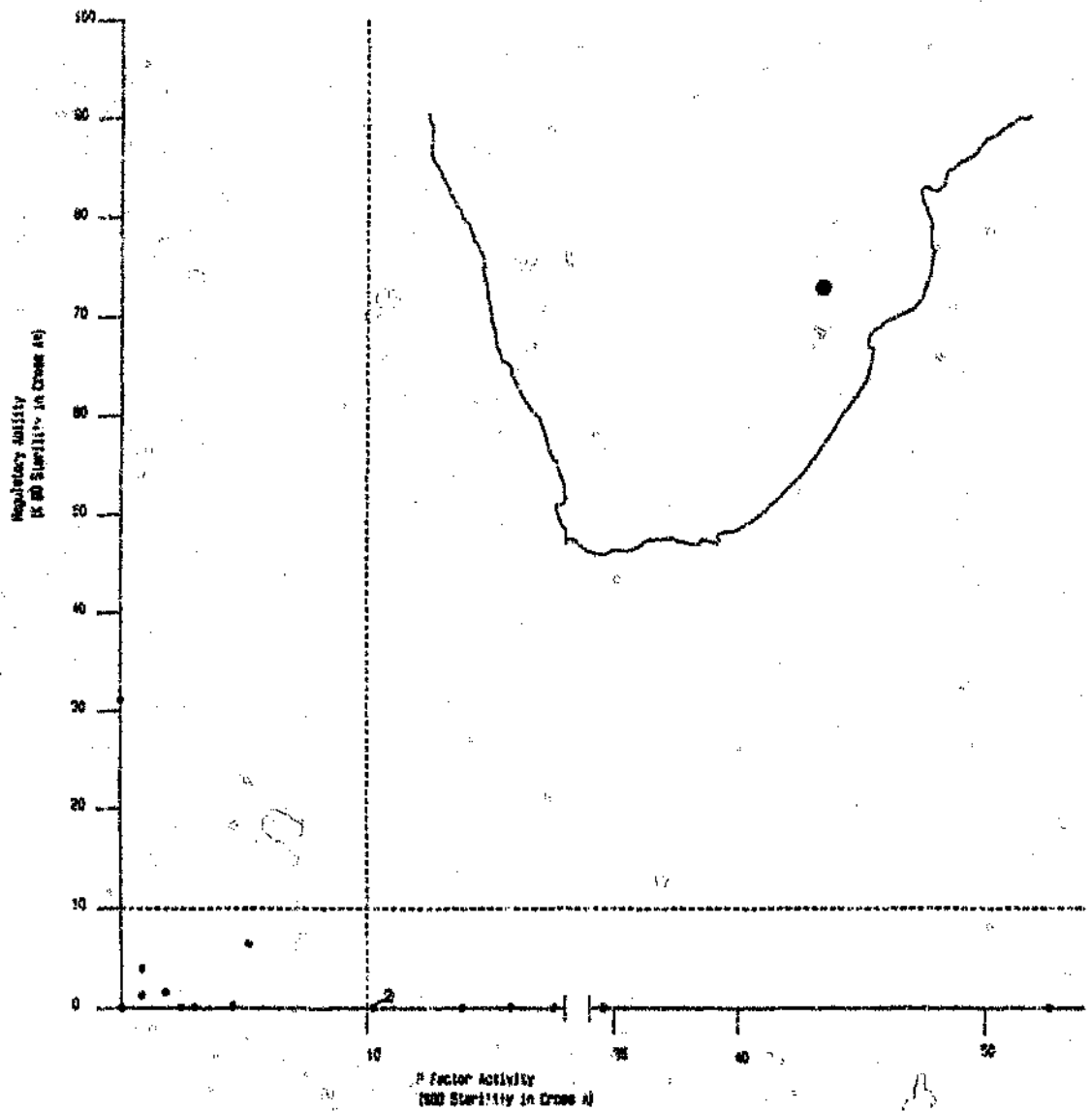


Fig. 2.3 L: Nelspruit

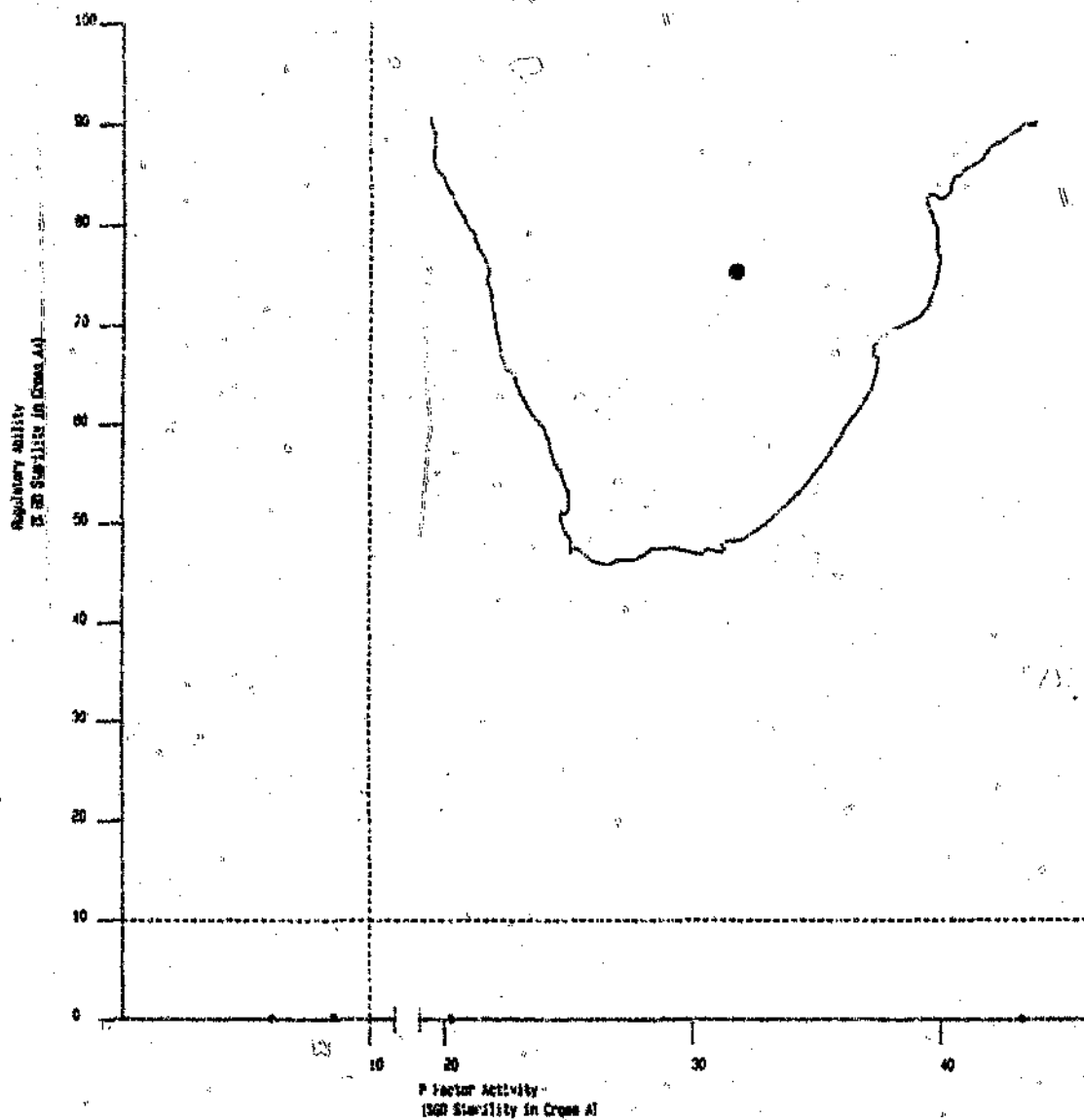


Fig. 2.3 N: Pietersburg

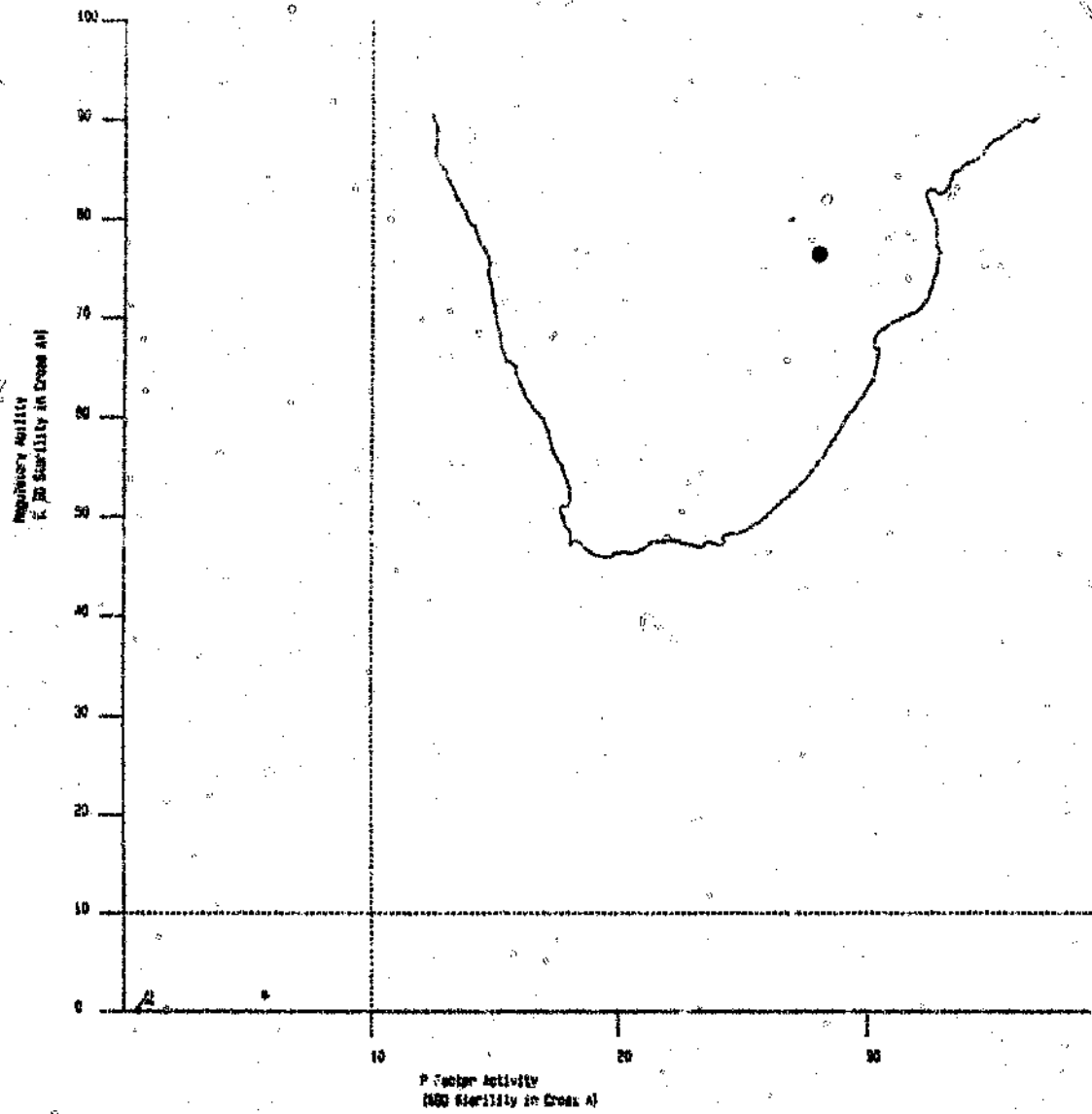


Fig. 2.3 W: Agutim

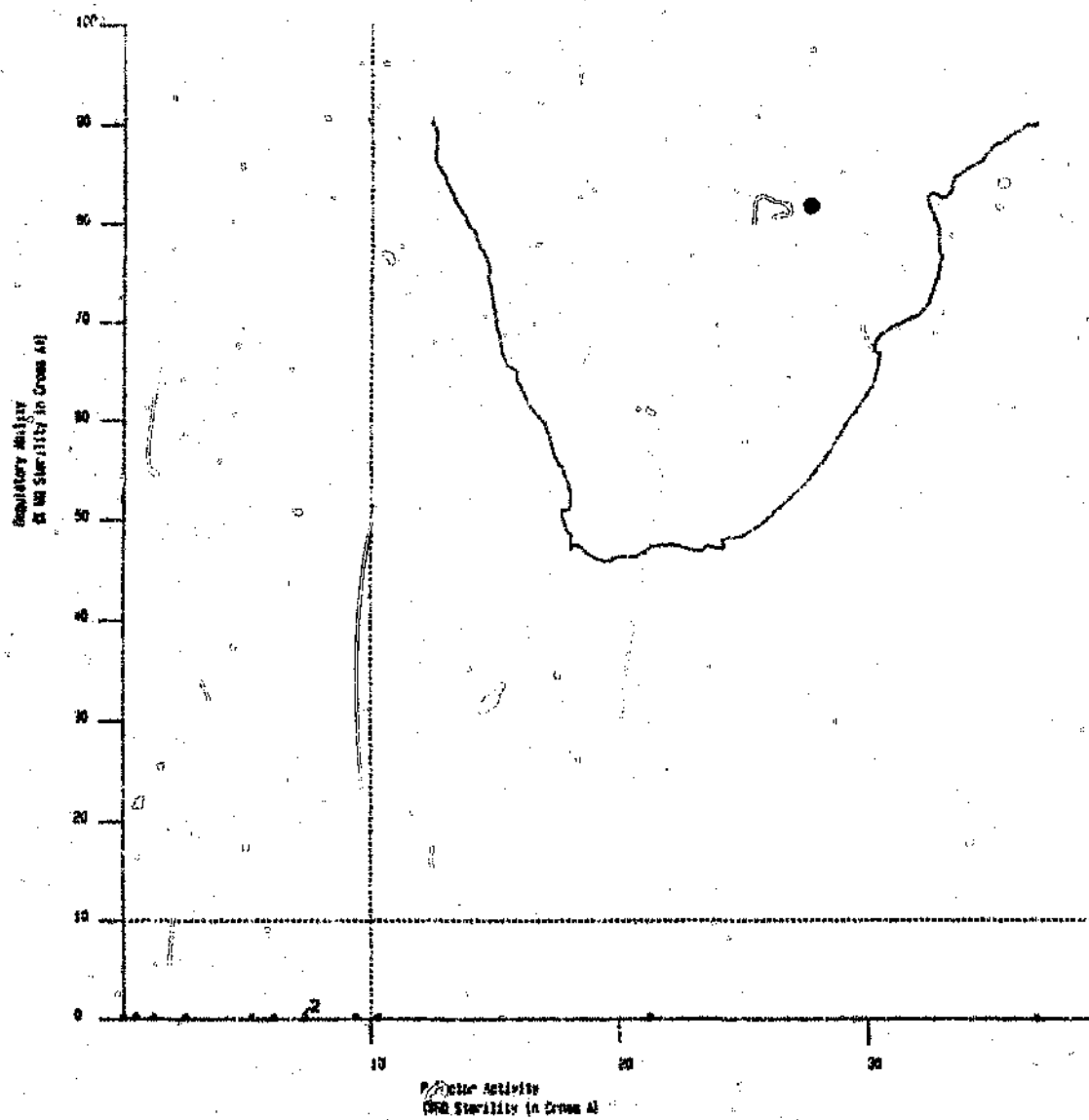


Fig. 2.3 O: Tachycondon

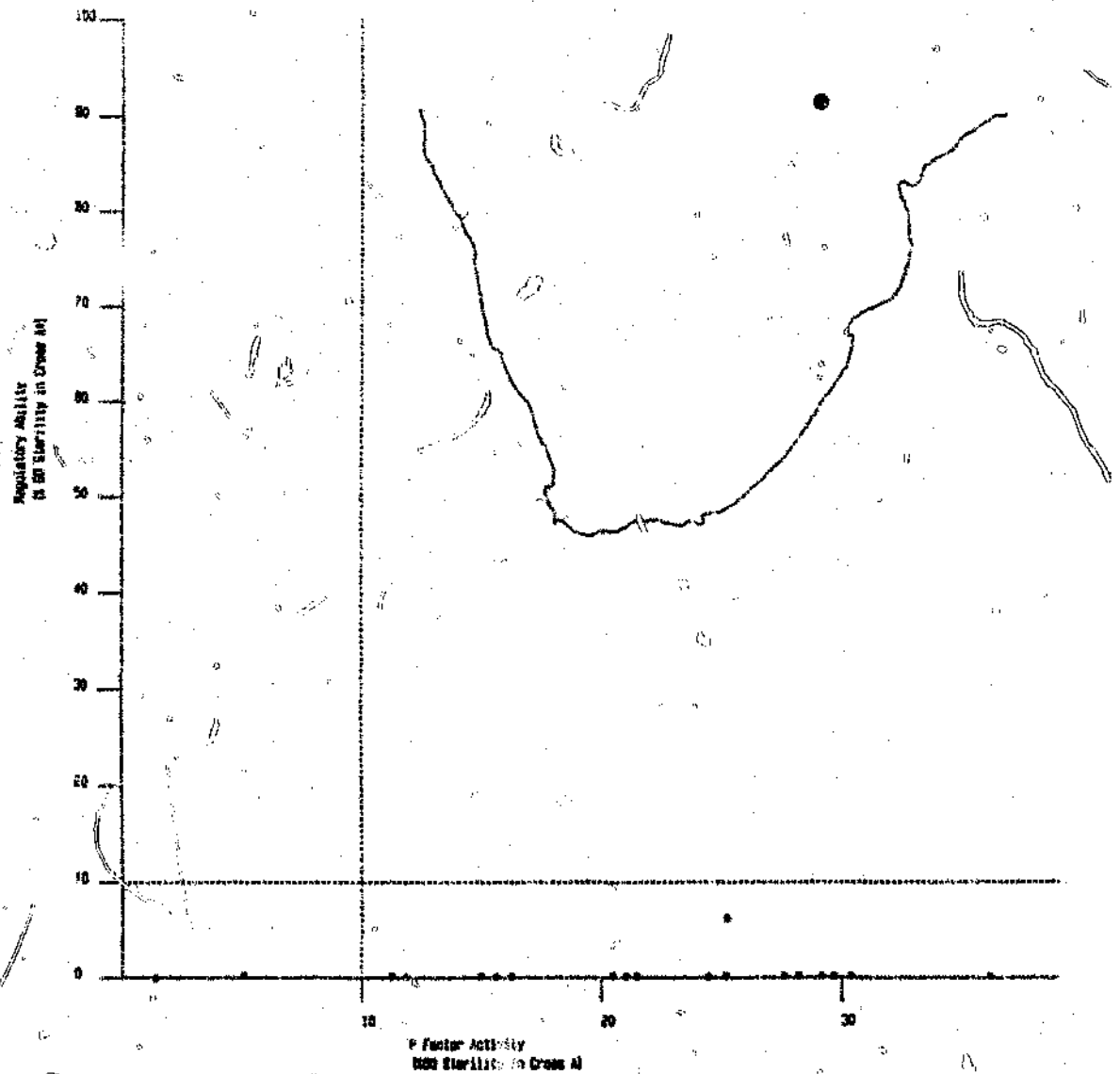


Fig. 2.3 P: Harera

sterility for Ais 14 was 14.6% and for Rut 2, GD sterility in the reciprocal cross B* was 7.9% (Appendix B). So both of these strains had high background sterility which would explain their inability to be classified. Altogether 204 out of the 206 wild Southern African strains had GD sterility <10% in either cross A or cross A* (or both). This distribution of GD sterility is not unique to Southern Africa; indeed it appears to be a worldwide phenomenon. For example in a series of studies spread over regions including Europe, Africa, Asia, India, China, Japan and the Americas, not one of approximately 150 different strains had GD sterility >10% in both cross A and cross A* (Anxolabéhère *et al.* 1982a, 1984, 1985b). With a few exceptions the same results were obtained in an investigation of approximately 100 strains collected from the Americas, France, Czechoslovakia and the USSR over the past 60 years (Anxolabéhère *et al.* 1988b), as well as in an extensive survey of strains from the Americas, Europe, Africa, the Near and Far East and Australia (Kidwell *et al.* 1983). Boussy and Kidwell (1987) observed that of 620 strains tested worldwide, only 56 were not within 10% of either axis. Their conclusions were drawn from work done in Japan (Yamamoto *et al.* 1984; Hihara *et al.* 1985), in North America (Kidwell and Novy 1985) and in Australia (Boussy 1987; Boussy and Kidwell 1987).

The P factor activity for P, P? and P-Q intermediate strains varied from low values of 10,3% GD sterility in Nelspruit and Thohoyandou (Nel 2, Nel 19, and Tho 3) and 10,4% in Durban (Ber 15 and Ber 22) to values high in the 20s. There were also a few exceptionally high values, namely 34,5% (Nel 12 in Nelspruit), 36,2% (Har 6 in Harare), 36,5% (Rut 4 in Johannesburg), 37,0% (Tho 1 in Thohoyandou), 43,2% (P'B 1 in Pietersburg) and 52,8% (Nel 20 in Nelspruit). The P susceptibility of M, M? and M-Q strains as manifested in the form of cross A* GD sterility showed even more variability. For example in Walmer, Port Elizabeth, GD sterility in cross A* was 100% for Wal 3 and 10,3% for Wal 7.

2.3.1.2 Indefinite strains and intermediate strains

Of the 204 classifiable strains, seven were indefinite (two of type M? one of type Q? and four of type P?) and 41 were intermediate (9 of type M-Q, 6 of type Q-M, 14 of type P-Q and 12 of type Q-P). Berea (Durban) had the most intermediates (9); the rest were scattered fairly evenly over all collection sites. The indefinite strains showed negligible or no intrastrain/reciprocal cross sterility. Thus there was no interference background sterility. These crosses were classified as indefinite owing to the confidence regions overlapping the 10% GD sterility boundaries. It is possible that these strains were in a state of transition at the time of testing. The two M? strains showed too much P factor activity (within the 95% confidence region) to be classified as M. They could thus be in the process of changing from M to P (due to the recent acquisition of P elements), or they could be changing the other

way. Strains of the P? type are those whose confidence regions overlap the 10% GD sterility cross A* boundary. Thus the P? strains found were lacking sufficient regulatory power to be classified as P strains. In this respect, Kidwell (1986a) stated that strains with high P factor activity and low regulatory ability have been shown to exist transiently in the laboratory. These types of strains tend to evolve into a stable P, Q or M state or to go to extinction. Further, Boussy and Kidwell (1987) stated that such strains (showing elevated P activity and low regulatory ability) were found in the wild in 1983 and were later shown to have changed to a state like that of the norm of the population from which they came.

2.3.2 Polymorphisms and Trends

Table 2.3 indicates the number of P/P?, Q/Q?, M/M? and intermediate strains obtained at each collection site. Some of the collection sites showed clear patterns of either a central P, Q or M tendency whereas others tended to be more polymorphic.

The M type sites were At Ais, Cape Town and Port Elizabeth. Mainly Q types (which tended to predominate throughout) were found at Swakopmund, Durban (Addington Beach), Johannesburg (northern suburbs 1988), Pretoria, Hammanskraal and Agatha. The only P type site was Harare. The rest of the sites tended to be polymorphic. The P/Q polymorphisms were obtained in Durban (Berea), Nelspruit, Pietersburg and Thohoyandou. Two of the collection sites, Plettenberg Bay and Johannesburg (northern

Table 2.3 The numbers of P, Q, M and intermediate strains obtained at each collection site.

Collection Site	P/P?	P-Q	Q-P	Q/Q?	Q-M	M-Q	M/M?	TOTAL
Swakopmund				7	1		2	10
Gobabeb			1	9	1	2	5	18
A1 Afs				2			11	14
Cape Town				3	1		10	14
Gordons Bay							4	4
Plettenberg Bay	1			1			2	4
Port Elizabeth:								
Summerstrand				2		1	16	19
Walmer						1	5	6
Maclear				2		1	1	4
Durban:								
Berea	2	5	4	4			1	16
Addington Beach			1	15				16
Johannesburg:								
N.subs 1988	1		1	5		1		8
N.subs 1984/87	3					1		4
S.subs				2	2			4
Pretoria				3		1		4
Hammanskraal			1	3				4
Nelspruit	4	3		7	1		1	16
Pietersburg	2		1	1				4
Agatha				4				4
Thohoyandou	3	1	3	5				12
Harare	12	5		2				19
Total	28	14	12	77	6	9	58	204

20

suburbs 1984/1987) had complete mixtures (P, Q, M and M? in the former and P?, M-Q and ? in the latter). Since these sites had only four and five strains respectively, this conglomerate of types may not be a true picture of the real distribution. A total of 28 P/P? type strains, 77 Q/Q? type strains, 58 M/M? type strains and 41 intermediate strains were obtained.

2.3.3 Intrastrain and Reciprocal Crosses

Reciprocal crosses (cross B and cross B*) were performed on all strains collected in 1987. These crosses were not performed on strains collected the following year, however, but most of the 1988 collections were tested for intrastrain sterility at 29°C. The results of the reciprocal and intrastrain crosses are presented in Appendix B.

2.3.3.1 Reciprocal crosses

In most cases the reciprocal crosses (B and B*) showed low levels of background sterility. There were, however, four cases where the levels of gonadal dysgenesis were abnormally high. Rut 2 (as mentioned above) was not classifiable due to its high levels of GD sterility in both crosses A and A*. This strain may have been tested when it was in a transitional state as evidenced by the high level of background sterility in cross B*. Den 1, a Q strain, had a rate of GD sterility in cross B of 9.0%, possibly due to a low level of P element activity in the progeny when the females of this strain were crossed to M type males. GB 2 and GB 3 (both M strains) had levels of GD sterility in cross B of 13.8% and 6.0% respectively. However, the high levels

of GD sterility, in cross A* (26.0% and 65.0% respectively) ruled out any likelihood that this sterility was due to P element activity. It is possible that these higher than average levels of GD sterility in cross B stem from a general background of infertility when the females of these strains were exposed to laboratory males. This is supported by the fact that the level of GD sterility in the A* crosses of these two Gordons Bay lines was higher than that obtained in the A* test crosses with the other two Gordons Bay strains, GB 1 and GB 4.

2.3.3.2 Intrastrain crosses

Most lines showed no or negligible intrastrain sterility at 29°C. High levels of intrastrain sterility could indicate the initial spreading of P elements within that strain prior to regulatory ability becoming established. Cytotype switching tends to lag behind P activity induction (Daniels *et al.* 1987a; Kidwell *et al.* 1988; Good *et al.* 1989). For example, the introduction of P elements into M strain genomes by chromosomal contamination resulted in initial (generations 9 and 13) high frequencies of intrastrain GD sterility which by generation 60 was reduced to the normal low frequencies of sterility observed in established lines (Kidwell *et al.* 1988). Similarly high levels of intrastrain sterility due to initial P element contamination which rapidly fell over the generations were also found by Black *et al.* (1987). Daniels *et al.* (1987a) using P transformation techniques found significant levels of intrastrain sterility in lines into which P elements had been newly introduced.

On the other hand intrastain sterility may be due in part to the "atrophie gonadique" (*ag*) character (Périquet 1976) which is a quantitative genetic trait determined by genes located on chromosomes 2 and 3 (Périquet 1978). The *ag* character, like GD sterility, can affect either one or both gonads and is temperature sensitive (Périquet 1976). The relationship (if any) between *ag* and the P-M system is uncertain (Périquet 1980). Anxolabéhère *et al.* (1988a) tested a strain containing the *ag* character (Marseillan 65) by molecular hybridization techniques using a P element containing probe, and found no evidence of P elements in this strain.

Six of the collection sites had strains showing intrastain GD sterility greater than zero. Table 2.4 shows all these strains. The results of the cross A and cross A* GD sterility tests are also given.

Apart from Ais 14, which could not be classified (discussed above) five of the six strains showing GD intrastain sterility greater than 6% were all of type M, the sixth was of type Q. Of the eight A1 Ais strains tested, four had intrastain sterility greater than zero and three showed fairly high levels of sterility (14.6%, 11.9% and 8.0%) indicating the possible presence of the *ag* character or some other cause of sterility in that area.

Table 2.4

Strains showing intrastain sterility greater than zero, their origin, type and GD sterilities obtained in crosses A and A*

Strain	Origin	GD sterility						Type
		Intrastrain		Cross A		Cross B		
		n ¹	% GD sterility	n ¹	% GD sterility	n ¹	% GD sterility	
Swak 12	Gakopmund	58	5,1	150	0	64	0	Q
Gob 1	Gobabeb	66	3,0	172	1,2	160	3,7	Q
Gob 7	Gobabeb	68	5,9	150	0	176	1,1	Q
Gob 10	Gobabeb	94	8,5	150	0	182	30,8	M
Gob 13	Gobabeb	60	3,3	150	2,7	222	23,4	M
Gob 21	Gobabeb	60	3,3	150	0	144	44,4	M
Ais 7	Ai Ais	84	11,9	136	1,5	156	29,5	M
Ais 8	Ai Ais	124	6,4	160	0	172	15,1	M
Ais 14	Ai Ais	82	14,6	158	10,1	118	31,5	?
Ais 16	Ai Ais	50	8,0	100	0	136	26,5	M
SS 4	Port Elizabeth	78	5,1	68	2,9	100	72,0	M
SS 10	Port Elizabeth	74	2,7	114	0	76	84,2	M
Ber 4	Durban	108	0,5	74	6,7	118	1,7	Q-P
Ber 6	Durban	128	21,1	118	0	98	38,8	M
Ber 8	Durban	96	4,2	158	17,7	76	2,6	P
Ber 15	Durban	96	4,2	134	10,4	84	0	P-Q
Ber 18	Durban	100	2,0	128	0	104	0	Q
Ber 19	Durban	100	2,0	134	11,9	108	0	P-Q
Nel 4	Nelspruit	100	1,0	228	14,0	128	0	P-Q
Nel 5	Nelspruit	100	1,0	220	1,8	214	0,9	Q
Nel 18	Nelspruit	100	2,0	268	0	178	31,5	M

¹n is the number of ovaries examined

A strain from South Africa called Limpopo (so presumably originating from somewhere near the Limpopo river which forms the boundary between South Africa and Zimbabwe) was found to have high intra-strain sterility (19% in females at 29°C) which appeared to occur independently of the high GD sterility produced in a P-M (cross A) reference cross (Daniels *et al.* 1987b). Ai Ais is in the Fish River Canyon and it is tempting to speculate that there may be some connection between the two. However, the two rivers (Fish and Limpopo) are separated by the Kalahari desert and so it is probably merely coincidental that high intra-strain GD sterility was observed in lines collected from these two river sites.

Five strains from Gobabeb revealed intrastrain GD sterility, but three of them were negligibly low. The other two were 5.9% for Gob 7 (a Q strain) and 8.5% for Gob 10 (an M strain). The Gobabeb collection site produced a mixture of Q and M type strains --so this GD sterility may be due to incipient P element activity. Stronger evidence for this is presented by the Ber 6 line. This is the line which showed the highest percentage of GD intra-strain sterility (21.9%). Ber 6 was the only M type strain present in the Berea collection. The other 15 lines from Berea were of type P or Q and the lines from Addington Beach, the other Durban site, were all of type Q. This might indicate that this intrastrain GD sterility was due to P element activity in a strain which was changing from M to P and had not yet developed its regulatory facilities.

2.3.4 Geographical Distribution

Fig. 2.4 shows the distribution of the 204 Southern African strains which were classifiable in the P/M/Q system. For this figure indefinite strains P? and M? were classified as P and M respectively while the intermediate strains M-Q, Q-M, Q-P and P-Q were classified in M, Q, Q and P respectively. Lines from Namibia in the west were all of type Q or M. There was no evidence of strong P activity in that area. There was also no evidence of P activity along the south coast of South Africa i.e. in Cape Town, Gordons Bay, Plettenberg Bay or Port Elizabeth with the exception of one strain from Plettenberg Bay (Plet 10) collected in 1987. To the east, P activity was first apparent in Durban (Berea) although that area is mainly of type Q. To the north three P strains were found in Johannesburg in 1987 but in the 1988 collection only one P strain was present. The 1987 collections in Pretoria and Hammanskraal were mainly Q. The number of P strains gradually increased in a northerly direction and on the eastern side between 25°S and 17°S, only one M strain was found to be present (in Nelspruit at 24.98°S). In Harare, the northern most collection point (17.83°S), 89% of the strains were of type P.

2.3.4.1 Clines

The mean GD sterilities obtained in crosses A and A* were evaluated for each collection site (Table 2.5). In order to detect any significant geographic patterns of distribution the Spearman rank correlation statistic on mean cross A and cross A* GD steri-

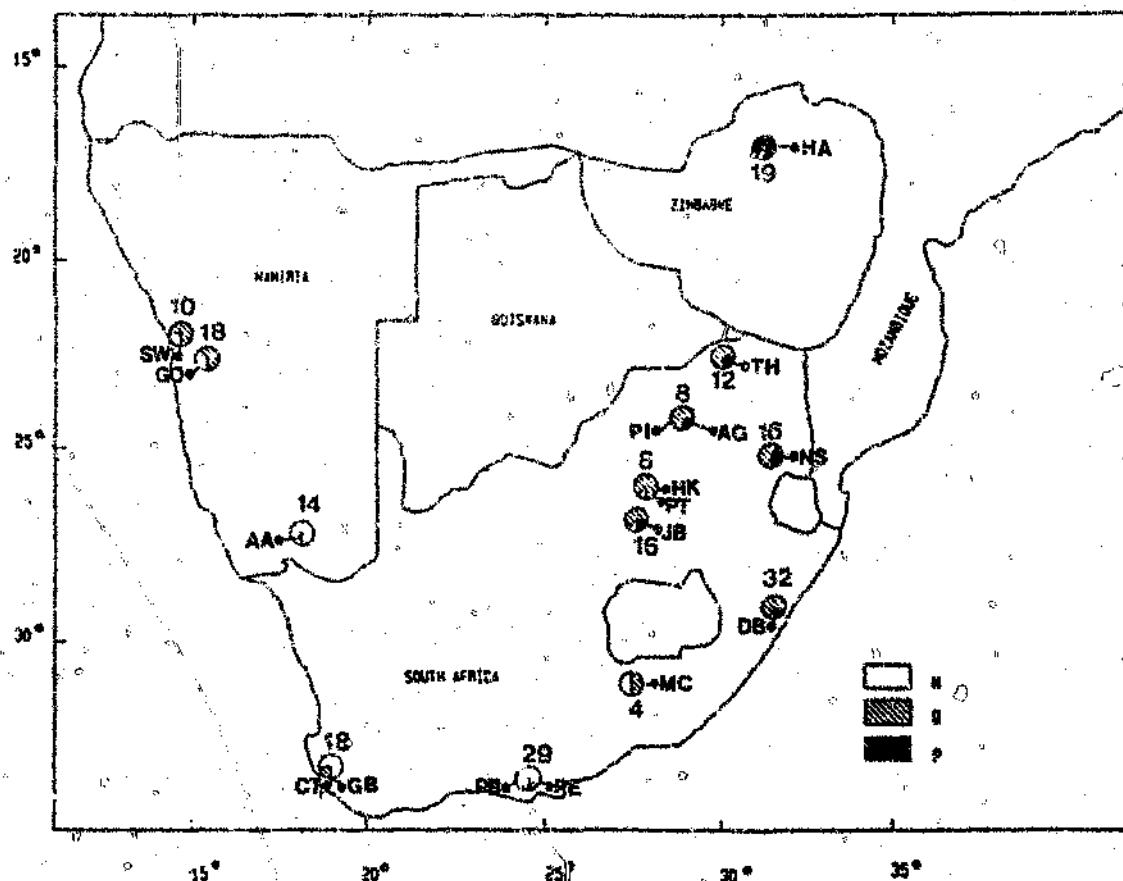


Fig. 2.4

Distribution of *D. melanogaster* populations sampled in Southern Africa. Numbers correspond to number of isofemale lines tested in each area.

For the sake of clarity, frequency pie diagrams have been amalgamated for the following collection sites:

- (i) Cape Town and Gordons Bay
- (ii) Plettenberg Bay and the two Port Elizabeth collection sites Summerstrand and Walmer.
- (iii) Berea and Kiddington Beach in Durban
- (iv) Northern suburbs 1984, 1987, 1988 and southern suburbs in Johannesburg.
- (v) Pretoria and Hammanskraal
- (vi) Pietersburg and Agatha

Abbreviations for collection sites are:

SW: Swakopmund; GO: Gobobab; AA: Ai Ais; CT: Cape Town; GB: Gordons Bay; PB: Plettenberg Bay; PE: Port Elizabeth; MC: Maclear; DB: Durban; JB: Johannesburg; PT: Pretoria; HK: Hammanskraal; NS: Nelspruit; PI: Pietersburg; AG: Agatha; TH: Tlokoeng; HA: Harare.

Table 2.5 Mean GD sterilities obtained in crosses A and A* for 21 collection sites in Southern Africa

Collection site	Number of isofemale lines	GD sterility	
		Cross A	Cross A*
Swakopmund	10	0	0,118 $\pm$ 0,174
Gobabeb	18	0,010 $\pm$ 0,010	0,132 $\pm$ 0,181
At Ais	15	0,015 $\pm$ 0,016	0,433 $\pm$ 0,180
Cape Town	14	0	0,328 $\pm$ 0,128
Gordons Bay	4	0,003 $\pm$ 0,008	0,341 $\pm$ 0,287
Plettenberg Bay	4	0,100 $\pm$ 0,175	0,208 $\pm$ 0,336
Port Elizabeth:			
Summerstrand	19	0,005 $\pm$ 0,007	0,565 $\pm$ 0,137
Walmer	6	0,004 $\pm$ 0,013	0,614 $\pm$ 0,391
Maclear	4	0,003 $\pm$ 0,007	0,078 $\pm$ 0,126
Durban:			
Berea	16	0,083 $\pm$ 0,036	0,033 $\pm$ 0,051
Addington Beach	16	0,007 $\pm$ 0,011	0
Johannesburg:			
N.subs 1988	8	0,029 $\pm$ 0,050	0,024 $\pm$ 0,035
N.subs 1987/1984	5	0,198 $\pm$ 0,158	0,099 $\pm$ 0,045
S.subs	4	0,012 $\pm$ 0,017	0,063 $\pm$ 0,067
Pretoria	4	0,004 $\pm$ 0,011	0,050 $\pm$ 0,101
Hammanskraal	4	0,037 $\pm$ 0,051	0
Nelspruit	16	0,109 $\pm$ 0,076	0,026 $\pm$ 0,042
Pieterburg	4	0,195 $\pm$ 0,233	0
Agatha	4	0,024 $\pm$ 0,036	0,005 $\pm$ 0,014
Thohoyandou	12	0,100 $\pm$ 0,067	0
Harare	19	0,208 $\pm$ 0,044	0,003 $\pm$ 0,007

For each mean, 95% confidence intervals are given.

lity was used to test for associations with latitude and longitude. A significant negative correlation ($r_s = -0,474$; $P < 0,05$) for latitude and a significant positive correlation for longitude ($r_s = 0,507$; $P < 0,05$) was obtained with cross A GD sterility. Thus P factor activity increased as the latitude decreased and the longitude increased in a northeasterly direction. As far as regulatory ability was concerned, a highly significant positive correlation ($r_s = 0,609$; $P < 0,005$) for latitude and a negative correlation for longitude ($r_s = -0,773$; $P < 0,001$) with cross A* GD sterility was obtained. So regulatory ability increased (cross A* GD sterility decreased) with decreasing latitude and increasing longitude i.e. in a northeasterly direction. Thus P factor activity and regulatory ability increased together (as would be expected) in a northeasterly direction.

Clinal patterns on both smaller and larger geographic scales than the present one have been observed elsewhere in the world. For example, investigations in the Languedoc-Roussillon area of France covering a distance of 250 km revealed a clinal pattern with weak P populations in the northern zone, Q populations in the centre and M' populations in the south (Anxolabéhère *et al.* 1988a). On a larger scale, a latitudinal cline along the east coast of Australia from Cairns (latitude $16,9^\circ\text{S}$) in the north to Cygnet ($43,2^\circ\text{S}$) in the south (a distance of approximately 3100 km) comprising P in the north Q in the centre and M type populations in the south was found to exist (Bous and Kidwell 1987).

A west-east analysis of 27 strains sampled from France in Western Europe to Kirghizistan in the USSR revealed a clear gradient from Q type strains to M type strains (Anxolabéhère *et al.* 1985b). Geographical patterns worldwide and hypotheses regarding the reasons for these distributions are discussed in Section 2.4.4.1.

2.4 Discussion

The pattern of geographical distribution for Southern Africa (Fig.2.4) revealed a preponderance of Q type strains. Of the 206 strains tested, 75 were unambiguously Q type strains and another 41 were intermediate Q type strains. The rest consisted of 56 M type strains, 24 P type strains and 9 strains which were of type M?, Q? or P? or were unclassifiable.

Of the 21 collection sites examined, 12 showed a tendency towards either M, P or Q type dominance. The other 9 collection sites were polymorphic in nature, mainly P/Q or Q/M type polymorphisms. Such polymorphisms have been noted in the wild in populations around the world. For example in North America (Kidwell and Novy 1985) and in Australia (Boussy and Kidwell 1987). Indeed an extreme P/M cytotype polymorphism was shown to be stable over a period of five years in Nasr' Allah (an oasis in South Tunisia) (Anxolabéhère *et al.* 1982b; Izaabel *et al.* 1987).

A clinal pattern was evident in the Southern African distribution in that the P strains tended to dominate in the northeast and M strains were present in large numbers in the southern areas. The samples from Namibia in the west, consisted of M or Q type strains.

A significant correlation was obtained with both latitude and longitude with cross A GD sterility and a highly significant correlation with latitude and longitude and regulatory ability (cross A* GD sterility) indicating an overall northeastern to southwestern decrease in regulatory ability and P factor activity. Knowledge regarding the origins of P elements their dynamics and the mechanisms of P element regulation may help to reveal the reasons for this observed pattern.

2.4.1 Origins of P elements

D. melanogaster is a member of the *melanogaster* subgroup of the *melanogaster* species group of the subgenus *Sophophora*. Many dispersed repetitive DNAs including *copia*, 297 and 412 which are found in *D. melanogaster* have also been found to be present in two of the sibling species, *D. simulans* and *D. mauritiana* (see, for example, Young and Schwartz 1981; Dowsett and Young 1982; Dowsett 1983; Martin *et al.* 1983; Brookfield *et al.* 1984). However DNA hybridization studies have failed to detect any P homologous sequences in the sibling species *D. simulans*, *D. mauritiana*, *D. yakuba*, *D. teissieri*, *D. areolaris* or *D. erecta* (Brookfield *et al.* 1984; Anxolabéhère *et al.* 1985a). On the other hand, P homologous sequences have been shown to occur in four of the species groups within the subgenus *Sophophora*. A non-Sophophoran species from the genus *Drosophila* which showed homology to P elements was *D. nasuta* from the subgenus *Drosophila* (Stacey *et al.* 1986). Even further afield, DNA from a species from the closely related genus *Scaptomyza* (*S. pallida*) was found to hybridize positively with a P element probe (Anxolabéhère *et al.* 1985a). A listing

of the species of the Sophophoran subgroup which showed positive hybridization with P element probes and those which tested negative are given in Tables 2.6A and B (Brookfield *et al.* 1984; Daniels *et al.* 1984; Anxolabéhère *et al.* 1985a; Lansman *et al.* 1985; Daniels and Strausbaugh 1986; Stacey *et al.* 1986). To try to assess the significance of these findings it is interesting to look at them in the light of the phylogenetic development of the Sophophoran species.

The Sophophoran radiation, which is the second of five successful radiations of the Drosophilidae and is thought to have occurred before the end of the Eocene Epoch, consists mainly of four groups: the *saltans* and *willistoni* lineages in the New World tropics, the *melanogaster* lineage in the Old World tropics and the *obscura* group which is more widely distributed than the other three groups, being found in both the New World and Old World north temperate zones i.e. the Holarctic temperate zones (Throckmorton 1975).

The *saltans* and *willistoni* groups are the products of a single Sophophoran lineage and the primary tropical disjunction of the *melanogaster* group in the Old World tropics and the *saltans-willistoni* lineage in the New World tropics is hypothesized by Throckmorton (1975), on the basis of geographic distribution and continental drift, to have taken place in Oligocene times. On the other hand, Beverley and Wilson (1984), using the evolutionary changes in a larval hemolymph protein (LHP), estimated the time of divergence of the *melanogaster* and *willistoni* lineages to be 53 million years (MYR) ago. The *obscura* group is most

Table 2.6

A : Species of the subgenus *Sophophora* containing P homologous sequences

Species group	Subgroup	Species
<i>melanogaster</i>	<i>melanogaster</i> <i>montium</i> <i>fiensphila</i> <i>suzukii</i>	<i>melanogaster</i> * <i>kikkawai</i> ^{4, 6} * <i>fiensphila</i> ^{4, 6} <i>lucipennes</i> ⁶
<i>obscura</i>	<i>obscura</i>	<i>maderensis</i> ³ <i>subobscura</i> ³ <i>bifasciata</i> ³ * <i>pseudo-obscura</i> ^{3, 4, 6}
	<i>affinis</i>	<i>affinis</i> ^{4, 5} <i>helvetica</i> ³
<i>willistoni</i>	<i>willistoni</i>	<i>willistoni</i> ^{4, 5, 6} <i>equinoxialis</i> ^{4, 5, 6} <i>tropicalis</i> ^{4, 5, 6} <i>pauistorum</i> ^{2, 4, 5, 6} <i>pavlovskiana</i> ⁵
	<i>basatinensis</i>	<i>nebulosa</i> ^{4, 5, 6} <i>succinea</i> ^{4, 5, 6} <i>capricornii</i> ^{4, 5, 6} <i>unipennis</i> ^{5, 6}
<i>saltans</i>	<i>saltans</i>	<i>saltans</i> ^{4, 5, 6} <i>australo-saltans</i> ^{4, 5, 6} <i>prosaltans</i> ^{4, 5, 6} <i>lusaltans</i> ⁵
	<i>sturtevantii</i>	<i>sturtevantii</i> ^{4, 5, 6} <i>milleri</i> ⁵

1. Brookfield *et al.* 19842. Daniels *et al.* 19843.† Anxolabéhère *et al.* 1985a4. Lansman *et al.* 1985

5. Daniels and Strausbaugh 1986

6. Stacey *et al.* 1986

* Different researchers reported differing results for these species.

† In this study two different P elements probes were employed. Only species which hybridized positively with both probes are listed in A.

Table 2.6 **B : Species of the subgenus *Sophophora* which did not show positive hybridization with P element probes**

Species group	Subgroup	Species
<i>melanogaster</i>	<i>melanogaster</i>	<i>similans</i> ^{1, 3}
		<i>mauritanica</i> ^{1, 3}
		<i>yakuba</i> ^{1, 3}
		<i>teissieri</i> ^{1, 3}
		<i>orena</i> ³
		<i>erecta</i> ³
	<i>montium</i>	<i>locqueti</i> ³
		<i>burlai</i> ³
		* <i>kikkawai</i> ³
		<i>vulkana</i> ³
<i>obscura</i>	<i>ficusphila</i>	* <i>ficusphila</i> ³
	<i>ananassae</i>	<i>ananassae</i> ^{3, 4}
		<i>malerkotliana</i> ³
	<i>elegans</i>	<i>elegans</i> ^{3, 4}
	<i>eugracilis</i>	<i>eugracilis</i> ^{3, 4}
	<i>takahashi</i>	<i>takahashi</i> ⁴
	<i>obscura</i>	* <i>pseudoobscura</i> ¹
	<i>willistoni</i>	<i>insularis</i> ⁵
	<i>cordata</i>	<i>neocordata</i> ^{4, 5, 6}
		<i>emarginata</i> ^{4, 5, 6}
<i>salans</i>	<i>elliptica</i>	<i>neoelectica</i> ⁵

1. Brookfield *et al.* 1984

2. Daniels *et al.* 1984

3. Anxolabéhère *et al.* 1985a

4. Lansman *et al.* 1985

5. Daniels and Strausbaugh 1986

6. Stacey *et al.* 1986

* Different researchers reported differing results for these species.

† In this study two different P element probes were employed. Only species which hybridized positively with both probes are listed in A.

Table 2.6

B : Species of the subgenus *Sophophora* which did not show positive hybridization with P element probes

Species group	Subgroup	Species
<i>melanogaster</i>	<i>melanogaster</i>	<i>simulans</i> ^{1, 3}
		<i>mauritiana</i> ^{1, 3}
		<i>yakuba</i> ^{1, 3}
		<i>teissieri</i> ^{1, 3}
		<i>orena</i> ³
		<i>erecta</i> ³
	<i>montium</i>	<i>bocqueti</i> ³
		<i>burlai</i> ³
		* <i>kikkawai</i> ³
		<i>vulkana</i> ³
	<i>fioussphila</i>	* <i>fioussphila</i> ³
	<i>ananasae</i>	<i>ananasae</i> ^{3, 4}
		<i>malerkotliana</i> ³
	<i>elegans</i>	<i>elegans</i> ^{3, 4}
	<i>eugracilis</i>	<i>eugracilis</i> ^{3, 4}
	<i>takahashi</i>	<i>takahashi</i> ⁴
<i>obscura</i>	<i>obscura</i>	* <i>pseudoobscura</i> ¹
<i>willistoni</i>	<i>willistoni</i>	<i>insularis</i> ⁵
<i>saltans</i>	<i>cordata</i>	<i>neocordata</i> ^{4, 5, 6}
	<i>elliptica</i>	<i>emarginata</i> ^{4, 5, 6} <i>neoelectica</i> ⁶

1. Brookfield *et al.* 1984

2. Daniels *et al.* 1984

3. Anxolabehère *et al.* 1985a

4. Lansman *et al.* 1985

5. Daniels and Strausbaugh 1986

6. Stacey *et al.* 1986

* Different researchers reported differing results for these species.

† In this study two different P element probes were employed. Only species which hybridized positively with both probes are listed in A.

closely related to the *melanogaster* group, and is believed to have originated together with this group from a common proto-*melanogaster* lineage in the Old World tropics (Throckmorton 1975). A molecular estimate of 46 MYR ago as the divergence time of the *obscura* and *melanogaster* groups was obtained by Beverley and Wilson (1984).

In many cases the patterns of DNA hybridization obtained in the *willistoni* and *saltans* groups reflect phylogenetic relationships (Daniels and Strausbaugh 1986) thus suggesting that these sequences have a long evolutionary history in these groups. Further the presence of P homologous sequences in the three species subgroups *willistoni*, *saltans* and *obscura* of the subgenus *Sophophora* would seem to indicate the presence of P elements in these groups prior to the Sophophoran radiation. However, certain anomalies present themselves under this hypothesis, namely the discontinuous and patchy spread of P elements throughout all the species of the subgenus *Sophophora* and the difference in the conservation of the P elements present in certain species. For example P homologues from the *melanogaster* species were only weakly detectable, whereas those from the *willistoni* and *saltans* groups showed most similarity to P elements and the *obscura* group species lay somewhere between these two (Lansman et al. 1985; Stacey et al. 1986). Thus the most highly conserved P element homology (relative to *D. melanogaster*) occurred in the *willistoni* and *saltans* species groups which are more distantly related to *D. melanogaster* than the species in the

melanogaster and *obscura* groups. These patterns seem to suggest that P element invasions have occurred at different times and with different strengths, perhaps over a long period of evolutionary time. It is possible that certain species of the *Sophophora* subgenus are particularly susceptible to these invasions, while others are not. The source and mechanism of these invasions is still unknown. Engels (1986) has hypothesized that transfer of a transposable element from one species to another could be brought about by a virus in that the transposable element present in the one species could be integrated into the viral genome (Miller and Miller 1982) and subsequently transferred into another of the virus host species. Whatever the mechanism, if the transfer takes place between species, then one would expect it to have occurred between species whose geographic ranges overlap. The ranges of the *saltans*, *willistoni* and *obscura* groups overlap in South and Central America, where also are found the cosmopolitan *D. melanogaster* and the near cosmopolitan P element containing species of the *montium* subgroup *D. kikukawai* (Throckmorton 1975). Further, the ranges of many of these species extend into areas of North America which today contain many P strains of *D. melanogaster*. Daniels and Strausbaugh (1985) state that "some of the members of the *willistoni* species group (namely *D. nebulosa*, *D. willistoni* and *D. sucinea*) have ranges that extend into the southern portion of North America, the geographical area in which the strongest early P-factor activity is suspected to have occurred in *D. melanogaster*". Although the *willistoni* group is mainly neotropical, *D. willistoni* extends further north up to central Mexico and Florida whereas

D. nebulosa extends to Texas and may even be found in Nebraska (Ehrman and Powell 1982). These are two members of the *willistoni* species group which contain P homologous sequences (Table 2.6A).

According to Lakovaara and Saura (1982), the *obscura* group is very widespread with ranges overlapping *D. melanogaster* worldwide. For example both *D. pseudoobscura* and *D. affinis* which displayed homology with P elements are found on the North American continent. The range of *D. pseudoobscura* extends from British Columbia in the northern part of the American continent to Guatemala in the south. The species has also been found in places like Texas and Nebraska. *D. affinis* is distributed over the Eastern United States from the Canadian border. In the west its range extends to the Nebraska-Wyoming border.

This evidence seems to point to an infectious (horizontal) transmission rather than a hereditary (vertical) transmission of P elements. This hypothesis was favoured by Stacey *et al.* (1986) after they had calculated the divergence rate of P elements on the basis of their acquisition by vertical transmission, and found it to be 7-19 times higher in some of the *melanogaster* species than in the *willistoni*, *saltans* or *nasuta* species. These authors suggest that the source for P elements may originate in South America. They point out however that this hypothesis does not account for the presence of P elements in *D. nasuta*, an old-world species. Since *D. nasuta* is afrotropical and has been found in Benin, the Congo and Zaire as well as several Indian Ocean islands (Tsacas *et al.* 1981), it is possible that

this species is directly responsible for the presence of P elements in *D. melanogaster* in sub-Saharan Africa, in that a horizontal transmission took place on the African continent between these two species in the same way as it is hypothesized to have arisen on the American continent between the *willistonii-saltans* species group and *D. melanogaster*. The other possibility is that the transmission occurred in the other direction i.e. that *D. melanogaster* in Africa obtained its P elements via gene exchange with western or eastern P element contaminating *D. melanogaster* strains and that a horizontal interspecies transmission between African drosophilids was responsible for *D. nasuta* acquiring the P elements. It may be, as mentioned previously, that *D. nasuta* is one of those species which is particularly susceptible to P element invasion. Finally *S. pallida*, the species from outside the genus *Drosophila* which was found to contain P hybridizing sequences (Anxolabéhère et al. 1985a), is a cosmopolitan species and so could certainly be considered as a potential source of P element transfer from outside the genus.

2.4.2 Dynamics of P elements

There are at present two (not necessarily conflicting) theories as to why P elements are found only in wild strains and not in older laboratory strains.

The recent loss hypothesis was originally proposed by Engels (1981a) who suggested that the lack of P elements in long held laboratory stocks was due to random drift in small laboratory

populations and that in fact, P elements have been present in *D. melanogaster* for an exceedingly long time. Experimental evidence for this theory was obtained by Hihara *et al.* (1985) who found type conversions to occur in the laboratory in isofemale lines recently established from wild caught flies. These conversions were from P and Q type strains to M type strains. No spontaneous changes to P type ever occurred although type conversions from M to Q did occur. (These were explained by assuming that a few P or "Q" factors initially present in the M type strains, increased through transposition). Unfortunately, strain classifications were performed by ovarian dissection and no molecular hybridization studies were performed. Thus it is not possible to know from these results whether the M strains were entirely free of (degenerate) P elements.

Engels (1986) pointed out the problems associated with his (Engels, 1981a) stochastic loss theory. Essentially the amount of time that was thought to have elapsed since the original setting up of laboratory stocks would seem to be much less than the theoretical required time for a random loss process to have taken place (Kaplan *et al.* 1985). Also since the majority of P elements in a genome are the non-autonomous type (O'Hare and Rubin 1983), the likely outcome of a stochastic loss process would be M populations, since once all autonomous elements have been removed, there would be no more transposase function which is essential for excision (Karess and Rubin 1984).

The rapid invasion hypothesis (Kidwell 1979, 1983) proposes that P elements were only introduced to or became common in *D. melanogaster* from about 1950. Since then they have spread rapidly and increased in frequency in the wild owing to the high efficiency of replicative transposition with or without positive or negative selection. However Engels (1986) points out that this hypothesis implies an unreasonably high rate of acquisition of new transposition families over evolutionary time.

The I factors are a second family of transposable elements in *D. melanogaster* with similar properties to those of the P family. In particular, laboratory strains collected in the 1920s were all shown to lack I elements, whereas, more recently collected wild strains have the I element present in them. Thus, at least two invasions are known in this century and if one extrapolates backwards in time to one million years ago, since the minimal time of the divergence of *D. melanogaster* from *D. simulans* (which lacks P and I elements), then *D. melanogaster* would have gained at least 20 000 families of transposable elements. Since the entire genome is estimated to contain only 20-50 transposable element families (Engels 1981a; Rubin 1983), this is clearly not feasible. A counter argument is that elements might be gained and subsequently lost from the genome, but in that case a mechanism for ridding the genome of all elements of a given family must be found.

Studies in temporal and geographic distributions of P elements in *D. melanogaster* worldwide support an invasion theory. In America, strains collected before 1960 were predominantly of type M; P and Q type strains began to appear during the 1960s and today there appear to be very few M type strains in America (Kidwell 1983; Kidwell et al. 1983; Kidwell and Novy 1985).

In Europe, strains collected prior to 1970 are mainly of type M. Today the Q populations appear to predominate with the odd scattering of P type strains which appear to decrease in frequency as one moves eastwards (Kidwell et al. 1983; Anxolabéhère et al. 1984, 1985b, 1988b).

In Japan and elsewhere in the Far East, strains were mainly M type before 1970. Today P and Q populations predominate there (Ohishi et al. 1982; Kidwell et al. 1983; Yamamoto et al. 1984). A recent investigation of Australian populations (Boussy 1987; Boussy and Kidwell 1987) has revealed a clinal pattern along the eastern seaboard from P type strains in the north to M type strains in the south.

Africa today is populated by predominantly M type strains in the north, and by P/Q type strains in the central sub-Saharan region (Anxolabéhère et al. 1982a, 1984; Woodruff et al. 1989).

In South Africa, one strain from Graaff-Reinet (in the Cape) collected in 1949 and five strains from Cape Town collected in 1954 were all of type M. (Kidwell *et al.* 1983; personal communication M G Kidwell). The Cape is the area where the largest collection of M type strains were obtained in this study (Cape Town, Gordons Bay and Port Elizabeth). The earliest signs of P elements in South Africa appeared in 1963 in a strain collected from the Magaliesberg (approximately 80 km west of Johannesburg). This strain was of type Q (Kidwell *et al.* 1983; personal communication M G Kidwell).

Mention is made in Daniels *et al.* (1987b) of two true M strains collected in the early 1970s. One is called Kalahari and the other Limpopo. These names presumably indicate the geographic origins of these strains. However, these names refer to two vast areas and the specific locations of the collection points are not given. Further north in Botswana, 30 isofemale lines collected in 1985 all contained P elements (Eanes *et al.* 1988).

The above information from past collections together with this study suggests a possible southward radiation of P elements, perhaps originating in central Africa and spreading down through Zambia, Botswana and Zimbabwe to the northern and eastern parts of South Africa. In the west, it is possible that the Namib and Kalahari deserts formed barriers to the spread of P elements. This could then explain the lack of P type strains in places like Gobabeb and Swakopmund which are in the same latitudinal regions as the Northern Transvaal where P type strains were found (Pietersburg and Thohoyandou).

This pattern is not inconsistent with a theory of a worldwide invasion of P elements, for which the following scenario has been suggested (Anxolabéhère *et al.* 1988b). The invasion started in the Americas in the 1950s and spread to most regions of these continents during the 1960s and 1970s. During this period (the 1960s and 1970s) the eastward bound P element invasion of Europe was initiated. It seems likely that, also in the 1960s or possibly earlier, an invasion of central Africa and central Australia took place. In Africa the Sahara desert prevented the northward spread of these elements. However there were no barriers to a southern and eastern movement, and they are gradually spreading in those directions. This would then account for the present distribution of P elements.

2.4.3 Regulatory Ability

This study has revealed the presence of a clinal pattern in both P factor activity and regulatory activity of Southern African strains of *D. melanogaster*. There is a highly significant increase in regulatory ability together with a significant increase in P factor activity both of which occur in a south-westerly to northeasterly direction. In order to analyse the reasons for this it is necessary to examine the various modes of regulation of P factor activity.

2.4.3.1 Autonomous P elements involved in regulation

Engels (1979b) suggested the cytotype mode of regulation. Nearly all females could be classified as P or M cytotype on the basis of their ability to produce sterile daughters. Thus daughters of M cytotype mothers were susceptible to the sterilizing effects of P strain chromosomes whereas daughters of the P cytotype mothers were immune. The rules of inheritance of cytotype appeared to be a complex mixture of both cytoplasmic and chromosomal transmission.

Engels (1981a) suggested that an extrachromosomal element capable of self replication to some extent and homologous to certain chromosomal sequences, possibly P factors, may be present in either M or P cytotype. Presumably, such episomes could be generated by precise excision of chromosomal P factors. O'Hare and Rubin (1983) have suggested a mechanism whereby circular P elements containing the 8 bp host target DNA could be generated (described in Section 1.1.3.2).

Simmons and Bucholz (1985), elaborating on the hypothesis that non-chromosomal P elements are important in the determination of cytotype, suggested that these extrachromosomal P elements might titrate the transposase produced by intact chromosomal P elements in that the transposase binds to the terminal repeats of defective (or transcriptionally inactive) extrachromosomal

P elements. Thus, in the presence of a small amount of transposase and a large number of extrachromosomal elements, a shortage of transposase would occur leading to a reduction of P mediated transposition events. On the other hand, Sved (1987) in an experiment which replaced the entire P chromosome complement by M strain chromosomes demonstrated the absence of true maternal inheritance for P cytotype. The author thus argued against P factor repression being dependent on the autonomous replication of non-chromosomal elements. However the author noted that repression could still be explained by the presence of non-chromosomal copies of the P elements provided that these copies were continually replenished from chromosomal P elements.

An alternative suggestion to the episomal model was put forward by O'Hare and Rubin (1983). They suggested that the 2.9 kb P elements code for two functions, a transposase and a regulator that can lead to the suppression of transposition. A feedback mechanism ensures the maintenance of P and M cytotype as follows: In P cytotype, transposition is prevented by the amount of regulator present being such that it negatively feeds back on transposase and positively feeds back on itself and, in so doing, stimulates its own activity. When P elements are introduced into the M cytotype there is insufficient regulator for efficient negative feedback on transposase and so transposition takes place. However, accumulation of the regulator eventually (after one or more generations) leads to a sufficient quantity of regulator building up to ensure positive feedback on itself. Consequently the change from M to P cytotype occurs. Possible

candidates for the transposase and regulator proteins are an 87 000 dalton protein and a 66 000 dalton protein derived from different P factor transcripts, the former polypeptide being derived from a transcript with all three introns removed, and the latter being derived from the same transcript but with the third intron in place (Rio *et al.* 1986).

2.4.3.2 KP and other deletion derivatives

All of the above models suggested a maternal mode of inheritance and the participation of the complete P element. The existence of M' strains which apparently lack autonomous P elements and yet are able to repress hybrid dysgenesis to a degree (Kidwell 1985) suggests that alternative modes of regulation may be taking place. In a genetic analysis of M' strains, Kidwell (1985) showed that suppression of GD sterility was chromosomally based and transmitted through both sexes.

The concept of a mutant transposase (derived from defective P elements) that binds to autonomous P elements and thus inhibits their transcription or alternatively blocks their access to normal transposase has been put forward by Simmons *et al.* (1987) who found that partial suppression of P element activity occurred in a strain which lacked complete P elements and transposase activity. Further experimental evidence of a truncated transposase repressor was obtained in Japan: Sakoyama *et al.* (1985) isolated a 2.1 kb element from a Q strain which had the same DNA sequence as that of the complete P element, apart from a missing central

region and a trivial unsequenced region of 38 bp. Defective P elements (deleted between bases 1991 and 2448) were also present in a Q strain of Japanese origin WY113 (Nitasaka *et al.* 1987). Since Q strains have P cytotype, the role of defective P elements as encoders of repressor proteins was consequently hypothesized. Consistent with this hypothesis are the findings of Robertson and Engels (1989). These authors identified a product of two modified P elements which alters the expression of cytotype dependent alleles and which can repress P elements mobility. These modified P elements each contain mutations which affect the fourth exon; one is a frameshift mutation in exon 3, the other is a mutation in the splice junction of the 2-3 intron. The protein which could be produced by these defective P elements would presumably be the same as, or similar to the 66 000 dalton protein (containing 576 amino acids) observed by Rio *et al.* (1986) to have been encoded by the mRNA retaining the third intron.

Finally, a shorter 207 amino acid polypeptide has been hypothesized to be involved in hybrid dysgenesis repression (Black *et al.* 1987). This polypeptide is thought to be encoded by the KP element found to be present in natural M' strains. KP elements result from an internal deletion of 1753 nucleotides (nucleotides 808 to 2560) of the complete P element and flies carrying KP elements and P factors contain an abundant 0.8 kb transcript. This corresponds to the expected transcript of the KP element and is capable of being translated into a 207 amino acid polypeptide (Black *et al.* 1987). The above-mentioned authors hypothesized

that the KP elements constitute a system of suppression of hybrid dysgenesis in natural populations. They assayed for the regulation of hybrid dysgenesis by scoring the GD sterility induced in crosses of KP containing strains to P cytotype and M cytotype strains and also by scoring intrastain sterility at 29°C. They found that GD sterility in Cross A was lower in the presence of KP elements. Further they found that the inheritance of KP induced repression was transmitted through both sexes as opposed to the maternal inheritance of repression in lines lacking KP elements. Black *et al.* (1987) suggested that the putative protein encoded by the KP element might interfere with transposition, not by competing with transposase since it lacks the amino acid region believed to be involved with DNA binding (Rio *et al.* 1986) but by competing for a co-factor required for P transposition, or by forming a nonfunctional heteropolymer with P factor encoded peptides.

Up to 30 copies of KP elements have been shown to occur in M' and Q populations sampled from various parts of the world (Black *et al.* 1987). The strong cline observed in regulatory ability in Southern African populations (as opposed to the weaker cline exhibited in P factor activity) strongly suggests that something other than P elements on their own are responsible for the observed pattern. A south bound wave of KP (or some other deleted P) elements in conjunction with that of autonomous P elements would explain the difference of the clinal pattern strength of the two effects. Indeed, the presence of autonomous P elements has been shown to increase the KP copy number and subsequent repression of P factor activity (Black *et al.* 1987; Jackson *et al.* 1988).

2.4.4 Movement of P elements

Laboratory experiments have been used to monitor the spread of P elements in populations. These experiments have indicated that, although in general P element contamination takes place swiftly and efficiently, the end results are often dependent on the initial conditions of the population as well as on the ambient temperature.

2.4.4.1 Initial conditions and P element dynamics

Populations initiated with mixed P and M type strains evolved rapidly towards the P type (Kidwell *et al.* 1981; Kiyasu and Kidwell 1984; Good *et al.* 1989). Kidwell *et al.* (1981) found, however, that with initial P/Q or Q/M mixed cultures the changes were not so clearcut. Populations that were initially of mixed P and Q types also attained equilibrium for P factor activity but at a lower level than the P/M initiated populations. Finally mixed Q/M strains showed negligible P factor activity. The tendency in both these experiments was for P cytotype to replace M cytotype but a significant frequency of females with M cytotype remained after 17 generations. In similar types of contamination experiments Anxolabéhère *et al.* (1986) found in a set of 10 experiments, that, in populations initiated with a mixture of P, M or Q wild strains, dysgenic sterility potential was lost and the result was Q type populations after 30 generations. When populations consisting of strong P laboratory strains and Q or M' wild strains were set up, the authors found that the M cytotype disappeared during the first five generations. On the other hand the P activity potential reached an equilibrium

Kaplan *et al.* (1985) proposed a theoretical model of transposable elements which predicts their eventual extinction. This model is based on the assumption that wild type transposable elements can mutate (to functionally distinct mutant elements) as well as undergo deletions, whereas mutants can only undergo deletions. The model suggests that either an element becomes extinct in the first few generations after infection or else the copy number of the complete element increases rapidly initially and then decreases slowly as the average copy number of the mutant starts to grow. Ultimately, after a period of stabilization the complete elements disappear leaving only the mutant. At this point transposition must cease since the excision function is lost and the mutant eventually becomes extinct due to the forces of deletion and drift. Now, if Q type strains are assumed to possess mutant P elements such as KP elements (Black *et al.* 1987) or the elements found in Q strains in Japan (Sakoyama *et al.* 1985; Nitasaka *et al.* 1987), then the model of Kaplan *et al.* (1985) predicts a progression from true M to P to Q to 'M' strains. This could explain the changes that took place in Johannesburg between 1987 and 1988.

Table 2.7 Comparisons between 1987 and 1988 proportions of P, Q and M type strains†

Year	%P	%Q	%M	Collection Site
1987	0	0	100	Gordons Bay
1988	0	29	71	Cape Town
1987	50	50	0	Johannesburg
1988	12,5	75	12,5	Johannesburg
1987	25	75	0	Agatha/Pietersburg
1988	44	50	6	Nelspruit

†The strain Rut 2 (Johannesburg 1987) and all P-Q and P? strains were classified as P; All Q-P and Q-M strains were classified as Q.

CHAPTER THREE

3. *sn^W* Mutability

3.1 Introduction

The *singed* (*sn*) locus is on the X chromosome and is associated with female sterility (Bender 1960) as well as with bristle phenotype. This locus has many alleles (Engels (1979a) mentions 52 independent *sn* mutations) with varying phenotypes ranging from slightly wavy bristles or hooked tips to extremely twisted, gnarled diminished bristles with curved thoracic hairs. On the basis of their inordinately high spontaneous mutation rates (reversions to wild type occurring at frequencies of 10^{-4} and 10^{-3}) many of these alleles were judged more than twelve years ago to be insertion mutations (Golubovsky et al. 1977; Green 1977a,b).

A weak phenotypic allele of the *singed* bristle locus, *singed-weak* (*sn^W*) was identified by Engels (1979a). This allele was recovered in the progeny of a P-M hybrid dysgenic male and occurred on an X chromosome which, although originally derived from π_2 , had acquired a complex chromosomal rearrangement (Engels and Preston 1981). The *sn^W* allele was found to be hypermutable in the M cytotype, mutating to either a more extreme allele (*sn^e*) or to a nearly wild type allele (*sn⁺*) (Fig. 3.1) at a rate which could exceed 50% (Engels 1979a).

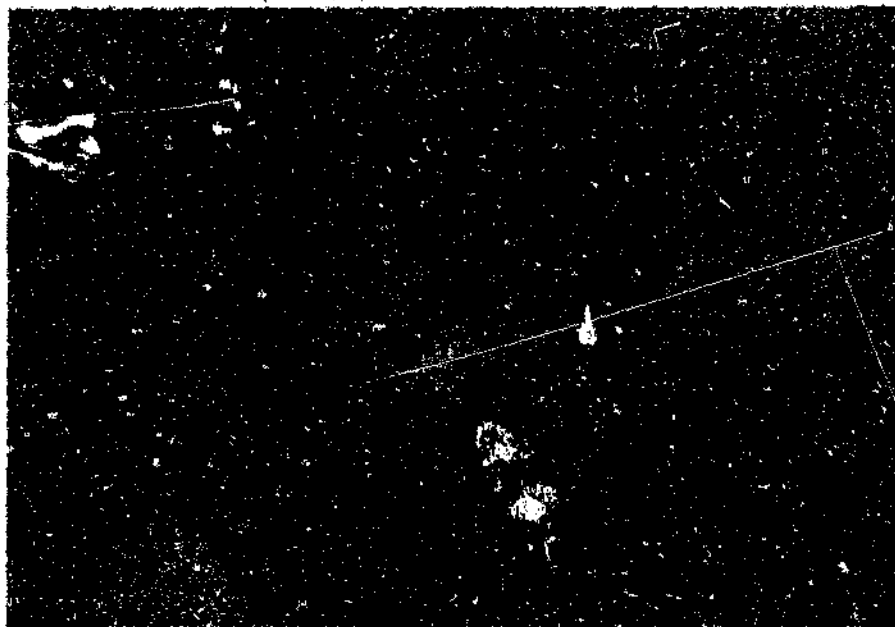


Fig. 3.1 B (ii) Lateral view of sn males.

level which was positively correlated with the P regulatory level of the Q or M' strain with which the population was initiated. These results, together with the temporal and geographical distribution of P elements in the world, led to the stepping stone concept of a progressive diffusion of autonomous P elements from America with an accompanying decrease in their activity potential (Anxolabehère *et al.* 1985).

This theory is consistent with the recent invasion hypothesis in that the primary invasion of P elements in America resulted in strong as well as moderate P strains (due to the internal deletions of P elements). The migration of moderate P individuals into Eurasia resulted in weak P/Q populations due to the continuing process of P element deletion and the accumulation of such elements. The subsequent movement of these individuals into the rest of Asia and North Africa resulted in the M' strains found there. This would also explain the distribution and strength of P type strains found in this study to be present in Southern Africa. The initial invasion into Africa would have taken place somewhere in Central Africa and then radiated outwards from there. The Sahara would have blocked the northward movement and the Kalahari desert could have impeded movement in a westerly direction. However a southern and eastward radiation would account for the weak P and Q type strains in Zimbabwe, Northern Transvaal and Natal. Finally the M' strains in the Cape would have arisen due to the introduction of defective P elements lacking P factor activity for the most part, but providing some regulatory ability. The clinal pattern exhibited in Australia could be similarly accounted for.

2.4.4.2 Temperature

A crucial element in the final outcome of P element contamination is temperature. Whereas populations maintained at 20°C attained fairly high levels of sterility (Kidwell *et al.* 1981, 1988), those raised at higher temperatures of 25°C (Kidwell *et al.* 1988) and 27°C (Kiyasu and Kidwell 1984) attained relatively lower levels of P activity. As pointed out by both Kiyasu and Kidwell (1984) and Kidwell *et al.* (1988), this temperature effect suggests that under conditions of high temperature, those flies containing P element combinations which have the ability to cause high levels of sterility when the temperature is high, will be selected against. Thus under conditions of high temperature, selection would favour those flies with moderate P activity potential. No strong P strains were obtained in this study and one of the causes may be a temperature dependent selection of flies containing less active P elements. During the periods of population increase of *D. melanogaster* that occur in the summertime, moderate to high temperatures (23°C - 30°C) are prevalent in Southern Africa.

2.4.4.3 Stochastic effects

Although the efficiency of P element contamination is high, the outcome of mixed population experiments is not always predictable. The introduction of P elements into M strain genomes under conditions of P-M hybrid dysgenesis led to the rapid evolution of these cultures into one of three relatively stable types corresponding closely to P, Q and M' strains (Kidwell *et al.* 1988).

The reason for this disparity in evolution of individual lines governed in part by temperature and early conditions appeared overall to be due to stochastic factors. This would then explain populations like Plettenberg Bay and Nelspruit where all three (P, Q and M) strain types were found.

2.4.4.4 Temporal trends

In the present study sampling took place over two years in the summers of 1987 and 1988. Although collections were not made in identical sites, some collections in 1988 were taken fairly close to those of 1987 and it is interesting to observe the differences between the two years. Thus the following places can be compared: Gordons Bay 1987 and Cape Town 1988; Johannesburg 1987 and Johannesburg 1988; Agatha/Pietersburg 1987 and Nelspruit 1988 (Table 2.7).

No real patterns of movement are evident. Whereas Q strains were absent from Gordons Bay in 1987 and appeared in Cape Town in 1988, the proportion of Q strains decreased in the Eastern/Northern Transvaal area (Agatha/Pietersburg 1987 - Nelspruit 1988) with a subsequent increase in the proportion of P and M type strains. However in Johannesburg (where the 1987 and 1988 collections were the closest geographically) an increase in the proportions in Q and M and a decrease in the proportion in P type strains occurred in 1988.



Fig. 3.1 A (i) Scanning electron microscopy photograph of thornic bristles of an insect.



Fig. 3.1 A (1) Scanning electron microscopy photograph of thoracic bristle of an sn male.



Fig. 3.1 A (iii) Scanning electron microscopy photograph of thoracic bristles of an sn mole.

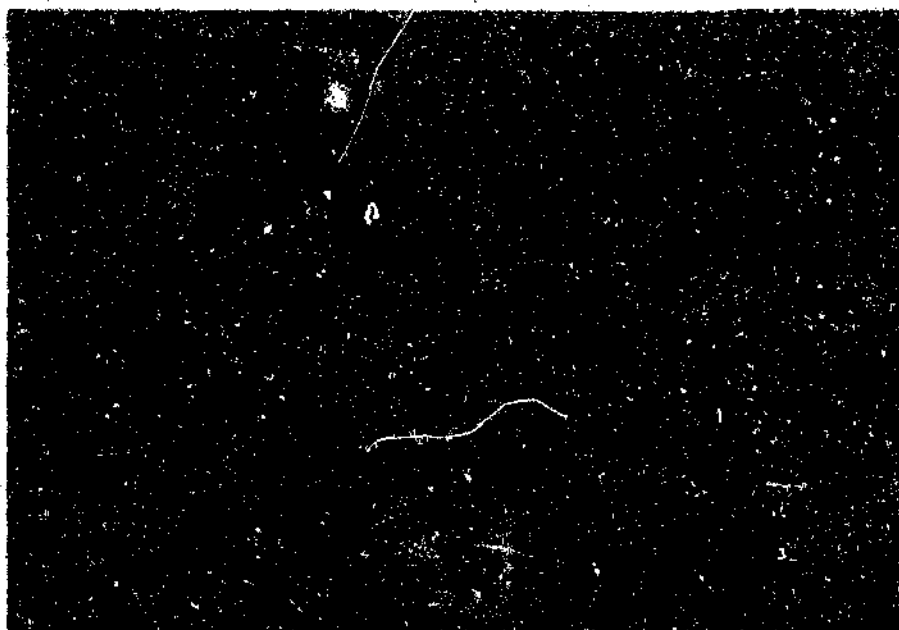


Fig. 3.1 B (i) Lateral view of sn^+ mules.

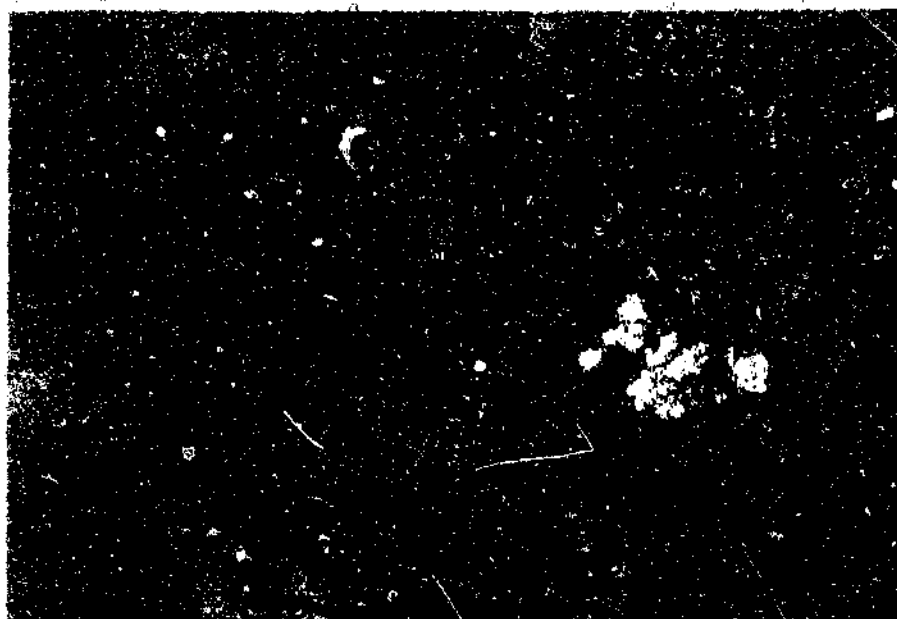


Fig. 3.1 B (111) Lateral view of sn^c males.

In the P cytotype sn^W was practically stable (Engels 1981b). This cytotype control of sn^W hypermutability suggested that the putative inserted genetic element in sn^W might be a P element (Engels 1981b). Finally Engels and Preston (1981), investigating chromosome breakage points as places of P element insertion ("hotspots") observed a breakage site coinciding with the appearance of sn^W mutation. The presence of P elements in alleles of the *singed* locus were later confirmed by molecular techniques (Nitasaka and Yamazaki 1988; Roiha *et al.* 1988).

The sn^W allele contains two defective P elements which are inserted in reverse directions in a "head-to-head" arrangement. The two alternative mutational events (sn^e and sn^+) are caused by the excision of one or the other of these P elements (Nitasaka and Yamazaki 1988; Roiha *et al.* 1988). Another P element insertion, sn^{cm} which mutates at high frequencies to sn^+ and sn^{ext} was analysed by Hawley *et al.* (1988). They found that wild type revertants appeared to represent precise (or nearly precise) excisions of the P element and that some of the sn^{ext} alleles resulted from the insertion of a duplicate copy of this P element into a nearby site in an inverted orientation.

3.1.1 sn^W Destabilization as an Assay System for P factor

Activity

Not only does P element mobilization in the sn^W allele result in an exceptionally high mutation rate, but this destabilization has also proved to be sensitive to the transposase activity

of the functional P elements. For example Spradling and Rubin (1982) transformed *Drosophila* embryos by microinjection of varying concentrations of the plasmid p π 25.1 (Section 4.2.2) which contains the 2.9 kb P element flanked by about 1.8 kb of *Drosophila* DNA, cloned into pBR 322.

The authors found that the transformation induced *en^W* instability and also that the fraction of transformants showing germ line mutations at the *stinged* locus was highly dependent on the concentration of injected plasmid DNA. They showed further that the presence of a single autonomous P element in the chromosomes of a strain was sufficient to induce *en^W* instability. The *en^W* mutability requirement for autonomous P elements was demonstrated by Karess and Rubin (1984) who constructed an autonomous P element derivative and used *en^W* hypermutability as a screen to detect transposase activity in P element constructs containing frame-shift mutations. They showed that all of the identified ORFs were essential for transposition of P elements as monitored by *en^W* hypermutability.

In addition Engels (1984), in a genetic analysis, showed that *en^W* hypermutability required a *trans*-acting product (transposase) that could only be provided by factors on P containing chromosomes. Presumably this transposase activity is provided only by the complete P factors themselves.

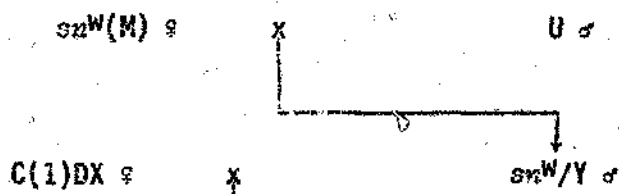
3.1.2 The Relationship between GD Sterility and sn^W Mutability

There is a great deal of variability in the induction of GD sterility due to P factor activity. Where the amount of GD sterility in P strains can vary from 10 to 100%, some strains produce no or very little GD sterility and yet they also possess the P cytotype. Thus, this variability may be due to differences in the amount and activity of complete P factors (as opposed to repressor activity) present in the genome. Since sn^W mutability is caused by P factor activity (Engels 1984), the correlation between sn^W mutability and cross A GD sterility has been examined by many workers (Engels 1984; Kocur *et al.* 1986; Hagiwara *et al.* 1987; Simmons 1987). Generally, a good correlation has been obtained between these two P element induced phenomena but with some exceptional cases arising.

This study, which examined the correlation between sn^W mutability and cross A GD sterility in 78 different lines of *Drosophila melanogaster* obtained from various parts of Southern Africa, was undertaken for two reasons: firstly, to assay the Southern African wild type strains for P factor activity using the technique of sn^W mutability which is possibly more sensitive than that of cross A GD sterility; secondly, to check the correlation between two methods of testing for P factor activity, sn^W mutability and GD sterility, and to examine any exceptional cases.

3.2 Materials and Methods

In order to determine the rate of sn^W mutability the following crosses were made:



score male progeny for sn^W , sn^e and sn^+ phenotypes

C(1)DX is a laboratory attached X line - the full description being C(1)DX, $y^+ y^+ / Y / fB$.[†]

$sn^W(M)$ is as described in section 2.2.

U is the unknown strain.

All crosses were performed at 20°C in order to prevent GD sterility.

The sn^W mutation rate was then determined from the second generation males as:

$$sn^W \text{ mutability} = \frac{(\# sn^e \sigma + \# sn^+ \sigma)}{(\# sn^e \sigma + \# sn^+ \sigma + \# sn^W \sigma)}$$

The sn^W mutation rate was then evaluated for 78 wild populations and two controls π_2 and $sn^W(M)$.

With regard to this method two points should be noted.

[†]All genetic symbols are in Lindsley and Grell (1968).

Firstly the sn^W mutation rate is measured only through the male germ line. Measurements through the female germ line may produce slightly different results (Kocur *et al.* 1986; Simmons *et al.* 1985). Secondly, since I used $sn^W(M)$ rather than Canton-S females in testing for induction of GD sterility, the female progeny examined in cross A (Chapter 2) have the same parental genomes as the males whose germ line is being tested for sn^W mutability. Thus a comparison of cross A GD sterility and sn^W mutability is a meaningful one.

3.3 Results

Appendix C shows the relative frequency of sn^W mutability in the second generation males obtained from the crosses as described in Section 3.2. Altogether 78 wild isofemale lines and two control laboratory strains ($sn^W(M)$ and π_2) were tested. The mean number of wild male progeny tested per line was 443. Unfortunately, the yield was poor on the Harare lines tested (possibly due to the action of P elements in these lines producing sterile F_1 males) and so the number of males tested in these lines varied from 26 (Har 5) to 137 (Har 10). Excluding Harare, the average number of males tested per line was 467.

Apart from a few exceptions, the number of sn^e mutants was approximately equivalent to the number of sn^+ mutants. The exceptions were Ber 10, Ber 15, Rut 4, Pta 6, H'K 2 and Nel 10. Since there are sometimes problems in distinguishing the phenotypes of sn^W and sn^+ flies, strains which showed an

excess of sn^+ phenotypes were checked two or three times. (Flies were retained and rechecked after a period of one week). It can thus be assumed that the scoring of phenotypes was accurate.

The levels of sn^W mutability obtained varied from zero per cent. for 14 of the 78 lines to a maximum value of 27.2% for a P strain Har 15 (GD sterility in cross A was 0.283 and in cross A* was 0 for this line). Thirteen of the 14 lines with zero levels of sn^W mutability were M lines. However, one strain (Mac 1) was an M-Q line. Because of this and also because this line gave a medium (MD) hybridization signal when probed with the *Acc 1* fragment of p π 25.1 (Section 4.3.2), indicating a fair number of P homologous sequences, a large number (930) of flies were tested in this line - however not one sn^+ or sn^e mutant was found among these males.

Fig. 3.2 shows the relationship between GD sterility obtained in cross A and the sn^W mutability rate. In general, strains that were able to induce GD sterility were also able to induce sn^W mutability. A highly significant correlation between P factor activity and sn^W mutability was obtained using Kendall's rank correlation coefficient with ties (Kendall 1962). ($\tau = 0.47$; $P < 0.001$, $n = 78$). Four of the strains gave values which lay outside the point cloud of the scatter diagram. They were Nel 20, P'B 1, Har 5 and Har 15. Nel 20 and P'B 1 had exceptionally high levels of cross A GD sterility (0.528

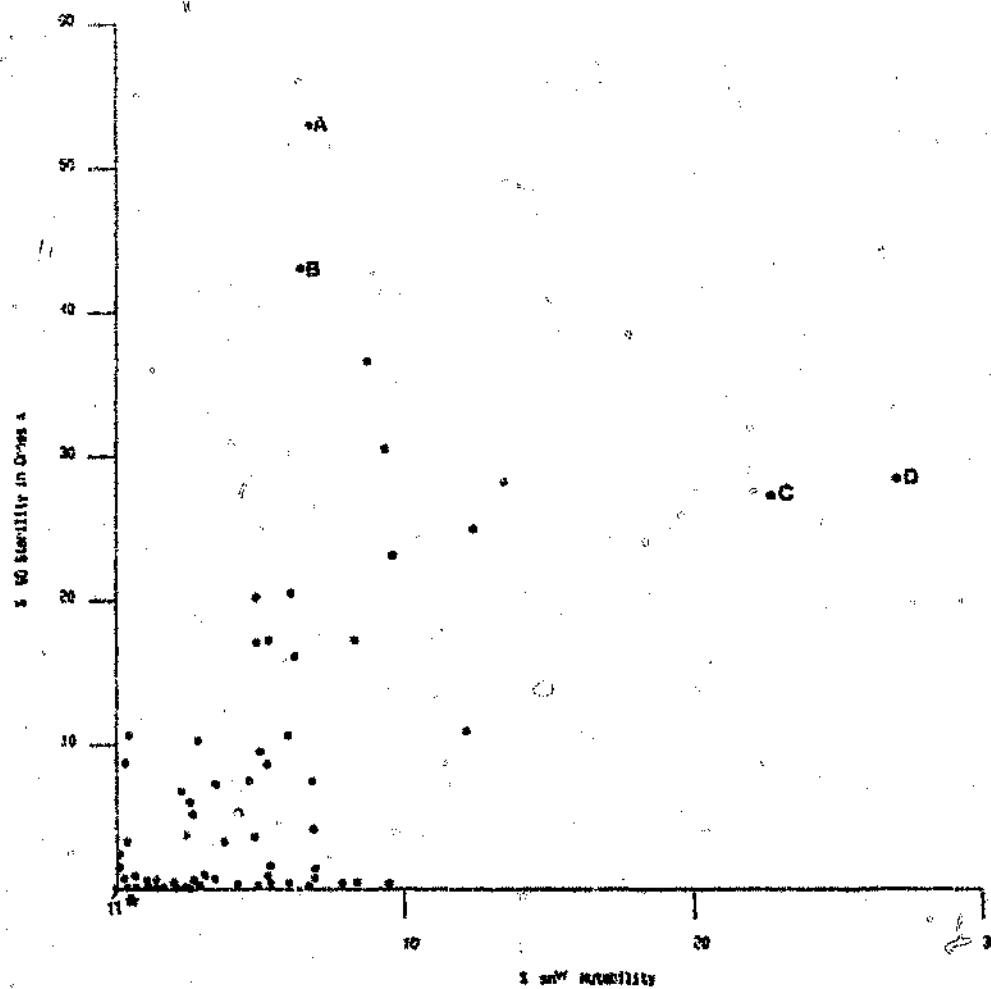


Fig. 3.2

Relationship between Cross A GD sterility and sm^M autability

The four strains lying outside the cloud of points are:

A Mel 20

B P'B 1

C Mar 5

D Mar 15

*** 11 strains have coordinates (0,0)**

and 0.432 respectively) when compared to their *sn^W* mutability levels (0.068 and 0.065 respectively). For Har 5 and Har 15, the opposite was true. Here the levels of *sn^W* mutability were high (0.221 and 0.272 respectively) whereas the levels of cross A GD sterility were relatively low (0.276 and 0.283 respectively).

Several Q strains Gob 5, AB 15, JHB 15 and one P-Q strain Ber 20 had levels of *sn^W* mutability less than 10^{-2} . However no Q strains showed zero mutability. On the other hand a number of M strains had *sn^W* mutability levels greater than zero (Table 3.1). Levels of 1% mutation could be due to the spontaneous mutation rate for an insertion mutation. Engels (1979a) reported mutation rates between 10^{-3} and 10^{-2} at alleles (other than *sn^W*) of the *singed* locus when backcrossing dysgenic hybrid males to M type females. Kolubovsky *et al.* (1977) found reversions of *singed* mutations to wild type to occur at frequencies of 10^{-4} - 10^{-5} . Alternatively this mutation rate may be indicative of autonomous P elements in M type strains or of defective P elements, retaining some type of transposase function (see Section 3.4.1).

3.4 Discussion

The close correlation obtained between the ability to activate *sn^W* mutability and the potential to cause GD sterility seen in this study has been obtained by many other researchers. Engels (1984) obtained a value of $r = 0.779$, $P = 0.00023$ when testing a set of 12 P strains which were selected so that they represented

Table 3.1

Southern African strains which showed levels of sn^W mutability greater than zero and which tested as M or M-Q in GD sterility tests

Strain	sn^W mutability	GD sterility		Type
		Cross A	Cross A*	
Gob 10	0,012	0	0,308	M
Ais 5	0,004	0	0,198	M
Plet 6	0,006	0,033	0,485	M
SS 25	0,002	0	0,759	M
Wal 3	0,003	0	1,000	M
Wal 4	0,011	0	0,300	M
Mac 3	0,008	0	0,180	M
Br	0,070	0,038	0,140	M-Q
Pta 6	0,070	0,016	0,158 ^B	M-Q

the entire range of GD sterility potential. Kocur *et al.* (1986) tested 45 inbred lines derived from flies trapped in North America between 1978 and 1983 and also obtained a significant correlation (using Kendalls τ) between cross A GD sterility and sn^W mutability. Simmons (1987) used 12 lines which had been constructed approximately one and a half years prior to testing by extracting single X chromosomes from a wild population into a laboratory background free of P elements. The author obtained a strong positive correlation between GD sterility obtained in cross A and sn^W mutability. Finally, Hagiwara *et al.* (1987) using 35 wild lines from Japan and elsewhere also found that there was a strong relationship between cross A GD sterility and the sn^W mutation rate. Despite the high correlation between sn^W mutability and P factor induced GD sterility there are still several anomalous results. These will be discussed below.

3.4.1 M Strains and Q strains

Several M strains displayed levels of sn^W mutability exceeding 10^{-2} (Table 3.1). If autonomous P elements are responsible for sn^W mutability at these levels (Karess and Rubin 1984; Engel 1984), then it would appear that complete P elements can exist in the M cytotype. In particular, the two M-Q strains Br and Pta 6, in addition to producing 7% sn^W mutability, also gave medium (MD) hybridization signals when probed with the *Aad 1* fragment of p π 25.1 (Section 4.3.2) thus indicating the presence of a number of P homologous sequences in these

strains. Further evidence of the possible existence of P factors in M cytotype strains was given by Hagiwara *et al.* (1987). They identified a strain which had M cytotype but which was able to induce a high (17%) *sn^W* mutation rate. In subsequent tests the cytotype of this strain remained M but the rate of *sn^W* mutation decreased to about 0.9%. The authors suggested that some intact P elements may have existed in this strain, but not sufficient to convert the cytotype to P, i.e. these elements were not able to repress GD sterility in a P background.

On the other hand, the rate of *sn^W* mutability in several of the Q strains was extremely low. Simmons and Bucholz (1985) suggested that the binding of transposase to the terminals of defective P elements would create a shortage of transposase in the vicinity of the elements responsible for the *sn^W* mutations thus lowering the mobilization frequencies of these elements. In this regard Simmons *et al.* (1987) found that dysgenic flies that had one or more paternally derived chromosomes containing defective P elements from an inbred line of the Muller-5 Birmingham M' strain, were partially able to repress the instability of the *sn^W* mutation. Thus the low rate of *sn^W* mutability in these Q strains could be caused by the titration of the limited amount of transposase present in these strains due to the presence of a number of internally deleted P elements.

3.4.2. Exceptional Cases

Although Engels (1984) found one exceptional case (a strain which produced more sn^W instability than would have been expected from the amount of GD sterility it induced), he nonetheless concluded that both GD sterility and sn^W hypermutability are caused by the P transposase and further "that P strains vary widely in the amount or kind of transposase they produce".

Kocur *et al.* (1986) on the other hand, found that one strain in their sample consistently induced too much GD sterility for its ability to destabilize sn^W . This, coupled with the results of Engels (1984) suggested once more that the relationship between transposase production and sn^W mutability may not be a simple one. This is borne out by results obtained in this study where both types of exceptional cases occurred. Nel 20 and P'B 1 produced extremely high levels of GD sterility compared to their relatively low levels of sn^W mutability while the opposite was true for Har 5 and Har 10 (Appendix C).

Low levels of sn^W mutability might be due to the titration of transposase in the presence of many defective P elements (Simmons and Bucholz 1985). One more piece of evidence for this possibility was the strain Mac 1 in which all of the 930 flies tested were of type sn^W . Yet Mac 1 was an M-Q strain (Appendix A) and produced a median (MD) hybridization signal when probed with the *Aco 1* fragment of p25.1

(Section 4.3.2). This strain then may have a large number of defective P elements present which are titrating the limited amount of transposase present in the strain and preventing sn^w destabilization.

The case of high levels of sn^w mutability combined with low levels of GD sterility induction is difficult to explain. However an investigation of the relationship between the MR and P-M mutator systems showed that although P elements are present in MR stocks and that MR and P-M are part of the same mutator system, nonetheless there were certain potent mutator chromosomes which generated little or no GD sterility (Green 1984).

Thus it seems likely that as well as P elements there may be other functions which enhance or depress sn^w mutability. These other functions may be dependent on the chromosomal complement of the strain as well as on the sites of P element insertion and the presence of other transposable elements in the genome.

CHAPTER FOUR

4. Molecular Analysis of P elements

4.1 Introduction

Phenotypic tests, such as those described in chapters 2 and 3, present an accurate picture of the distinction between strong P and strong M strains and can also serve to identify those Q strains which are typical of their class. However phenotypic tests are unable to discriminate between those M strains which are completely devoid of P elements and those which are not (M' strains). Also many strains are vaguely defined as lying between P and Q or between Q and M (Appendix A) and a knowledge of the quantity and types of P elements within these strains would help to enhance our knowledge of the mechanisms and control of P element transposition. The advent of molecular techniques and the introduction of P element probes has consequently opened up many new paths of investigation in the P-M system.

4.1.1 Cloning of P elements

Molecular cloning of P elements was achieved by Rubin et al. (1982). They made use of five mutations at the *white* locus, two of which had previously been generated by Simmons and Lim (1980) in P strain chromosomes and three of which they generated in M strain chromosomes. These *white* mutants were thought to be due to P element activity since they arose in crosses between P males and M females and were unstable in M cytotype but stable in P cytotype. The *white* locus was chosen

since, as well as being easily scorable, the entire wild type *white* locus had previously been cloned (Bingham *et al.* 1981; Levis *et al.* 1982).

The structure of the mutant *white* alleles was investigated in a series of molecular cloning and Southern blot hybridization experiments. The mutant *white* alleles were all shown to contain additional DNA sequences inserted into the locus. These sequences although heterogeneous in size, ranging from 0.5 to 1.4 kb, were all homologous and were designated P elements. Excision of the P elements occurred when these mutations were placed in the M cytotype and resulted in a wild type phenotype. The same mutations were stable when maintained in the P cytotype.

O'Hare and Rubin (1983) isolated and characterized several members of the P family of transposable element and they showed that these ranged in size from less than 500 bp up to a 2.9 kb autonomous P element. All P elements were shown to have a 31 bp inverted terminal repeat and to create a direct 8 bp duplication of genomic target sequence upon insertion. The smaller P elements were shown to be missing sequences internal to the terminal inverted repeats.

O'Hare and Rubin (1983) constructed the plasmid p π 25.1 which consists of the 2.9 kb P element flanked by genomic DNA from the 17C⁺ region of chromosome 1 of π_2 (described in Section 2.2.3) cloned into the *Eam* HI site of pBP 322. This plasmid, as well

as its restriction fragments, has been used extensively to probe for P elements in wild collections and in laboratory strains of *Drosophila* (Examples include Todo *et al.* 1984; Anxolabéhère *et al.* 1985b, 1988b; Sakoyama *et al.* 1985; Daniels *et al.* 1987a; Boussy *et al.* 1988; Good *et al.* 1989).

The technique of molecular probing, as well as being used to reveal the presence or absence of P element homologous material within *Drosophila* strains, has also been used to quantify the amount of this material present in the strain by examining the intensity of the hybridization signal. (Anxolabéhère *et al.* 1985b, 1988b; Boussy *et al.* 1988; Good *et al.* 1989). Further by means of restriction enzyme and Southern blot techniques the particular species of P homologous DNA present in the strains can be analysed. These (above-mentioned) techniques have been used to shed some light on the causative effects of P element activity and susceptibility.

In this study I probed for P homologous elements in 31 wild strains of *Drosophila melanogaster* which were also tested phenotypically for GD sterility (Chapter 2).

4.1.2 Objectives

The hybridization was performed with two objectives in mind: to determine whether South African strains which test as M phenotypically are entirely devoid of P elements and to try to ascertain the quantity of P homologous DNA present in these strains by visually comparing the intensity of the signals obtained between these strains and certain controls.

Several sections of the laboratory work were performed with the help and guidance of Susan Hearn under the supervision of Jocelyn Webster.

4.2 Materials and Methods

4.2.1 Extraction of *Drosophila* DNA (Jowett 1986)

Genomic DNA was extracted from wild collections and two laboratory strains using the "single fly DNA extraction" method of Jowett (1986). Eight adult flies (four males and four females) from each line were anaesthetized, placed in Eppendorf tubes, frozen in liquid nitrogen and stored at -70°C prior to use. Each sample was homogenized in 100 μl of 10 mM Tris-HCl (pH 7.5), 60 mM NaCl, 50 mM ethylenediaminetetra-acetate (EDTA), 0.15 mM spermine and 0.15 mM spermidine. Once the tissue was broken up, nuclei were lysed and DNA released by incubating at 60°C for 30-40 minutes with 100 μl of 1.25% sodiumdodecylsulphate (SDS), 0.3 M Tris-HCl (pH 9.0), 0.1 M EDTA, 5% sucrose and 0.75% freshly added diethylpyrocarbonate (DEP). The SDS and protein form complexes and DEP acts as a protein denaturant.

and nuclease inhibitor. The protein/SDS complexes were then precipitated by adding 30 μ l of 8 M potassium acetate and cooling on ice for 45 minutes. The tubes were spun for one minute in the microcentrifuge (Hager Design) and the supernatant containing DNA was carefully removed. The DNA was precipitated by adding two volumes of cold ethanol (-10°C). Samples were allowed to stand at room temperature for 15 minutes and then microcentrifuged for 5 minutes at 4°C . The supernatant was discarded and the pellets were washed with cold 70% ethanol and vacuum dried. The DNA was resuspended in 100 μ l of 10 mM Tris, 1 mM EDTA, pH 8.0 (TE) for at least 24 hours before restriction enzyme digestion. DNA samples of 20 μ l each were run on 0.7% agarose gels to ascertain that the DNA concentration obtained was uniform throughout (Fig. 4.1).

4.2.2 Preparation of the Probe

The plasmid $\text{px}25.1$ (O'Hare and Rubin 1983) was obtained from W.R. Engels who kindly supplied it in two forms: the purified plasmid DNA (from a recent plasmid preparation) dissolved in 1 M NaCl, 1.0 mM EDTA, 1.0 mM Tris as well as in a stab in *E. coli* DH1 (*recA1*, *gyrA96*, *thi-1*, *endA1*, *hsdR17*, *supE44*). $\text{px}25.1$ contains a complete 2907 bp P element which is flanked by approximately 1.8 kb of DNA from the 17C region of the X chromosome of the *D. melanogaster* strain of π_2 and inserted into the *Bam*HI site of the plasmid $\text{pBR}322$ (O'Hare and Rubin 1983). The plasmid consequently has the gene for ampicillin resistance. After

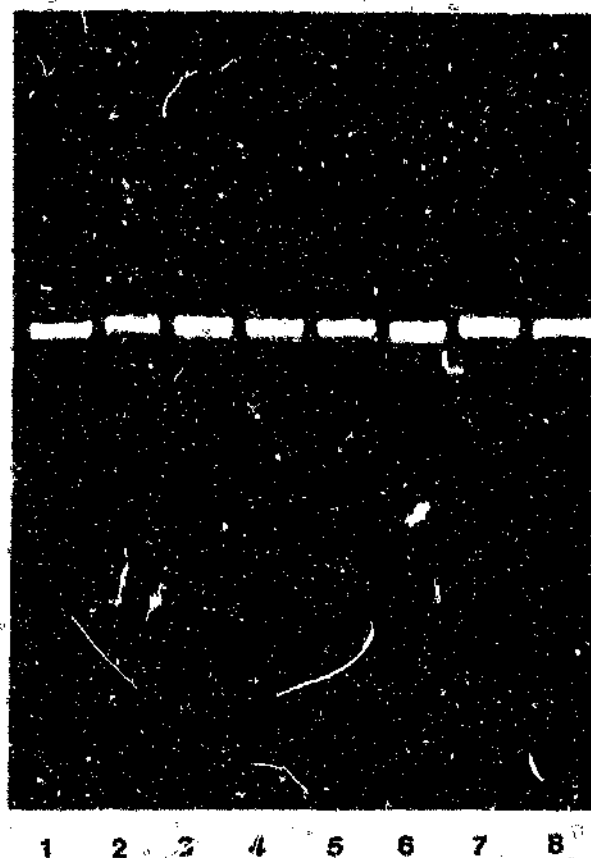


Fig. 4.1

Examples of Drosophila DNA extractions.

Genomic DNA was extracted using the "single fly DNA extraction" method of Jowett (19...) as described in Section 4.2.1. The two laboratory strains (lanes 2 and 3) as well as five of the thirty one wild strains (lanes 4-8) are shown in the photograph above. DNA samples of 20 μ l each were run on a 0.7% agarose gel.

- Lane 1: 200 ng λ DNA
- Lane 2: π_2
- Lane 3: Canton-S
- Lane 4: Rut 1
- Lane 5: Oen 1
- Lane 6: Pta 1
- Lane 7: H'K 1
- Lane 8: Ag 1

plating from the stab cultures the plasmid was amplified as follows: the *E. coli* strain DH1 containing p π 25.1 was inoculated into 10 ml of Luria Broth (LB) containing 1 μ l of ampicillin (50 mg/ml) and incubated overnight with vigorous shaking. A subsequent inoculation of 0,1 ml of the overnight culture into 25 ml of LB containing 25 μ l of (50 mg/ml) ampicillin in a 100 ml flask was performed. The flask was incubated at 37°C with vigorous shaking until the culture reached the late log phase (Optical density (OD) at 600 nm = 0,6). A final inoculation of 25 ml of the culture into a 500 ml of 37°C LB medium containing 0,5 ml of ampicillin was performed. The culture was incubated at 37°C with vigorous shaking until OD at 600 nm was = 0,4. Finally 2,5 ml of chloramphenicol (34 mg/ml in ethanol) was added to the culture flask which was then incubated at 37°C with vigorous shaking for 16 hours.

4.2.2.1 Plasmid Isolation (Clewell and Helinski 1969)

Plasmid material was obtained from Susan Hearn after being isolated by the following method. The amplified plasmid p π 25.1 was isolated from DH1 as follows: A 40 ml culture of *E. coli* DH1 containing plasmid p π 25.1 was harvested at 6 000 rpm for 10 minutes. The pellet was washed in 10 ml of cold TE, then pelleted and resuspended in 2,0 ml of cold 25% sucrose, 0,05 M Tris (pH 8,0). The sample was transferred to a 100 ml flask and 0,25 ml of fresh lysozyme solution (10 mg/ml in 0,25 M Tris HCl (pH 8,0)) was added and the sample was swirled gently on ice for 5-15 minutes. Lysozyme is used to catalyze the hydrolytic

cleavage of complex polysaccharides in the protective cell walls of the bacteria. Occasional swirling on ice was continued for five minutes after the addition of 0,25 ml of 0,5 M EDTA pH 8,0.

To promote cell lysis, 2,5 ml of cold detergent (20% Triton X-100, 0,5 M EDTA (pH 8,0), 1 M Tris (pH 8,0) and distilled water to required volume) was added and the sample was gently swirled on ice for 10-20 minutes until the solution was somewhat cleared and highly viscous. The solution was transferred to a Beckman JA-20 centrifuge tube and centrifuged at 18 000 rpm for 45 minutes. The supernatant was then removed and the volume measured and 0,1 ml EtBr (10 mg/ml) and 0,95 g CsCl per ml of supernatant was then added. A 10 000 rpm spin for 15 minutes at 4°C produced a cleared lysate which was loaded into Beckman quick-seal ultracentrifuge tubes. The sealed tubes were spun in the Beckman ultracentrifuge at 45 000 rpm for 16-17 hours at 10°C. The (lower) plasmid bands were observed under ultraviolet light and the DNA was removed using a sterile syringe.

Finally, the DNA was purified. To remove the EtBr, an equal volume of isopropanol saturated with 5 M NaCl was added to the DNA solution, the two solutions gently mixed and spun for 30 secs in the microcentrifuge at 4°C. The solutions separated into two distinct layers and the upper (pink) layer containing EtBr was discarded. This procedure was repeated until a colourless solution was obtained. The CsCl was removed by dialyzing against cold (4°C) TE buffer for about 12 hours.

The plasmid DNA obtained was subjected to electrophoresis on a 0.7% agarose gel together with the p π 25.1 plasmid isolated in the laboratories of W.R Engels in order to confirm that the correct size plasmid had been extracted. At the same time the DNA concentration was assessed.

Two methods of ethidium bromide fluorescent quantitation of the amount of double stranded DNA (Maniatis *et al.* 1982) were used to estimate the concentration of plasmid isolated from DH1. They were the plastic wrap method (results not shown) and the minigel method. Both methods make use of the difference in the intensity of fluorescence of different concentrations of DNA/EtBr under the ultraviolet transilluminator. The plastic wrap method spots DNA samples onto a sheet of plastic stretched over a UV transilluminator whereas in the minigel method the samples are subjected to electrophoresis before viewing on the transilluminator. The estimation of the concentration of plasmid DNA isolated from DH1 using the two methods was the same viz. 10 ng/ μ l (Fig. 4.2).

4.2.2.2 Digestion of p π 25.1

To 25 μ l of purified p π 25.1 (250 ng) was added $\frac{1}{2}$ unit[†] Aca I and 3 μ l of the recommended commercial incubation buffer.

[†]The catalytic activity of the restriction endonuclease is based on the minimum amount of enzyme required for the generation of the "enzyme-specific" final fragment pattern of λ DNA. Thus, one unit is the enzyme activity that completely splits 1 μ g of λ DNA in one hour at 37°C. The pH of the incubation buffer is adjusted to the test temperature (in this case 37°C).

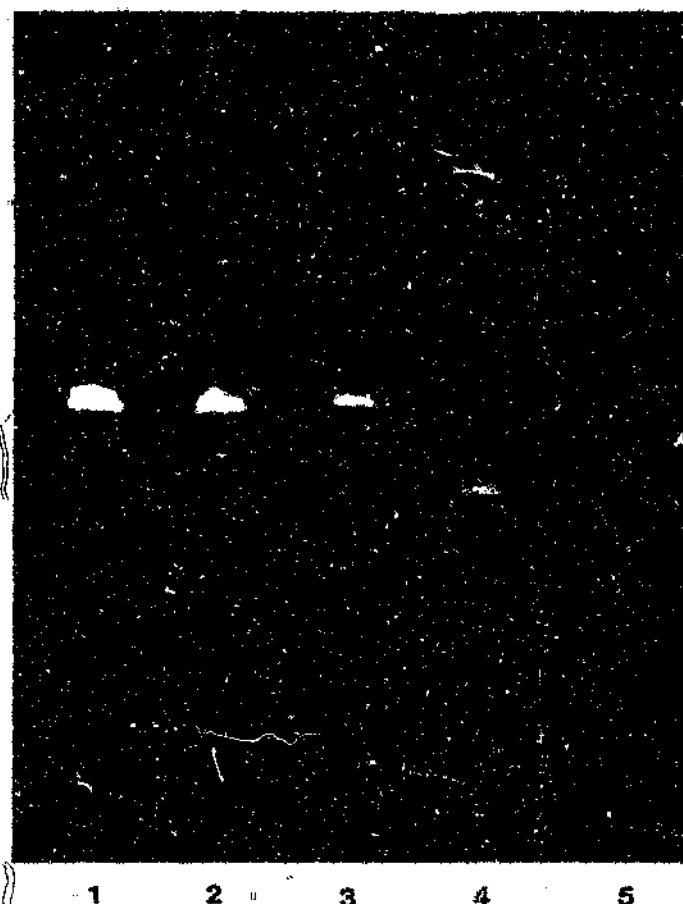


Fig. 4.2 Estimation of the concentration of plasmid DNA using ethidium bromide fluorescent quantitation methods (Maniatis *et al.* 1982).

Lanes 1, 2 and 3 contain 200 ng, 100 ng and 50 ng respectively of λ DNA.

Lane 4 contains 10 μ l of plasmid pr25.1 DNA isolated from the *E. coli* strain DH1 as described in Section 4.2.2.1.

Lane 5 contains 1 μ l of the purified plasmid pr25.1 as obtained from the laboratories of W.R. Engels.

Comparison of the intensity of the fluorescence of the λ DNA standards with that of the DNA in lane 4 gave a concentration estimate of 10 ng/ μ l for the isolated plasmid DNA. (Intensity of fluorescence of the sample in lane 4 was judged to be equal to that of the λ DNA standard in lane 2; i.e. the sample contained 100 ng of DNA).

Digestion was allowed to proceed for at least four hours at 37°C. The reaction was halted by addition of 1 µl of 0,5 M EDTA pH 7,5. The restricted DNA samples were subjected to electrophoresis together with a λ III marker on a 0,65% low melting point agarose gel at 4°C.

4.2.2.3 Radioactive labelling of DNA

The Amersham RPN1601 multiprime DNA labelling system was used. The band of agarose containing the Acc 1 2,4 kb P element fragment (see Section 4.3.1 for a full description) was melted and the DNA denatured by heating the agarose/DNA at 95°C for 10 minutes in a water bath. Thirty (30) µl of the molten agarose/DNA were mixed with the following components of the labelling kit: 15 µl buffer, 7,5 µl primer and 5 µl enzyme. Finally 4 µl of deoxycytidine 5'-[α -³²P]triphosphate was added and the reaction was allowed to proceed overnight at room temperature.

The level of radioactive incorporation was assayed by blotting 1 µl of labelled DNA onto the nitrocellulose filter, fixing with 10% trichloroacetic acid (TCA) and washing three times with 5% TCA. The approximate level of incorporation was read with a Geiger counter. A more accurate measure was obtained by reading the sample (1 µl of labelled probe added to 5 ml of scintillation mix) in the scintillation counter (Packard Instruments - Protea).

The level of incorporation of ^{32}P for the 2,4 kb P element DNA probe was found to be $1,5 \times 10^6$ cpm/ μg .

4.2.3 Agarose Gel Electrophoresis

DNA samples were run on 0,7% agarose minigels. These gels were prepared by dissolving the agarose powder in Tris-Borate-EDTA (TBE) (0,089 M-Tris base, 0,089 M boric acid, 0,002 M EDTA). After pouring the gel, it was allowed to set for at least two hours before electrophoresis. One x TBE was used as the running buffer. Samples to be run were mixed with a gel loading buffer containing 0,25% Bromophenol as the tracking dye and 40% (w/v) sucrose in distilled water to sink the samples in the well. Gels were run at 8 v/cm and then stained with (0,5 $\mu\text{g}/\text{ml}$) EtBr. An ultraviolet transilluminator was used to visualize the DNA bands.

A 0,65% low melting point agarose gel was used for electrophoresis of the restricted plasmid DNA. Electrophoresis was performed at 4°C and the gel was subsequently viewed under the transilluminator where the appropriate bands were excised. The cut out pieces of agarose containing the DNA bands were stored at 4°C.

4.2.4 Hybridization

4.2.4.1 Binding of DNA to nitrocellulose (Maniatis *et al.* 1982)

Denaturation of DNA samples was accomplished by heating at 95°C for 10 minutes. The samples were then quickly chilled on ice, an equal volume of 1 M NaOH added, and incubated at room temperature for 20 minutes. The DNA samples were then neutralized by the addition of 0,5 volumes of 1 M NaCl, 0,3 M Na citrate, 0,5 M Tris HCl (pH 8,0) and 1 M HCl. The solutions were well mixed and samples were immediately chilled on ice.

Denatured DNA was bound to nitrocellulose by spotting 5 μ l of the DNA solution onto the filter, allowing it to absorb and then spotting a further 5 μ l. This process was repeated until ~ 40 μ l of the original sample had been spotted. The filter was then baked for 1 hr 20 min at 80°C and stored at room temperature prior to use.

4.2.4.2 Prehybridization

Prior to prehybridization, the nitrocellulose blots were washed in 3 x SSC (1 x SSC = 0,15 M NaCl, 0,015 M Na citrate) for 10 minutes at room temperature. They were then immersed in a prehybridization solution consisting of 50% deionized formamide, 4,37 g NaCl, 2,2 g Na citrate, 25 ml of 20 x Denhardt's (0,4% Ficoll, 0,4% Bovine Serum Albumin (BSA), 50 mM NaPO₄ buffer (1 M Na₂HPO₄, 1 M H₂PO₄ (pH 6,5)) and water up to 100 ml) salmon sperm DNA (1,5 ml of 1 mg/ml) was denatured by boiling for 10 minutes. This DNA was then chilled on ice and added to the prehybridization solution. The dots were prehybridized with shaking for 4-5 hours at 42°C.

4.2.4.3 Hybridization

The prehybridization solution was decanted and 50 ml of hybridization solution (50% deionized formamide, 4.37 g NaCl, 2.2 g Na citrate, 25 ml of 20 Denhardt's, 50 mM NaPO₄ buffer (pH 6.5), 20 ml Dextran sulphate (50% solution) and water up to 100 ml) was added. The labelled probe together with the salmon sperm DNA (1. of 1 mg/ml) was boiled for 10 minutes and then chilled on ice and added to the warm hybridization medium. The blots were hybridized with shaking for 18-24 hours at 42°C.

4.2.4.4 Posthybridization

Following hybridization, the nitrocellulose filters were washed with 50 ml of 5 x SSC at 42°C for 10-15 minutes. The filters were then dried lightly on clean paper towels and sealed in plastic bags in preparation for autoradiography.

4.2.4.5 Autoradiography

Nitrocellulose blots sealed in plastic were placed in cassettes and exposed to Fuji X-ray films for 1 to 3 days at -70°C. X-rays were developed, fixed and washed and hybridization patterns were observed.

4.3 Results

4.3.1 Digestion of Plasmid px25.1

Fig. 4.3 depicts the *Acc* I and *Bam* HI restriction sites of the plasmid px25.1. The 2.4 kb *Acc* I fragment of the P element makes up 83% of the full sized 2.9 kb P element. This fragment

4.2.4.3 Hybridization

The prehybridization solution was decanted and 50 ml of hybridization solution (50% deionized formamide, 4.37 g NaCl, 2.2 g Na citrate, 25 ml of 20 x Denhardt's, 50 mM Na₂ buffer (pH 6.5), 20 ml Dextran sulphate (50% solution) and water up to 100 ml) was added. The labelled probe together with the salmon sperm DNA (1.5 ml of 1 mg/ml) was boiled for 10 minutes and then chilled on ice and added to the warm hybridization medium. The blots were hybridized with shaking for 18-24 hours at 42°C.

4.2.4.4 Posthybridization

Following hybridization, the nitrocellulose filters were washed with 50 ml of 5 x SSC at 42°C for 10-15 minutes. The filters were then dried lightly on clean paper towels and sealed in plastic bags in preparation for autoradiography.

4.2.4.5 Autoradiography

Nitrocellulose blots sealed in plastic were placed in cassettes and exposed to Fuji X-ray films for 1 to 3 days at -70°C. X-rays were developed, fixed and washed and hybridization patterns were observed.

4.3 Results

4.3.1 Digestion of Plasmid p π 25.1

Fig. 4.3 depicts the *Acc* I and *Bam* HI restriction sites of the plasmid p π 25.1. The 2.4 kb *Acc* I fragment of the P element makes up 83% of the full size 2.9 kb P element. This fragment

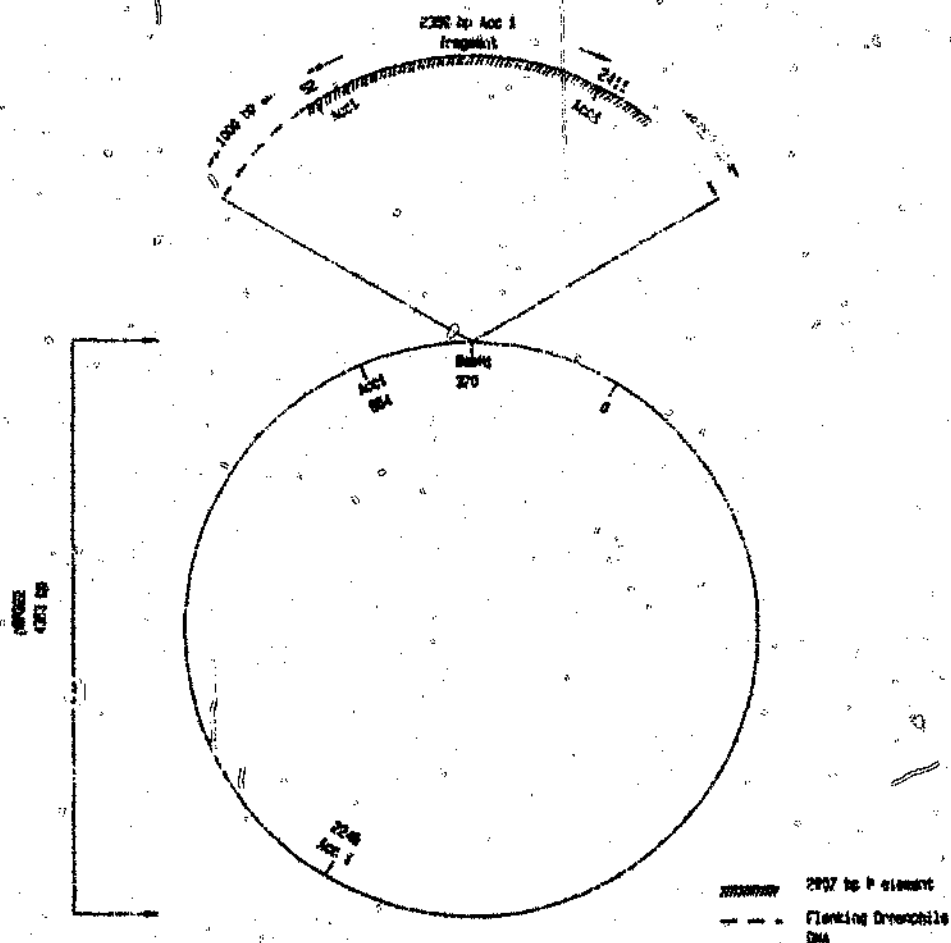


Fig. 4.3 Partial restriction map of the plasmid pR25.1 (O'Hare and Rubin 1983)

The restriction sites shown on the 2.9 kb P element were based on the published sequences of O'Hare and Rubin (1983). The restriction sites shown on pR25.1 were obtained from the Boehringer Mannheim (1987) manual. The 2.9 kb P element is flanked by about 1.8 kb of *Drosophila* DNA from the 17C region of chromosome 1 of π_2 .

can be used to detect intact P elements as well as those degenerate P elements having either a deletion between the two *Acc* I sites or having one of the *Acc* I sites deleted. Theoretically the only elements that will not be detected using the *Acc* I fragment as a probe are those degenerate P elements containing deletions that span the entire *Acc* I interval. The restriction of P elements by *Acc* I, in order to obtain a 2.4 kb probe and the digestion of genomic DNA by *Acc* I, have been used successfully in hybridization studies involving P elements (Daniels *et al.* 1987a; Boussy *et al.* 1988).

Fig. 4.4 shows the *Acc* I digest of the plasmid p π 25.1. From the *Acc* I restriction map (Fig. 4.3) it was deduced that four fragments of size ca 3789 bp, 2359 bp, 1595 bp and ca 1327 bp should theoretically be obtained. In order to assess the actual size of the fragments, the digested plasmid was subjected to electrophoresis in a lane adjacent to the λ DNA molecular weight marker III, and use was made of the relationship between migration distance and molecular weight as follows: the distance, d , moved in a fixed time by a DNA fragment of molecular weight M is given by $d = a - b(\log M)$ where a and b are empirically determined constants depending on the buffer, gel concentration and temperature. Thus the standard curve (log molecular weight (in bp) versus distance migrated) was fitted using the molecular weight marker III and was used to obtain the size of the p π 25.1

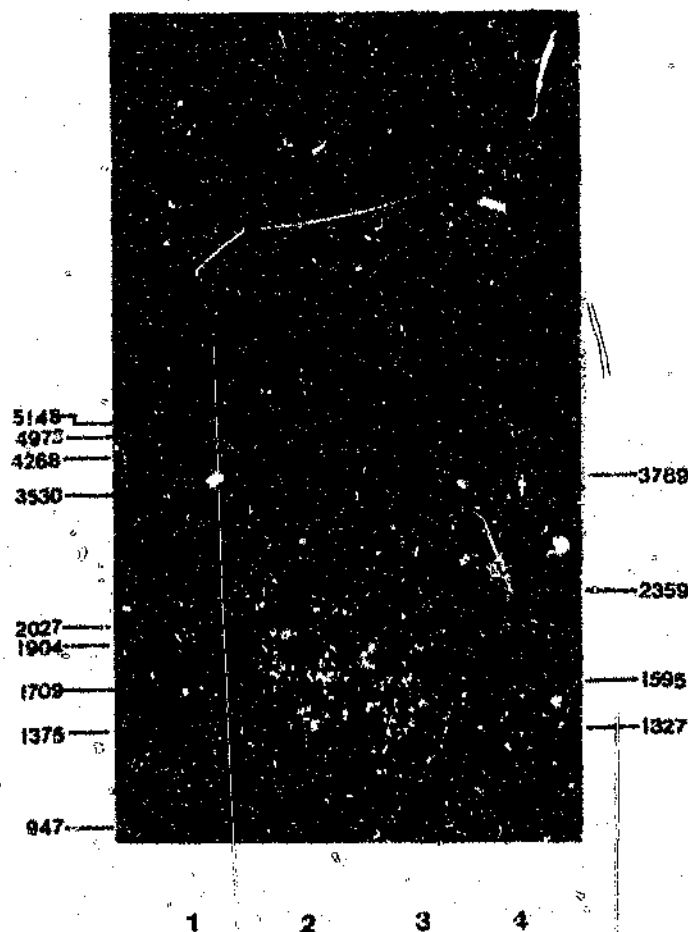


Fig. 4.4

The plasmid p25.1 digested with Acc 1 on a 0.65% low melting point agarose gel.

Lane 1: λ DNA molecular weight marker III (12 μ l)

Lanes 2 and 3: p25.1 DNA restricted with Acc 1 as described in Section 4.2.2.2. (25 μ l)

Lane 4: p25.1 DNA undigested (10 μ l)

The fragment lengths of λ DNA molecular weight marker III as obtained from Boehringer Mannheim (1987) are indicated on the left hand side of the photograph. On the right hand side of the photograph are the expected fragment lengths obtained from the Acc 1 restriction of the plasmid p25.1 (Fig. 4.3). The faint bands present are due to incompletely digested DNA or concatamers that formed subsequent to digestion.

fragment (Fig. 4.5). A good agreement was obtained between the theoretical expected fragment size and those estimated from the standard curve Table 4.1).

4.3.2 Dot Blot Hybridization

In 31 different wild populations collected mainly in 1987, samples of eight flies (four females and four males) were tested for the presence of P elements by direct hybridization with the *Acc* 1 2.4 kb fragment of the P element. The probe used is free of the *Drosophila* chromosomal sequence surrounding the P element in π 25.1. Hence no hybridization signal would be due to chromosomal non P element DNA. Three controls were routinely tested in each blot (except for one in which the one control dot broke off). The controls were the plasmid π 25.1, the strong P strain π ₂ (described in section 2.2.3) and the M strain, Canton-S. Canton-S is a laboratory stock, probably imported from the United States of America to South Africa in the 1950s, and subsequently kept in the laboratory of the Genetics Department of the University of the Witwatersrand. The intensity of hybridization signal obtained for the wild flies was visually compared to the signals obtained with reference M (Canton-S) and P(π ₂) strain flies on the autoradiograph. Photographs of the dots are shown in Figs. 4.6 A-F.

In all blots where Canton-S was tested, a faint hybridization signal was obtained. This could be due to either non-specific hybridization or to the presence of P elements in Canton-S. The latter seems the more plausible of the two explanations since certain Canton-S strains have been discovered to harbor defective P elements (Woodruff *et al.* 1987a,b).

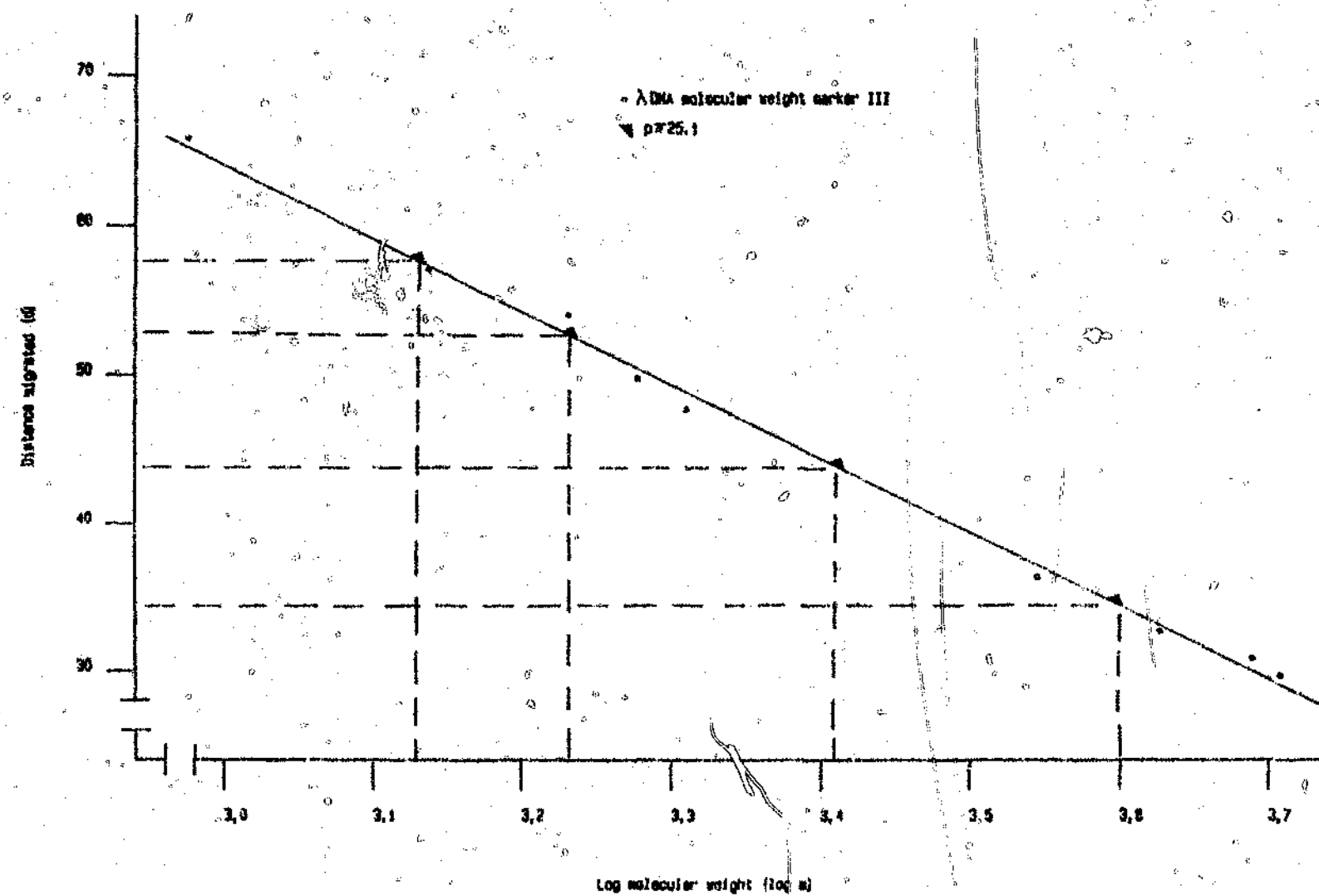


Fig. 4.5 Standard curve of log molecular weight (log M) versus distance migrated (d) for the lambda DNA molecular weight marker III. The size of the pX25.1 fragments were obtained from the graph by reading off the abscissa corresponding to the distance migrated for each fragment.

Table 4.1 The expected and estimated fragment sizes obtained from the *Acc* I digestion of p25.1

The regression equation of the straight line obtained by plotting distance migrated (d) against molecular weight (m) (in bp) for λ DNA molecular weight marker III is
 $d = 210,3 - 48,7 \log m$.

Theoretical, expected size	Estimated size from $d = 210,3 - 48,7 \log m$
ca 3789	3981
2359	2570
1595	1698
ca 1327	1349

Fig. 4.6

Responses to hybridization of genomic Drosophila DNA with the 2.4 kb Acc I fragment of pr25.1 as described in Section 4.2.4. Each blot except for the one shown in Fig. 4.6B had three control dots: Canton-S, π_2 and pr25.1. In the blot shown in Fig. 4.6B there were only two controls present as the Canton-S dot broke off.

The blot shown in Fig. 4.6A had 2 μ g (20 ng DNA) of the pr25.1 fragment. The blots shown in Figs 4.6 B-F had 5 μ g (50 ng DNA) each of the pr25.1 fragment.

Six levels of signal intensity were recognized.

ES - extremely strong (50 ng DNA of the Acc I fragment of pr25.1)

VS - very strong (π_2)

S - strong

MD - medium

W - weak

Wf - very weak (Canton-S).

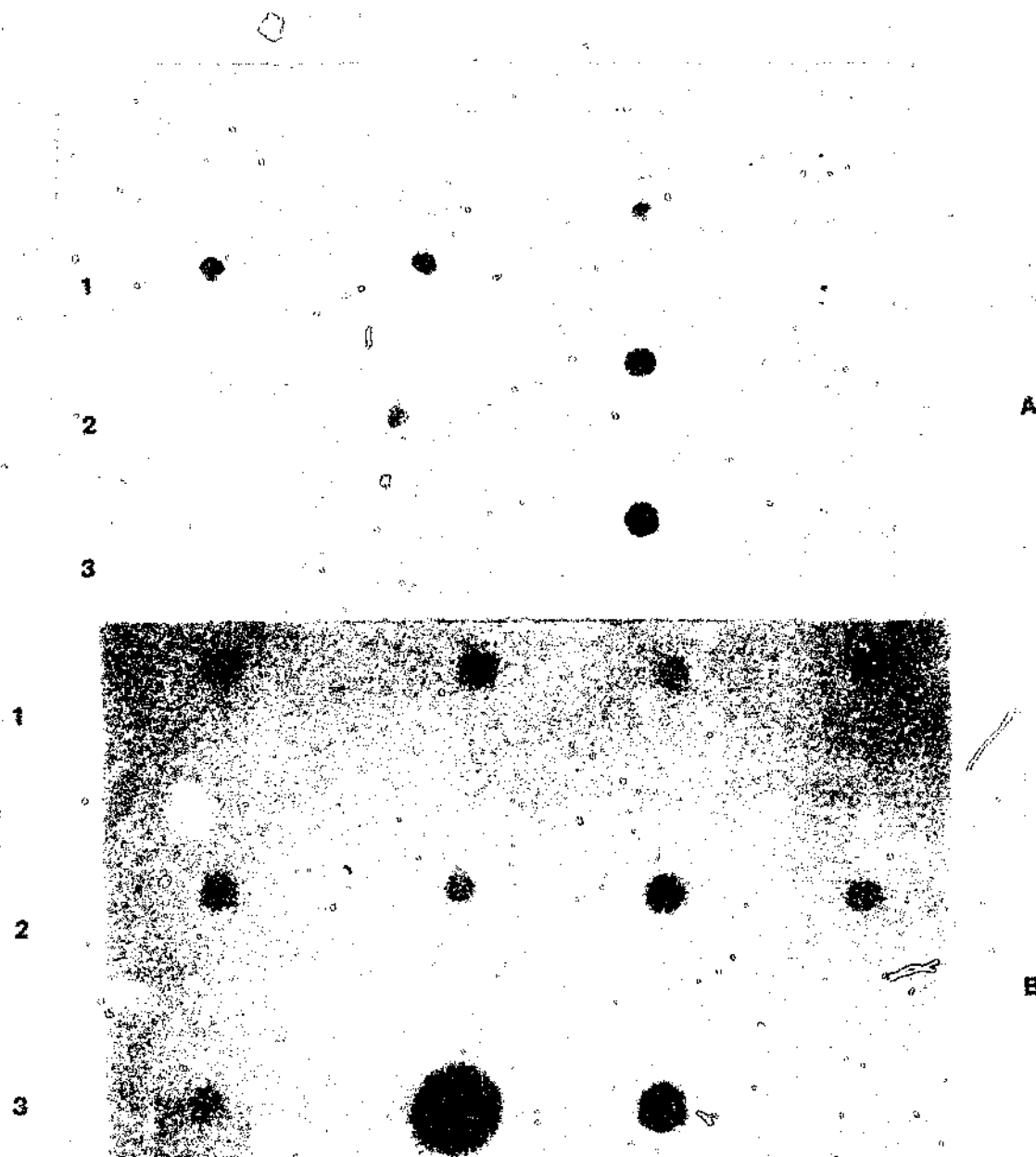


Fig. 4.6 A: Row 1: Rut 1(MD); Rut 2(MD); Ag 1(MD); Ag 2(W).
 Row 2: Plet 6(W); Plet 8(MD); π_2 (VS); Canton-3(W).
 Row 3: Pta 1(VN); Pta 2(VN); pw25.1(2 μ).

B: Row 1: Pta 2(W); Pta 5(MD); Ag 1(VN); Ag 2(MD).
 Row 2: P'B 1(MD); P'B 3(MD); P'B 4(S); P'B 5(W).
 Row 3: GB 1(W); pw25.1(5 μ); π_2 (VS).

1

2

3

1

2

Fig. 4.6**C: Row 1: Br(MD); Plet 6(V); Plet 8(W); Plet 12(VH).****Row 2: Den 1(VH); Den 2(W); Den 3(W); Den 4(VH).****Row 3: H'K 1(W); pr25.1(5 μ l); π_2 (VS); Canton-S(VH).****D: Row 1: GB 2(W); GB 3(VH); GB 4(W); Mac 1(MD); Mac 3(MD).****Row 2: Mac 4(MD); Mac 5(MD); pr25.1(5 μ l); π_2 (VS); Canton-S(VH).**

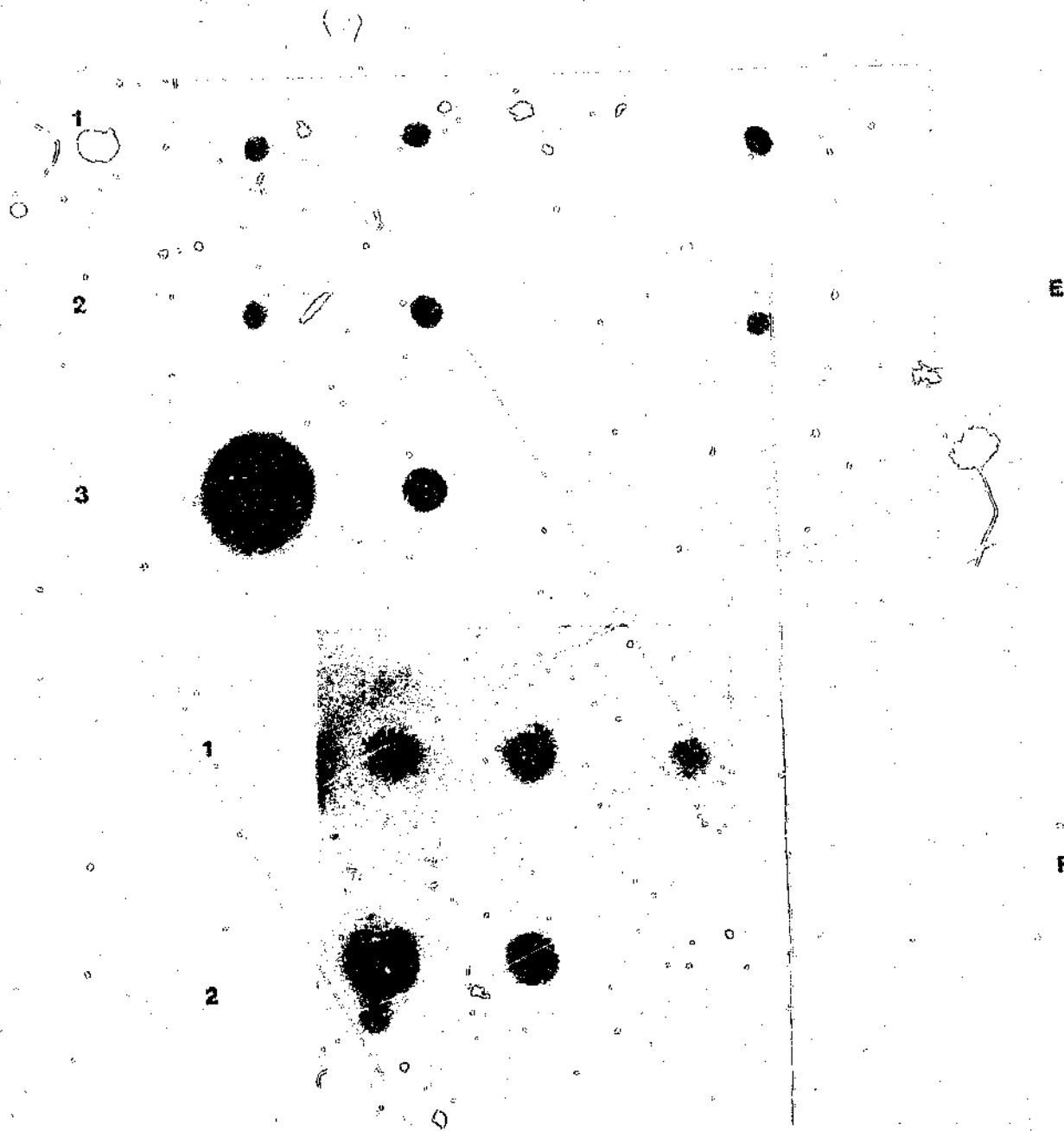


Fig. 4.6

**E: Row 1: H'K 4(MD); H'K 3(MD); P'B 5(VH); P'B 4(S)
 Row 2: Mac 5(MD); Mac 4(S); Rut 2(W); Plot 8(MD)
 Row 3: pr25.1(5 μ l); π_2 (VS); Canton-S(VH);
 Canton S(VH)**

**F: Row 1: H'K 1(MD); H'K 2(MD); H'K 4(W)
 Row 2: pr25.1(5 μ l); π_2 (VS); Canton-S(VH)**

The probe gave a strong signal with π_2 and a much stronger one with the plasmid p π 25.1 (Fig. 4.6). In order to compare the strains six different levels of signal intensity were allocated (in the style of Anxolabéhère et al. 1985b):

- ES - extremely strong (p π 25.1)
- VS - very strong (π_2)
- S - strong
- MD - medium
- W - weak
- VW - very weak (canton-S).

Replicates of several of the populations were tested in order to compare results. Table 4.2 lists the phenotypic and molecular characteristics of the tested strains.

Generally good agreement was obtained between replicates. Apart from two Q strains H'K 1 and Ag 1, (both of which read MD on one blot and VW on another), readings were either identical or else differed by a single gradation. Three of the strains which tested phenotypically as P/P? (excluding π_2) measured either S or MD on this system. However, the sample size of the P type of strain was very small (four strains, and possibly three since Rut 2 was uncertain). Rut 2, which was unclassified, but regarded as a potential P strain, measured MD on one blot and W on another. Of the ten M/M? (or M-Q) wild strains four were MD, four were W and two were VW.

Table 4.2 Molecular characteristics of 31 wild South African strains.

Strain	P-Q-M Designation	Signal-intensity
GB 1	M	W
GB 2	M	W
GB 3	M	VW
GB 4	M	W
Plet 6	M	W/W*
Plet 12	M?	VW
Mac 1	M-Q	MD
Mac 3	M	MD
Br	M-Q	MD
Pta 6	M-Q	MD
Plet 8	Q	W/MD/MD*
Mac 4	Q	MD/S*
Mac 5	Q	MD/MD*
Den 1	Q	VW
Den 2	Q	W
Den 3	Q-M	W
Den 4	Q-M	VW
Pta 1	Q	VW
Pta 2	Q	VW/W*
H'K 1	Q	VW/MD*
H'K 2	Q-P	MD
H'K 3	Q	MD
H'K 4	Q	W/MD*
P'B 3	Q	MD
P'B 5	Q-P	VW/W*
Ag 1	Q	VW/MD*
Ag 2	Q	W/MD*
Rut 1	P?	MD
Rut 2	?	W/MD*
P'B 1	P	MD
P'B 4	P	S/S*
Controls		
"2	P	VS
Canton-S	H	VW

1. VS = very strong; S = strong; MD = medium; W = weak; VW = very weak

* Replicate results

The 17 Q (or Q-M or Q-P) strains varied greatly, four were of intensity MD, two were W and three were VW. The rest measured differently on different blots: two were of intensity VW/W, two were of intensity VW/MD, three were of intensity W/MD, one was of intensity MD/S.

Thus, if one equates intensity of signal with number of P elements then the general pattern that emerges is that overall P strains do have more P elements present than M strains while Q strains are intermediate between the two, although they span the entire range from S to VW. However, it must be pointed out that the sample size of the P strains was very small. The implications of these results are discussed below.

4.4 Discussion

4.4.1 Wild M Strains Possess P elements

All of the wild strains tested gave a positive signal (ranging in intensity from strong to very weak) when probed with the *Aac 1* fragment of the P element. If this signal is interpreted as evidence of the presence of P homologous sequences (see section 4.3.2) then all of the wild flies probed in this study, including those which tested as M on GD sterility tests, possess P element hybridizing sequences. This is in accord with results obtained worldwide. Molecular analyses of wild populations from Japan, Korea, Eurasia, North and South America, Czechoslovakia, the

USSR and Australia all show that recent collections contain P element homologous sequences. (Todo *et al.* 1984; Anxolabéhère *et al.* 1985b, 1988b; Boussy *et al.* 1988). Indeed an examination of strains collected during the past 60 years from several continents revealed that the most recent appearance of true M strains in North America, France and the USSR occurred between 16 and 24 years ago, in 1956, 1969 and 1974 respectively (Anxolabéhère *et al.* 1988b).

4.4.2 Intensity of Hybridization Signal

The intensity of the hybridization signal obtained from the wild strain dot blots varied from strong to very weak. This variation was generally in good agreement with the phenotypic strain classification (Table 4.2).

Similar grading techniques have been employed by other researchers in this field. Anxolabéhère *et al.* (1985b, 1988b) used the "squash blot" technique (whole flies were squashed onto a nylon filter and then probed), and visually classified the variation in the signal intensity. Good *et al.* (1989) squashed single fly ovaries onto a nitrocellulose filter prior to hybridizing. They scored the hybridization signal visually as positive or negative although they did not attempt to quantify the intensity of the signal.

Since the *Aca* 1 fragment has homology with a large section of the complete P element, the variation in intensity of the signals should theoretically be governed by the number of P element

hybridizing sequences present in the strain, i.e. the intensity of the signal should be proportional to the number of P homologous sequences present. However, the proportionality may not be so simple. When the complete P element was used as a probe it was found by some researchers that the overall amount of P element-hybridizing DNA (Boussy *et al.* 1988) or the number of P hybridizing elements (Todo *et al.* 1984) was virtually the same for all strains tested (ranging from M to P).

An alternative explanation for the variation in intensity then is that the *Aco* 1 fragment is preferentially hybridizing to the complete P element and/or to a specific sub-collection of P elements. Probes which are made up of internal segments only (of the P element) will hybridize to fewer P homologous sequences than probes containing end regions. This is because a large number of P elements contain internal deletions (Rubin *et al.* 1982) and in these cases the region of homology to an internal fragment could be very small or non-existent.

Different probes do give different results. Anxolabéhère *et al.* (1985b, 1988b) used two different probes. Their probe P₁ contained the 0.85 kb *Hind* III fragment of the P element while P₂ contained the internal 1.5 kb *Hind* III/*Sal* I P element fragment, (Fig. 4.7). The signal intensity obtained was generally stronger and more uniform throughout the tested range with the shorter P₁ end probe than with the longer P₂ internal probe (Anxolabéhère *et al.* 1985b -- their Table 1; Anxolabéhère *et al.*

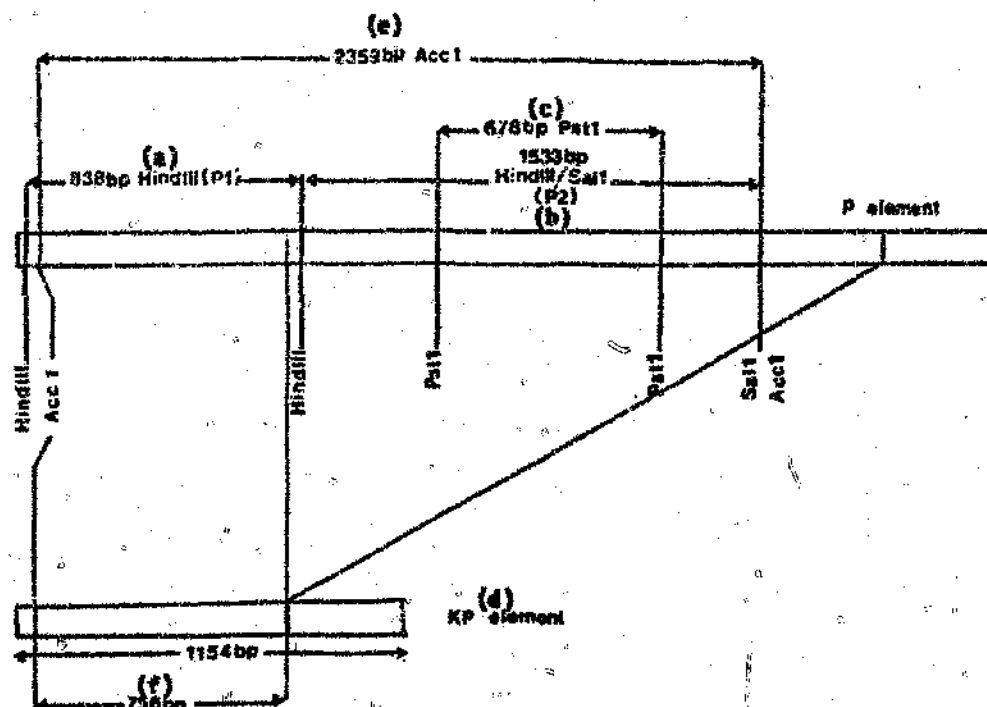


Fig. 4.7

Structure of the 2.9 kb P element in p25.1

The above diagram shows:

- a: The Hind III (P₁) probe (Amoladshere et al. 1985b, 1986b)
- b: The Hind III/Sal I (P₂) probe (Amoladshere et al. 1985b, 1986b)
- c: The Pst I probe (Tode et al. 1984)
- d: The KP element (Black et al. 1987)
- e: The Acc I probe (This study)
- f: The 756 bp of homology between the Acc I fragment and the KP element.

1988b - their Tables 1 and 3). Further Todo *et al.* (1984), using the central portion of the complete P element, the *Pst* I fragment of p π 25.1 (Fig. 4.7), found that the number of bands hybridizing on southern blots of genomic DNA was less than that obtained using p π 25.1 (which contains the complete P element) as a probe. These authors also obtained a correlation between the number of central portion retaining P elements and the phenotypic determination of the strain using this *Pst* I probe.

Thus, if the variation in the intensity is due to the preferential hybridization of the *Acc* I fragment to the complete P element, and if the total amount of P element hybridizing DNA is the same in all strains, then this would explain the strong and medium signals obtained for P strains. These strains have the complete P element present in higher copy number since only the complete P element is able to produce transposase (Rio *et al.* 1986) which is responsible for P element transposition and subsequent GD sterility. Further Boussy *et al.* (1988) in the study of wild Australian flies found the P activity of a line to be positively correlated with the number of full sized P elements in the line - so this would also explain the higher overall intensity of signals for P strains and the lower overall intensity for M strains (Table 4.2). The situation is however complicated by the possible presence of KP elements as outlined in the next section.

4.4.3 Susceptibility to P activity

When P and Q strain females are crossed to P strain males they are able to suppress GD sterility. Further, while true M strains do not suppress P factor activity, M' strains are able to reduce the level of hybrid dysgenesis when crossed to P strain males (Kidwell 1985).

This ability to suppress GD sterility was investigated by Engels (1979b) who used the term P cytotype to designate the ability to suppress GD sterility via the presence of P factors transmitted through the maternal cytoplasm. Further studies with M' strains suggest that suppression may also arise due to the presence of P element deletion derivatives. Kidwell (1985), in a genetic analysis of M and M' strains showed that a chromosomal component transmitted through both sexes was instrumental in suppressing GD sterility. Black *et al.* (1987) in a molecular and genetic analysis of KP elements (P factor internal deletion derivatives cloned out of the Russian strain Krasnodar (Fig. 4.7)) concluded that these elements contribute greatly to the suppression of hybrid dysgenesis in natural populations. They suggested that this type of suppression was distinct from the P cytotype system.

The molecular analysis of Australian lines (Boussy *et al.* 1988) revealed that the large majority of these lines contained only full sized P elements and KP elements. Further, KP elements have been shown to increase in number in M and Q lines (Boussy

et al. 1988) and to aid in the suppression of GD sterility[†] (Black *et al.* 1987). Thus, if signal intensity using the *Aac* 1 probe is in any way indicative of the number of KP elements present (the *Aac* 1 fragment and the KP element have 756 bp's of homology - Fig. 4.7), and if an increase in the number of KP elements implies an increase in regulatory ability, then the stronger the signal the greater the regulatory ability.

If the *Aac* 1 fragment is hybridizing to KP elements, then a good indication of the number of KP elements present could be obtained by examining M strains which are presumably free of complete P elements (due to their lack of P element activity) and also possibly free of deletion derivatives other than the KP element (Boussy *et al.* 1988). From Table 4.2 and Fig 4.8, it can be seen that, in general, as far as M strains are concerned, the greater the intensity of the signal the greater the ability of the strain to suppress GD sterility. (Using Spearmans rank correlation coefficient a highly significant ($r_s = 0.962$, $P < 0.001$) negative correlation was obtained between cross A* GD sterility and signal intensity for M strains.) A

[†]Although Boussy *et al.* (1988) found no correlation among the Q and M lines between regulatory ability and the number of KP elements, and Monastiroti *et al.* (1988) produced evidence which suggested that under certain circumstances KP elements may not be effective in suppressing hybrid dysgenesis.

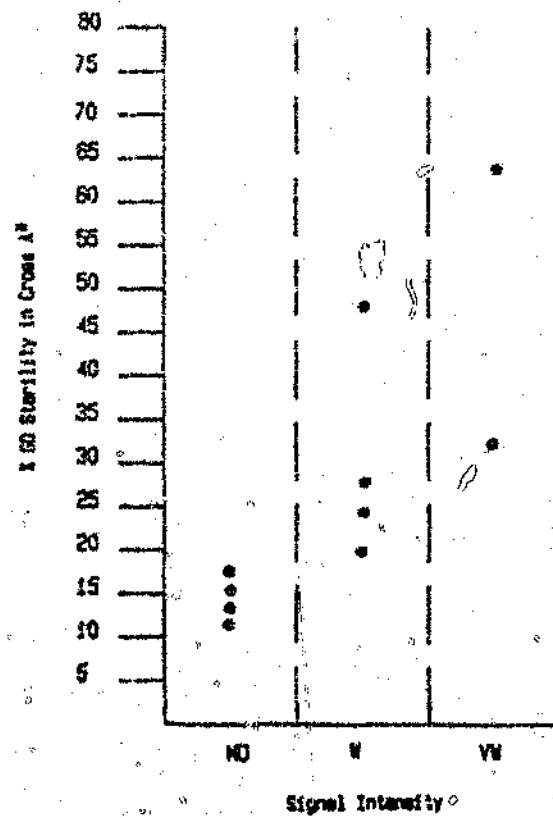


Fig. 4.8

Relationship between intensity of hybridization signal and cross A* percentage 6D sterility for H strains.

tentative explanation for this is that, the greater intensity of an M strain signal implies the greater number of KP elements present and consequently the greater the regulatory ability of that strain. However this is merely an hypothesis since, KP elements may not even be present in these strains.

4.4.4 Q Strains

Probing for P elements in Q strains has often resulted in high levels of hybridization (ingham *et al.* 1982; Sakoyama *et al.* 1985; Anxolabéhère *et al.* 1985b, 1988b; Boussy *et al.* 1988) revealing the presence of a large number of P homologous elements in these strains. In this study the Q strains gave signals ranging from strong to very weak, but most of them were in the medium to weak range. While the P factor activity in causing GD sterility is suppressed in Q lines, other P factor activities (such as MR and *en^W* mutability) is not (Simmons *et al.* 1987). Consequently the signal intensity obtained for the strains was compared with the percentage *en^W* mutability induced by them. No correlation was obtained between the two results and this lack of correlation between the number of P homologous elements (as evidenced by the intensity of the signal) and the amount of P transpositional activity (as evidenced by the percentage of *en^W* mutability) could be accounted for in terms of the transposase titration theory of Simmons and Bucholz (1985) (see section 2.4.3.1).

If transposase is in limited supply due to a small complement of full-sized P elements, then competition for transposase by defective (and intact) P elements could lead to a shortage of transposase which in turn would lead to a reduction of P transpositional events. (Experimental evidence for this was obtained by Kidwell (1985), who found that the presence of P elements in *M* strains suppressed P factor induced hybrid sterility to a certain degree.) Thus, a reduced *sn^W* mutability does not necessarily imply the presence of fewer P elements. It could be that the small amount of transposase being produced by the complete P element(s) is being competed for by a larger number of defective P elements and thus creating a shortage of transposase in the vicinity of the elements responsible for the *sn^W* mutability effects (Roiha et al. 1988).

A reasonable hypothesis is that, for *O* strains, the number of complete P elements is small and the number and kind of defective P elements is variable. The strength of the hybridization signal is proportional to the total number of intact and defective P elements present, but the amount of mutability displayed is dependent on a delicate balance between the amount of transposase being produced by the complete P elements and the number of all P elements complete and defective competing for this transposase. As a result *sn^W* mutability would not necessarily be proportional to the number of P elements present in the genome.

CHAPTER FIVE

5. P element Somatic Activity

5.1 Introduction

P element hybrid dysgenic traits have been hypothesized to be restricted to the germline (Engels and Preston 1979). The lack of phenotypic evidence of somatic effects together with the molecular and cytological indications of germline specificity seem to bear out this claim.

Firstly, Bingham *et al.* (1982) crossed dysgenic hybrid males carrying an M type X chromosome but heterozygous for M and P autosomes to attached X females and analysed the male F_1 and F_2 progeny by *in situ* hybridization of the salivary gland chromosomes with a P element containing probe. The authors detected no copies of the P element in the salivary gland X chromosome squashes of 15 F_1 males. On the other hand, in a total of 22 F_2 male X chromosomes, Bingham *et al.* (1982) observed 18 copies of the P element. These results indicated the rare or non-existent transposition of P elements onto M type X chromosomes in the somatic cell precursors of the salivary glands as opposed to the transposition of P elements at the rate of 0.8 copies per X chromosome in the germline cells.

Secondly, whereas transcription of P elements takes place in both somatic and germline cells, splicing of the third intron does not appear to take place in somatic tissue (Laski *et al.* 1986). As regards translation, Rio *et al.* (1986), using

immunochemical techniques, identified two proteins encoded by P transposable elements expressed in transformed *Drosophila* tissue culture cells. One, an 87 000 dalton polypeptide was encoded by a P element mRNA derived from the engineered DNA P element $\Delta 2-3$ ($\Delta 2-3$ is a full P element but has the third intron precisely excised). The other was a 66 000 dalton polypeptide which was encoded by P element mRNA that retained the third intron. Thus, it is not entirely clear whether *in vivo* translation of incompletely spliced mRNA takes place, although the 66 000 dalton protein was precisely the size polypeptide that would be encoded by the first three ORFs of the complete P element (translation of the mRNA being terminated due to the presence of stop codons introduced by the non-excision of the third intron).

Finally, transposition of P elements in somatic tissue *in vivo* was observed in the progeny of flies transformed with the $\Delta 2-3$ P element (Laski *et al.* 1986; Engels *et al.* 1987; Robertson *et al.* 1988). The above facts have led to the conjecture that the splicing of the third intron occurs only in the germline cells leading to the production of the 87 000 dalton protein which effects mobilization of P elements in these cells (Laski *et al.* 1986; Rio *et al.* 1986).

5.1.1 Motivation and Background

Although most phenotypic tests to date point to the tissue specific movement of P elements, and the molecular research (mentioned above) is strong evidence of this, nonetheless somatic mobilization of P elements cannot be totally ruled out. Previous

Fig. 5.1

Genetic development which could lead to formation of spots. Each diagram illustrates the left arm of chromosome 3 in a trans-heterozygous msh/fir³ cell prior to mitosis and the left arm of these chromosomes in the two cells formed subsequent to mitosis.

- A: Normal cells
- B: A small deletion gives rise to a msh single spot.
- C: A point mutation produces a msh single spot.
- D: A msh single spot is produced by recombination between the msh and fir³ markers.
- E: A twin spot is obtained by recombination proximal to the fir³ markers.
- F: A double crossover between msh and fir³ and between fir³ and the centromere leads to a fir³ single spots.

first two rows consisting of wings from female flies and the second two rows consisting of male fly wings. The wings were allowed to dry on the slide in a dust-free atmosphere for at least 24 hours. A droplet of Faure's solution was put on a coverslip which was placed on top of the wings. The coverslip was charged for at least 24 hours with several metal cubes (about 40 g each) while the preparation was drying and hardening.

The wings were analysed for the presence of spots under a compound microscope at 400 x magnification. Both dorsal and ventral surfaces were analysed simultaneously. The position of the spots were noted according to the sector of the wing (Graf *et al.* 1984). Only the distal wing compartment (Garcia-Bellido *et al.* 1976) was scored for clones (Fig. 5.2). The size and type of each spot was recorded. Scanning electron micrographs of *mw* and *flr*³ trichomes are shown in Fig. 5.3.

5.2.3.3 Scoring of wings

Large clones normally extend parallel to the longitudinal axes of the wing (Garcia-Bellido and Merriam 1971). Marked clones on the wing blade generally appear as contiguous non-interrupted spots. However, sometimes the spots are split into two or more cell groups separated by one or several rows of normal cells due to internal tissue pressure on independent cell movements during development (Garcia-Bellido and Merriam 1971). Consequently, spots separated by one or two wild-type cell rows were scored as a single entity. I followed Graf *et al.* (1984) for techniques and rules regarding spot classification.



Fig. 5.3

Scanning electron micrographs of wing blade trichomes supplied by Dr. K. Joon and Professor F.E. Mörzler from the Institute of Toxicology, Schwerzenbach.

A: Normal trichome

research studies designed to detect somatic effects of P elements in the wings (Thompson *et al.* 1978; McElwain 1986), in the eyes (Thompson *et al.* 1978) and in the abdomen (McElwain 1986) of dysgenic hybrids have proved to be negative. These experiments were carried out at developmental temperatures of 25°C. Since high temperatures (of approximately 29°C) play an important part in P element dysgenesis and since I do not feel that conclusive proof of the absence of P element activity in somatic tissue has been obtained, I investigated the possibility of P element induced mutational activity in the wings of hybrid flies. Because of the following two observations I have looked for evidence of a mitotically induced temperature sensitive event. Firstly, investigations of the dysgenic phenomenon of male crossing over have revealed that a considerable portion (if not all) of this event occurs in the mitotically dividing pre-meiotic germ cell lines (Hiraizumi *et al.* 1973; Kidwell and Kidwell 1976; Woodruff and Thompson 1977; Hiraizumi 1979; Isackson *et al.* 1981; Sinclair and Grigliatti 1985). Further the GD sterility temperature sensitive period begins in the late embryonic stages and continues through the first and second instars. It thus appears that GD sterility is caused by the failure in the early development of germ cells which may be taking place at the onset of mitosis (Engels and Preston 1979; Kidwell and Novy 1979; Niki and Chigusa 1986). Thus, somatic movement of P elements, if it exists, may be mitotically induced.

Secondly, apart from the well documented fact that a high developmental temperature of 29°C increases the occurrence of GD sterility (for example Kidwell and Novy 1979), high temperature has also been shown to enhance male recombination (Yannopoulos and

Pelecanos 1977) and to reduce male mating competitiveness in dysgenic hybrids (Colgan 1982). Further, Hihara (1987) observed that female lethality as a result of a developmental retardation in dysgenic females increased with increasing developmental temperature. So a high developmental temperature may be operative in inducing somatic mobilization of P elements. In this study, I scored the frequency of marked clones in the wings of dysgenic and non-dysgenic hybrid flies of the genotype *mwh/fir³* reared at temperatures of 25°C and 29°C, in order to test for somatic P element mobilization.

5.2 Materials and Methods

5.2.1 *Drosophila* Stocks

Genetic symbols not otherwise explained are in Lindsley and Grell (1968). The mutation *flare* is described by Garcia-Bellido and Dapena (1974).

w² : see section 2.2.3.

y an^w(M)/y⁺Y; bw; st (an^w(M)) : see section 2.2.3.

mwh : flies homozygous for the marker multiple wing hairs on chromosome three.

fir³/TM3 Ser : TM3 is a multiply inverted balancer chromosome. *Ser*, a dominant marker (*Serrate*), and *fir³* (an allele of *flare* (*fir*)) are both homozygous lethal.

This strain will hereafter be referred to as *flr³/Ser*.

The strains *flr³/Ser* and *moh* were obtained from the Institute of Toxicology, Swiss Federal Institute of Technology and University of Zurich, Schwerzenbach.

mmoh

The first and second chromosomes are derived from the π_2 line and so have a number of P elements present. Chromosomes three and Y are derived from the standard *moh* line. Chromosome four is uncontrolled. This strain was constructed by one of my supervisors (N. van Schaik) in 1985.

5.2.2 Testing of Stocks

GD sterility tests using cross A and cross A* (Section 2.2.4) were performed on the stocks *moh*, *flr³/ser* and *mmoh* in order to measure their potential for P factor activity and their regulatory ability. In addition, when testing *mmoh*, in order to ensure that the results obtained from the dissection of females in crosses A and A* were not due to spontaneous background or temperature induced gonadal dysgenesis, I performed reciprocal crosses (B and B*) on this strain (Section 2.2.4). Verification

of the status (dysgenic or non-dysgenic) of the hybrid flies whose wings were to be examined for mutant clones was obtained by dissecting the ovaries of samples of the progeny obtained from each of the relevant crosses. All crosses were performed at 29°C. Finally the P factor activity of the strain *mmh* was also measured using the *en*^W mutability test.

5.2.3. Testing for Somatic P element Activity in Dysgenic Flies

This assay system makes use of the two wing cell markers *mmh* (3-0,0) and *flr*³ (3-39,0). The wings of both dysgenic and non-dysgenic flies trans-heterozygous for *mmh* and *flr*² were scored for mutant clones. These flies were obtained from the following crosses, all of which were performed at both 25° and 29°C:

Cross I : *flr*³/*Ser* ♀ x *mmh* ♂

Cross II : *mmh* ♀ x *flr*³/*Ser* ♂

Cross III : *flr*³/*Ser* ♀ x *mmh* ♂

Cross IV : *mmh* ♀ x *flr*³/*Ser* ♂

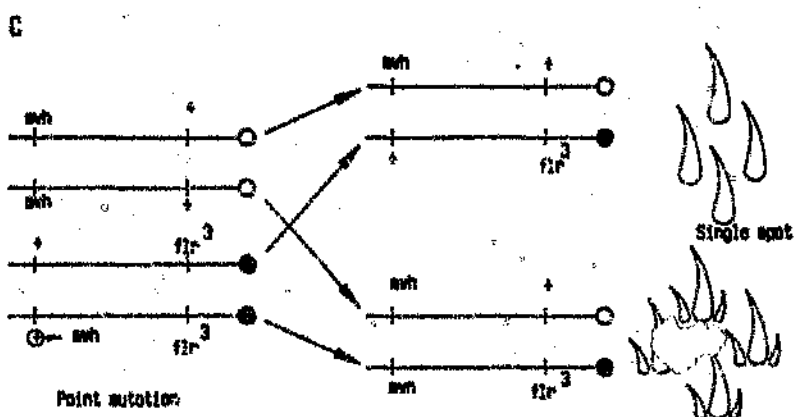
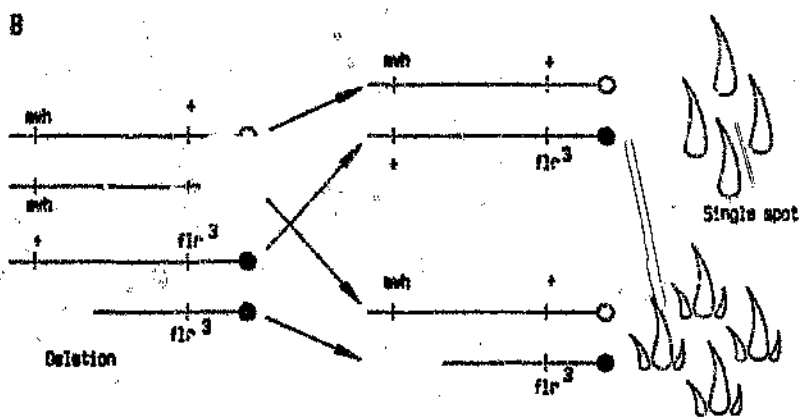
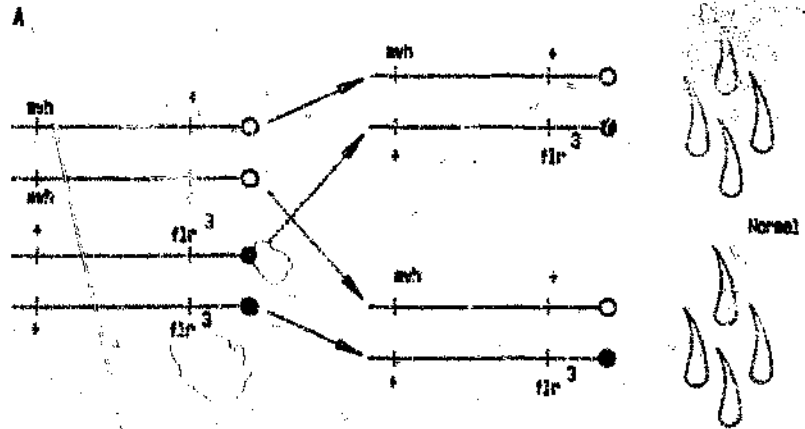
Cross I is the dysgenic cross, crosses II, III and IV are the controls. Crosses II and IV are the reciprocals of crosses I and III respectively. Progeny from cross III have the same maternal genetic background as well as the same paternal third chromosome as those from cross I.

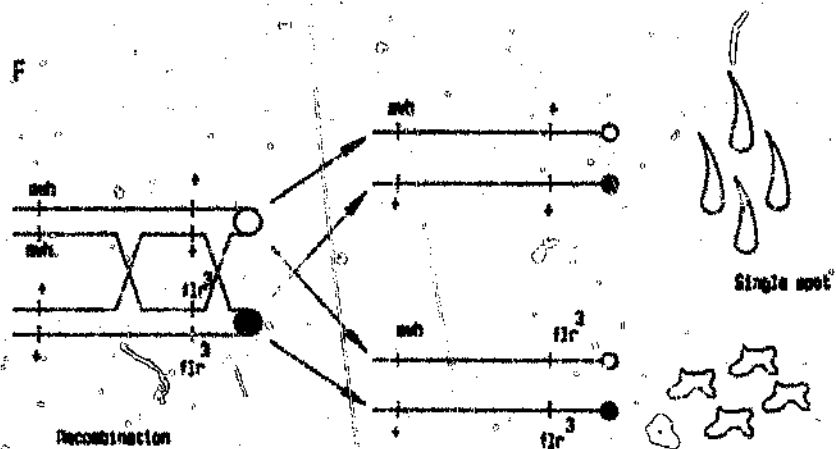
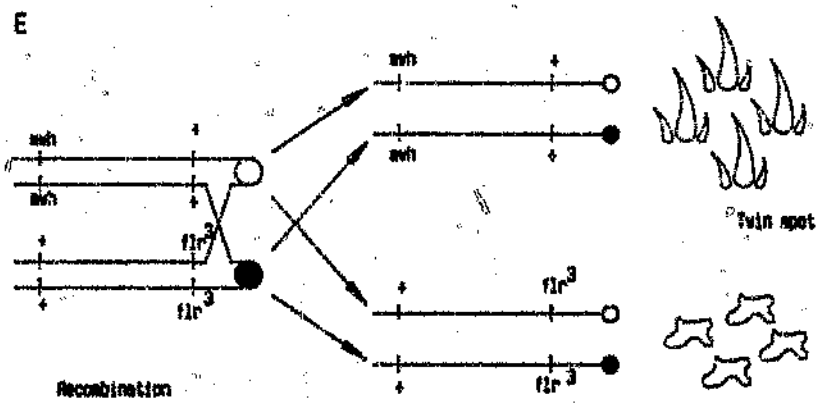
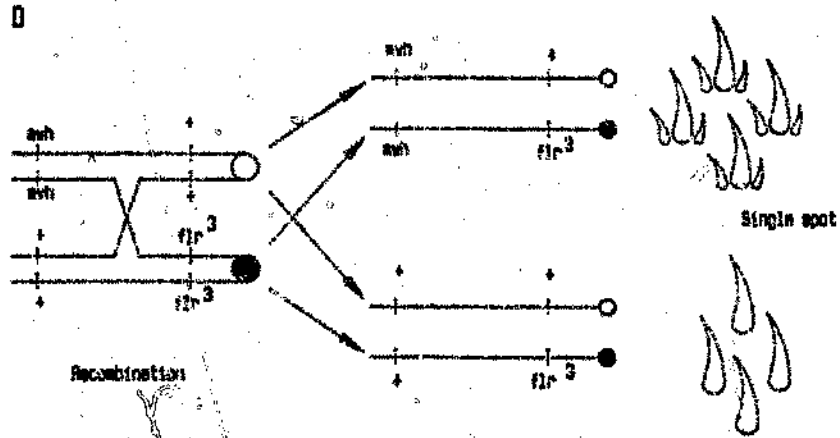
5.2.3.1 Clone formation

During larval development, any change in the wing imaginal disc cells resulting in a change in the genotype of the markers will, after metamorphosis, be observed as marked clones derived from these cells i.e. phenotypically mutant spots will be apparent on an otherwise wild type wing surface. Changes in genotype can be brought about in a number of ways (Fig. 5.1). Somatic recombination in the chromosome region between the centromere (at position 47,7) and *fLr³* leads to twin spots (*msh/fLr³*). Somatic recombination between the markers, *msh* and *fLr³*, leads to *msh* single spots. Single spots for *msh* and for *fLr³* may also arise from mutation or deletion of the respective wild type alleles in the double heterozygote. Also, the extremely rare event of a double crossover between *msh* and *fLr³* and between *fLr³* and the centromere could lead to single *fLr³* spots. Single spots could also arise if one of the cell lineages in an original twin spot is damaged in such a way that it becomes lethal.

5.2.3.2 Preparation and microscopic analysis of wings

For the preparation and analysis of the wings I have followed the methods of Graf *et al.* (1984). *Sex* flies were discarded. The remaining flies were stored in 70% alcohol until needed, at which time they were washed in distilled water and then transferred to a drop of Faure's solution (gum arabic 30 g, glycerol 20 ml, chloral hydrate 50 g, H₂O 50 ml) in a well slide. The wings were separated from the body and then placed flat on a clear slide, generally in rows of ten. The two wings of each individual were laid side by side and completed slides usually contained 40 wings made up of four rows of ten wings each. The





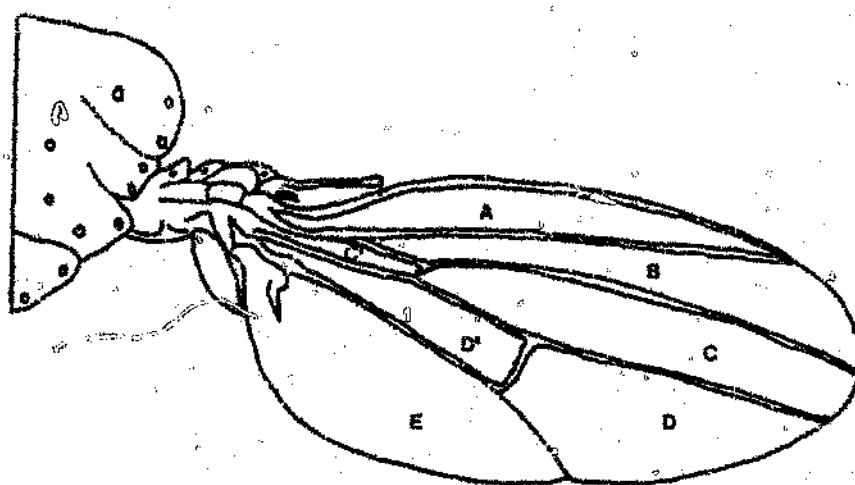
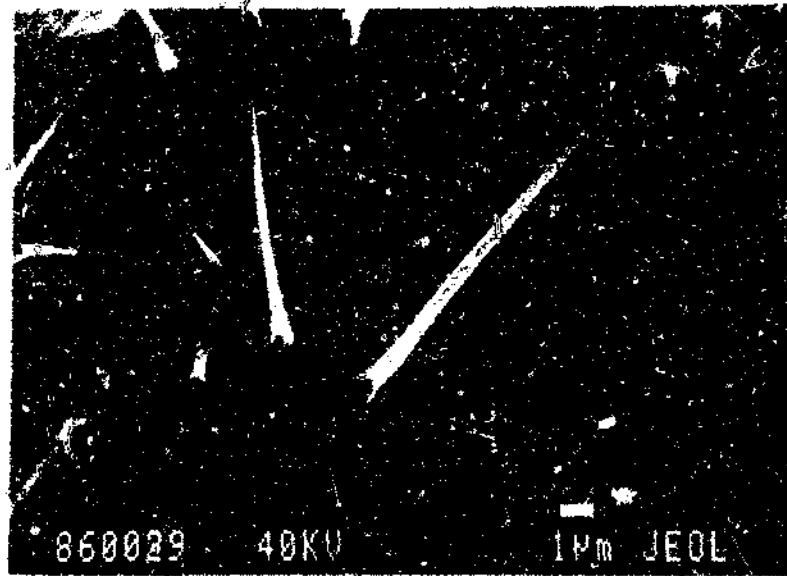


Fig. 5.2

Half mesothorax showing the regions A-E of the wing surface scored for spots (after Garcia-Bellido and Merriam 1971).

B



C

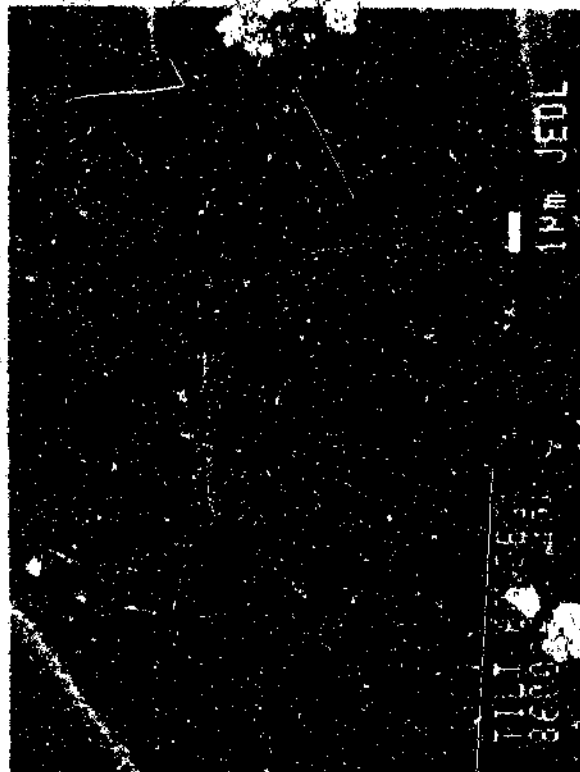


Fig. 5.3

B: mb trichomeC: flr³ trichome

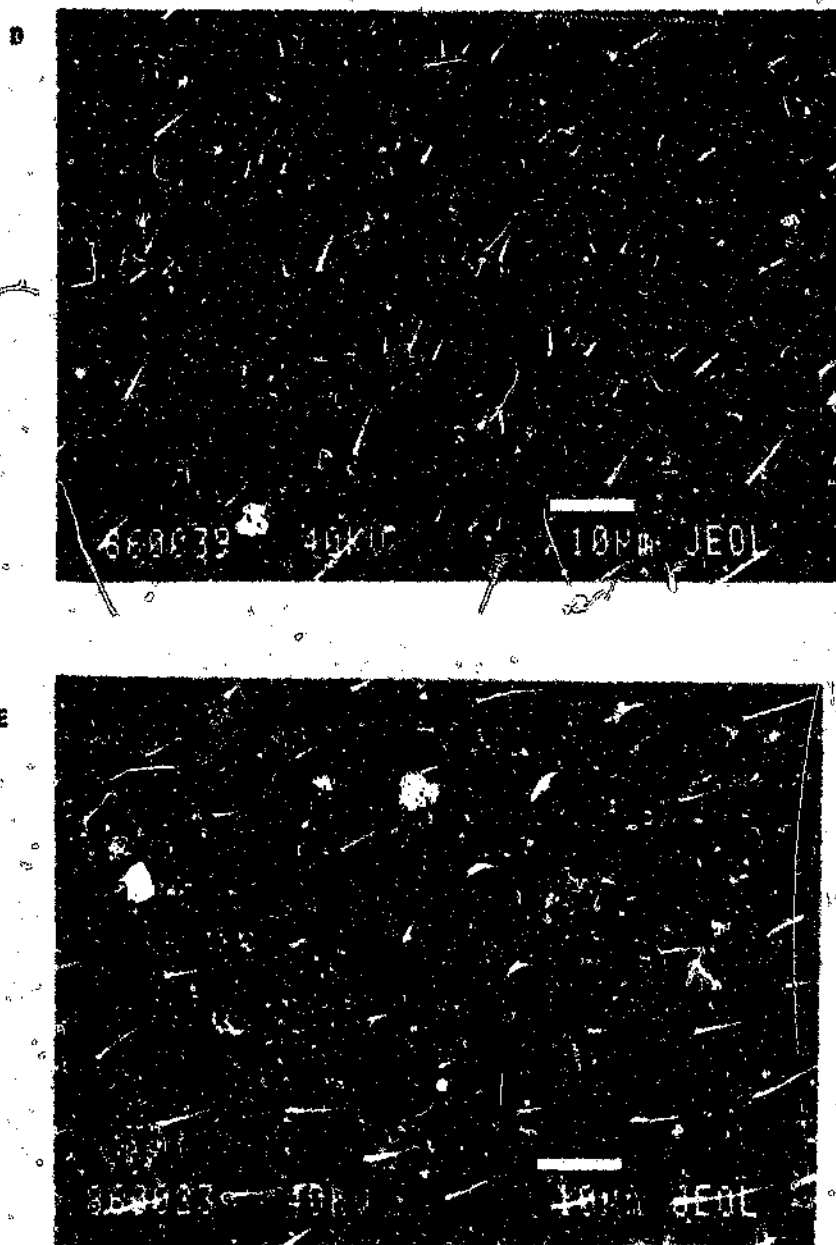


Fig. 5.3

- D: Multi-celled amo spot against a normal background.
- E: Multi-celled amo spot against a normal background.

5.2.3.4 Statistical procedures

After ascertaining the Poisson nature of the data distribution, χ^2 tests on males and females were performed separately at each temperature (25° and 29°C). If significant results were obtained I went on to test for a significant difference between the Poisson means obtained in the dysgenic cross and in each of the three control crosses, using the z statistic,

$$z = (\hat{\lambda}_1 - \hat{\lambda}_2) / (\hat{\lambda}_1/n_1 + \hat{\lambda}_2/n_2)^{1/2}$$

which is $N(0,1)$ under the null hypothesis ($\hat{\lambda}_i$ = Poisson mean; n_i = number of wings tested, $i=1,2$). I used the Bonferroni level of significance; that is, to obtain an overall significance level of α , each of the three controls had to be significantly different from the dysgenic cross at a level of $\alpha/3$.

5.3 Results

5.3.1 GD Sterility

Using the classification system of Kidwell (1983) (<10% GD sterility in cross A and >10% in cross A* indicates an M type strain whereas the converse indicates a P type strain) I was able to classify all the strains unambiguously. GD sterility tests (Table 5.1) classified the stocks *mwh* and *flr³/Ser* as M lines and the *nmwh* stock as a strong P line. The strong P factor activity of *nmwh* was confirmed by the *an^w* mutability test when 48 *an⁺* and 49 *an^e* out of a total of 324 individuals was obtained yielding a 30% *an^w* mutability rate. Further, the progeny of the cross *flr³/Ser* x *nmwh* σ were dysgenic and 96.7% of the ovaries of the hybrid females were atrophied; the progeny from the three control crosses were not dysgenic (Table 5.1).

Table 5.1

A Results of GD sterility tests performed on the stocks *flr³/Ser*, *msh* and *nmsh* (Section 2.2.4).

Classification of the (dysgenic or nondysgenic) the hybrid flies whose wings were examined clones.

A

Strain (T)	Cross ¹	n ²	% GD sterility	Type
<i>flr³/Ser</i>	A	100	0	M
	A*	100	100	
<i>msh</i>	A	100	6	M
	A*	100	100	
<i>nmsh</i>	A	100	100	P
	A*	100	0	
	B	100	0	
	B*	100	0	

B

female	Cross male	n ²	% GD sterility	State
<i>flr³/Ser</i>	<i>nmsh</i>	120	96,7	dysgenic
<i>nmsh</i>	<i>flr³/Ser</i>	120	0	nondysgenic
<i>flr³/Ser</i>	<i>msh</i>	120	0,8	nondysgenic
<i>nmsh</i>	<i>flr³/Ser</i>	120	1,7	nondysgenic

¹ Cross A : $sn^W(M) \text{ } \sigma \times T \text{ } \sigma$

Cross A* : $T \text{ } \sigma \times \pi_2 \text{ } \sigma$

Cross B : $T \text{ } \sigma \times sn^W(M) \text{ } \sigma$

Cross B* : $\pi_2 \text{ } \sigma \times T \text{ } \sigma$

T is the strain being tested; $sn^W(M)$ and π_2 are described in Section 2.2.3.

² n is the number of ovaries.

5.3.2 Mutant Clones

The wings of flies having the genotype *flr³/msh* were scored for mutant clones. The results are presented in Table 5.2 (for flies reared at 25°C) and Table 5.3 (for flies reared at 29°C). Results are given for males and females separately.

The first entries in Tables 5.2 and 5.3 give the results of the experimental cross *flr³/Ser* ♀ x *msh* ♂; the remaining three entries list the results obtained from the control crosses. In all the series analysed the most frequent type of spot found was the single spot expressing the *msh* phenotype. Less frequent were twin spots with both a *msh* and a *flr³* sub-clone. I found no evidence of single spots with a *flr³* phenotype. At 29°C the frequency of spots in general was much higher than that at 25°C, showing that the expression of *msh* spots is effected by high temperatures.

All data sets were verified to be Poisson distributed. The χ^2 analysis of the 25°C data set revealed no significant heterogeneity in the frequency of spots among the four crosses for either males or females separately or together. I have thus confirmed the results of Thompson *et al.* (1978) and McElwain (1986). Both of these authors performed similar experiments to this one and found no evidence of somatic P element activity at 25°C.

The χ^2 analysis of the data obtained for the flies reared at 29°C was highly significant ($P=0.005$; $\chi^2_6 = 19.8$). Subsequent tests showed that this high degree of heterogeneity was due to the

Table 5.2

Number of wings of flies reared at 25°C having 0, 1, 2 or ≥ 3 spots per wing.

Cross	Female Progeny						Male Progeny					
	0	1	2	≥ 3	N ¹	λ^2	0	1	2	≥ 3	N ¹	λ^2
I <i>flr</i> ² /Ser ♀ x <i>nmh</i> ♂	25	8	1	-	34	0,294	35	9	-	-	44	0,204
II <i>nmh</i> ♀ x <i>flr</i> ² /Ser ♂	25	14	1	-	40	0,400	33	7	-	-	40	0,175
III <i>flr</i> ² /Ser ♀ x <i>nmh</i> ♂	27	9	4	-	40	0,425	31	7	2	-	40	0,275
IV <i>nmh</i> ♀ x <i>flr</i> ² /Ser ♂	29	10	1	-	40	0,300	32	8	-	-	40	0,200

 $\chi^2_3 = 1,38$ for the females ($P > 0,05$)

 $\chi^2_3 = 1,11$ for the males ($P > 0,05$)

¹N is the total number of wings scored

² λ is the mean number of spots per wing.

Table 5.2 Number of wings of flies reared at 25°C having 0, 1, 2 or ≥3 spots per wing.

Cross	Female Progeny						Male Progeny					
	0	1	2	≥3	N ¹	χ ²	0	1	2	≥3	N ¹	χ ²
I flr ³ /Ser ♀ x nmah ♂	25	8	1	-	34	0,294	35	9	-	-	44	0,204
II nmah ♀ x flr ³ /Ser ♂	25	14	1	-	40	0,400	33	7	-	-	40	0,175
III flr ³ /Ser ♀ x nmah ♂	27	9	4	-	40	0,425	31	7	2	-	40	0,275
IV nmah ♀ x flr ³ /Ser ♂	29	10	1	-	40	0,300	32	8	-	-	40	0,200

$\chi^2_3 = 1,38$ for the females ($P > 0,05$)

$\chi^2_3 = 1,11$ for the males ($P > 0,05$)

¹N is the total number of wings scored

²χ is the mean number of spots per wing.

Table 5.3 Number of wing spots of flies reared at 29°C having 0, 1, 2 or ≥3 spots per wing.

Cross	Female Progeny						Male Progeny					
	0	1	2	≥3	N ¹	λ ²	0	1	2	≥3	N ¹	λ ²
I <i>flr</i> ³ / <i>Ser</i> ♀ x <i>nmsh</i> ♂	28	35	17	6	87	1,034	47	31	8	2	88	0,602
II <i>nmsh</i> ♀ x <i>flr</i> ³ / <i>Ser</i> ♂	30	19	9	-	58	0,638	34	18	6	-	58	0,517
III <i>flr</i> ³ / <i>Ser</i> ♀ x <i>nmsh</i> ♂	48	23	7	-	78	0,474	52	26	2	-	80	0,375
IV <i>nmsh</i> ♀ x <i>flr</i> ³ / <i>Ser</i> ♂	44	26	9	1	80	0,587	40	25	15	-	80	0,687

$\chi^2_{6} = 18,86$ for the females ($P < 0,005$)

$\chi^2_{6} = 11,92$ for the males ($P > 0,05$)

¹N is the total number of wings scored

²λ is the mean number of spots per wing.

contribution of the females ($P < 0.05$; $\chi^2_6 = 18.9$) but not the males ($P > 0.05$; $\chi^2_6 = 11.9$). The Poisson mean for the 29°C female dysgenic cross was significantly higher ($P < 0.005$) than that of each of the three 29°C controls ($z = 2.62$, $z = 4.18$, $z = 3.22$ for crosses II, III and IV respectively). Thus, for flies reared at 29°C there is a significantly higher ($P_{\text{Bonferroni}} < 0.02$) frequency of spots on the wings of female dysgenic flies than on the wings of female flies obtained from the corresponding non-dysgenic crosses. I postulate that this increase in frequency may be due to P element mobilization in the wing cells.

5.3.3 Proportion of Different Spot Sizes for Single Spots

In theory, the size of a mutant spot indicates the number of cell divisions that have taken place since the induction of the genetic change in the precursor cell from which the clone originated. Thus assuming a constant rate of cell division, a spot size of 2^n would result from n cell divisions. Although this relation is usually not maintained and clones of 3, 5, 6 and 7 cells (Garcia-Bellido and Merriam 1971; Graf *et al.* 1984) have been found, I nonetheless felt it was worthwhile to classify the spots according to size classes (1, 2, 3-4 and ≥ 5) and investigate the distribution of these classes.

If P element activity is induced by mitosis then one would expect to obtain a high percentage of cells manifesting P element induced events at times of high mitotic index. Stumpf (1956) detected a very high mitotic index in the wings at the time of puparium

formation (PF) or later, which reached a maximum at PF + 17 hours. Further, these mitoses, which take place after PF, probably represent the last ones occurring in the wing disc (Garcia-Bellido and Merriam 1971). Thus, if single cell spots indicate mutations which have arisen in the final mitotic division of a cell lineage, then P element functioning might be expected to induce a high percentage of one cell spots. Fig 5.4 shows the size distribution of clones at 29°C. For the males, there was no significant difference in distribution but 73% of the single spots in the females of the dysgenic (29°C) cross consisted of a single cell and there was significant heterogeneity ($\chi^2_3 = 8.5$; $P < 0.05$) in the spot size distribution at 29°C for the females. However a comparison of the proportions of one cell spots revealed no significant difference between the proportion of this distribution and two of the three (female, 29°C) controls. The only control which had a significantly different proportion of one cell spots from the dysgenic cross ($P < 0.05$) was Cross III (*fLr³/Ser* ♀ x *msh* ♂) where 53% of the single spots consisted of one cell only. No conclusions could therefore be drawn regarding the timing of P element activity.

5.4 Discussion

These experiments have examined the frequency of mutant clones in the wings of flies trans-heterozygous for the wing markers *msh* and *fLr³*. At developmental temperatures of 29°C, the frequency of these clones was found to be significantly higher in females derived from crosses between M strain females and P strain males than in females derived from the reciprocal crosses

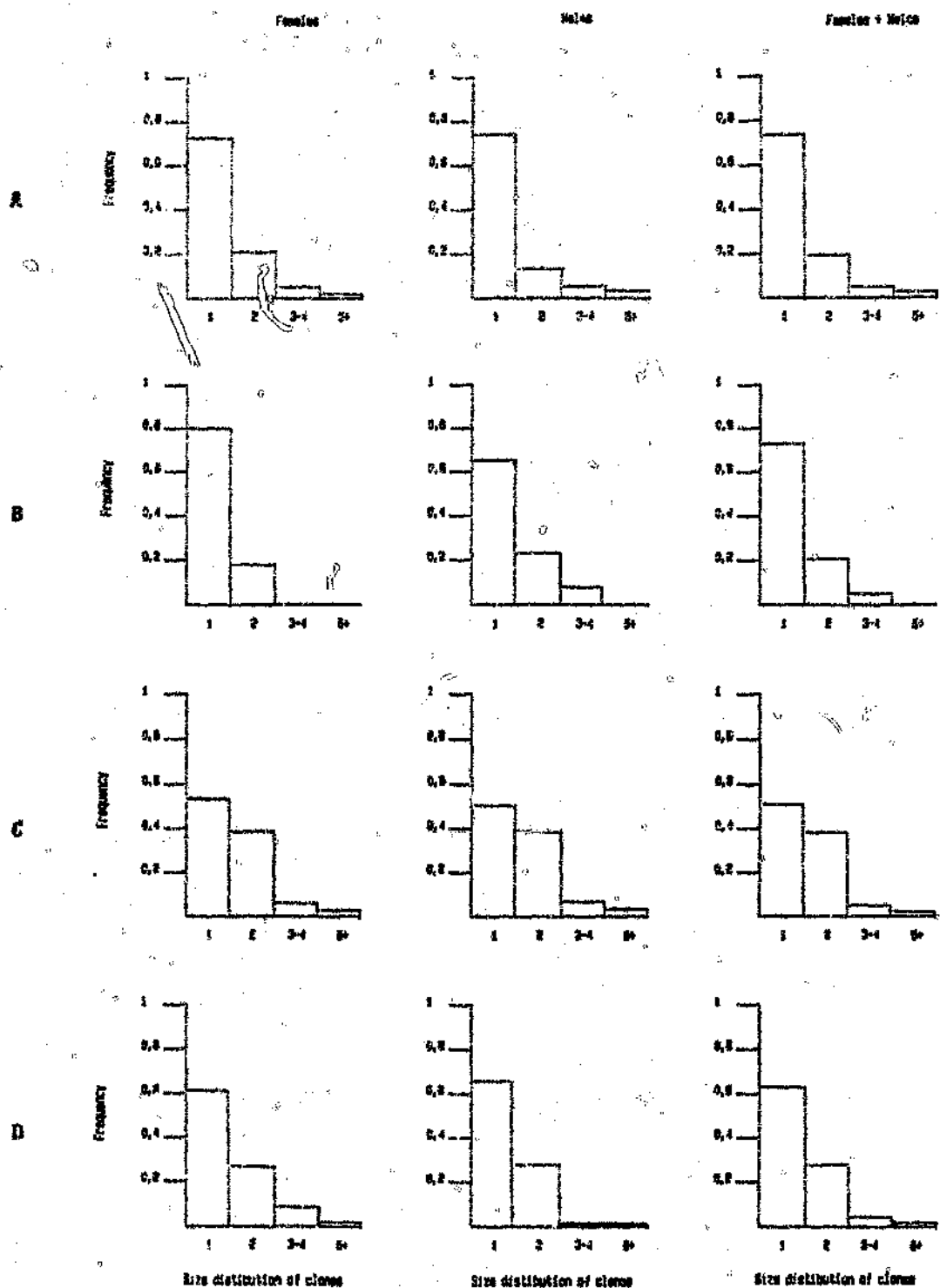


Fig. 5.4

Distribution of clones obtained at 25°C according to size classes. Frequencies were calculated by dividing the number of single spots of size 1-cell, 2-cell, 3-4 cell and 5-cell or larger by the total number of single spots.

A: $\text{flr}^3/\text{Ser} \varnothing \times \text{wmh} \sigma$

B: $\text{wmh} \varnothing \times \text{flr}^3/\text{Ser} \sigma$

C: $\text{flr}^3/\text{Ser} \varnothing \times \text{mh} \sigma$

D: $\text{mh} \varnothing \times \text{flr}^3/\text{Ser} \sigma$

or from crosses where P elements were entirely absent. I therefore suggest that P element activity may be taking place in wing cell somatic tissue as well as in the germline.

5.4.1 Developmental Temperature and Sex of Affected Progeny

I performed two parallel sets of experiments, at developmental temperatures of 25°C and 29°C respectively. In the experiments conducted at 25°C I found no significant increase in the frequency of the mutant clones in the wings of dysgenic flies. Since this confirms the findings of previous researchers (Thompson *et al.* 1978; McElwain 1986), it appears that if somatic movement is to take place it must be temperature induced.

At a developmental temperature of 29°C, although the frequency of mutant spots increased dramatically in all data sets, the significantly higher frequency of spots in the female progeny of the dysgenic cross indicates the possible somatic activity of P elements. A great body of research has demonstrated the importance of developmental temperature and the functioning of P elements. SD sterility is highly temperature sensitive. For example Kidwell *et al.* (1977b) found certain F₁ dysgenic females to be completely fertile when raised at 18°C and completely sterile when raised at 28°C.

Although male recombination was found to be most frequent at 25°C by some authors[†] (Kidwell *et al.* 1977b; Kidwell and Novy

[†] Kidwell *et al.* (1977b) cautioned however that their results should be interpreted with care since dessication had occurred in some of their tests.

1979; Engels 1979c), other researchers (Yannopoulos and Pelecanos 1977) showed that male recombination was enhanced at elevated temperatures of 29°C. Robertson *et al.* (1988) found, using $\Delta 2-3$ (the modified P element lacking the third intron) transformed stocks, that germline hypermutability of *on^W* increased with increasing temperature, exceeding 90% at 28°C.

As well as the abovementioned germline events, evidence of the temperature dependence of non-germline and (possibly) P induced events has also been obtained. Hihara (1987) found that developmental retardation and the degree of lethality of hybrid females increased at elevated developmental temperatures. Colgan (1982) found that male mating competitiveness was significantly reduced in dysgenic hybrids reared at 29°C. Finally, Engels *et al.* (1987) found that non-autonomous P elements caused lethality when mobilized in somatic cells by $\Delta 2-3$ and that the lethality showed a strong dependence on developmental temperature (the frequency increased at elevated temperatures). The authors postulated that this temperature sensitivity was due to the less efficient repair of P element induced chromosome breaks at elevated temperatures or that the production and/or activity of transposase may be temperature dependent.

Another possibility is that RNA splicing may be a temperature sensitive process, in that the P element primary transcript has the third intron differentially spliced at higher temperatures. However, if anything, the opposite could be the case, since work on heat shock proteins has shown that the splicing of mRNA precursors is blocked in *Drosophila* cells at temperatures of 37°C-39°C (Yost and Lindquist 1986).

Only the female progeny of the dysgenic cross showed an increase in the frequency of mutant spots. I do not know why this increase was restricted to the females. However, evidence that P element activity may affect females more than males was obtained by Hihara (1987) who found that developmental retardation occurred only in hybrid females (arising from dysgenic crosses). Also the incidence of GD sterility tends to be slightly greater in females than in males (see, for example, Kidwell and Kidwell 1975a).

5.4.2 Frequency of Spots

Male recombination is thought to be, if not initiated, then, at least accelerated by the premeiotic mitosis in the germline cells. Analogously if the induction of wing hair spots is due to the movement of P elements then it is likely that one of the triggering mechanisms is mitosis. However, the frequency of mutant cell clones per total number of cells in the dysgenic female hybrid was 4×10^{-5} (1,03/24 400), where 24 400 is the number of cells in the distal portion of the wing (Garcia-Bellido and Merriam 1971). On the other hand the frequencies obtained in male recombination experiments have been found to be to the order of 10^{-2} to 10^{-4} [†] (Hiraizumi 1971; Hiraizumi *et al.* 1973; Kidwell and Kidwell 1975b; Yannopoulos and Pelecanos 1977; Woodruff and Thompson 1977; Sinclair and Grigliatti 1985).

[†]The great variation observed in these male recombination frequencies is due to measurements being made on different chromosomes and also due to different methods of evaluating the recombination frequency i.e. either as the total number of recombinants divided by total number of progeny or as estimated number of recombination events divided by total number of progeny (Kidwell and Kidwell 1975b). The latter method of measurement is closer to the method of measuring the frequency of wing hair spots in that, a clone with more than one cell corresponds to a cluster of identical phenotypic progeny.

One reason for the discrepancy in the frequency of mutational events may be due to early cell death of affected cells in somatic tissue. Simpson and Morata (1981) found that slowly dividing cells of *Minute* mutations were progressively eliminated from heterozygous mosaic wing compartments due to cell competition which is the result of a local interaction between slow and fast growing cells (the *Minute* cells were invisible only when they were growing next to non-*Minute* cells). The authors point out that other mutations may be subject to a similar effect. Thus if P element mobilization takes place in certain wing cells then it is possible that the mutational events would bring about early cell death to these cells due to their physical interaction with juxtaposed healthy cells. In particular, since *fl²* is homozygous lethal, clones of these cells might not be viable. This situation may not occur in germline cells which tend to be more particulate in nature.

Another reason for the discrepancy could be due to the physical dissimilarity in the distribution of "hotspots". These are regions on the chromosome which are highly sensitive to P element activity. The existence of the P sensitive "hotspots" is evidenced by facts such as the dissimilarity in distribution of male recombination events when compared to that of female recombination events (Hiraizumi *et al.* 1973; Voelker 1974; Woodruff and Thompson 1977). Also most breakpoints of dysgenesis induced chromosomal rearrangements occur at a few highly specific locations and are non-random (Berg *et al.* 1980; Engels 1981a; Engels and Preston 1981).

It is possible that the *mtb* region on chromosome three is not sensitive to P element activity. In this regard Kidwell and Kidwell (1975b) noticed a particularly low frequency of male recombination in the left arm of chromosome three although female recombination was normal there.

Also the cell cycle period and the age of the cell lineage may also directly or indirectly affect the mobility of P elements. In the wings the cell cycle is estimated to be approximately eight and a half hours (Garcia-Bellido and Merriam 1971) whereas the stem cells in the germarium divide every 24 hours to give rise to a germarial cystoblast and a stem cell (King 1970). The stem cells in the testis have a cell cycle time of approximately ten hours (Lindsley and Tokuyasu 1980). Further the last division in the wing cells occurs at approximately P F + 21 hours (Garcia-Bellido and Merriam 1971) whereas in the male testis stem cell proliferation continues well into adult life (Lindsley and Tokuyasu 1980) and so spermatagonia with functional P elements may have been derived from cell lineages much older than those producing the final cells in the wing.

5.4.3 Conclusion

The results indicate that P element activity may be taking place in somatic tissue. This is apparently in contradiction with results obtained by Laski *et al* (1986) who found no evidence of third intron spliced mRNA in somatic tissue. There is however the possibility that this mRNA in somatic tissue is produced in such small quantities that it was not detected just as the spliced mRNA in germline tissue was not detected.

It is possible that the splicing of the third intron in *D. melanogaster* may be a relatively inefficient process which depends on factors such as the cell environment and external temperature. This would explain why the rates of germline $\Delta 2$ mutability observed for the artificially spliced $\Delta 2-3$ P element were found by Robertson *et al.* (1988) to be considerably higher than those caused by strong P strains such as π_2 (Engels 1984). Further, it would explain the exceptionally high level of transposase activity produced by $\Delta 2-3$ when measured in terms of somatic mosaicism (Laski *et al.* 1986; Robertson *et al.* 1988).

CHAPTER SIX

6. Allozymes

6.1 Introduction

Because of its commensal nature, *Drosophila melanogaster* was originally considered to be of not much value for ecological genetic studies. It was felt that because of the transport of this species with man, there would be very little genetic differentiation worldwide. In fact, the opposite proved to be true. Field studies over the past two decades have revealed a large amount of genetic variation amongst natural populations of *D. melanogaster*. This study is concerned with the genetic differences among populations of *D. melanogaster* in the southern part of Africa.

6.1.1. Selection and Neutralism

"There is one point connected with individual differences which is extremely perplexing: I refer to those genera which have been called 'protean' or 'polymorphic' in which the species present an inordinate amount of variation...These facts are very perplexing, for they seem to show that this kind of variability is independent of the conditions of life. I am inclined to suspect that we see at least in some of these polymorphic genera, variations which are of no service or disservice to the species, and which consequently have not been seized on and rendered definite by natural selection, as hereafter to be explained" - C Darwin (1859).

Although Darwin's concerns were aimed at the morphological rather than the molecular level, he was obviously puzzled by variation that was neutral, in the sense that it was not acted upon by some selective force. Twenty years of scientific investigation has failed to settle the "neutralist-selectionist" controversy. The neutral-mutation-random-drift theory first formally introduced by Kimura (1968) and subsequently elaborated on by King and Jukes (1969), postulates that most allozyme polymorphisms result from selectively neutral alleles being maintained in a balance between mutation and random genetic drift. This does not rule out natural selection, but the neutral-theory differs from neo-Darwinism in both a quantitative and conceptual sense. The orthodox view held that the momentum and direction of evolution is maintained primarily by positive natural selection which operates at all levels including the molecular level. Mutation and stochastic processes play only a minor role. It is now generally accepted that the amount of genetic variation (the "stuff" of evolution) between subpopulations and within them is determined by the combination of the micro-evolutionary forces of selection, migration, drift and mutation, their relative strengths and the period of time these forces have been acting. An analysis of the patterns and magnitudes of allozyme variation within and between subpopulations must help to reveal the relative importance

of these different processes. In an effort to determine the micro-evolutionary forces which are active in Southern African populations of *D. melanogaster*, I have concerned myself in this study with investigating patterns of gene frequency in relation to gene flow and spatial structure.

6.1.1.1 Gene flow

If a group of populations were completely isolated, genetic drift would tend to fix different selectively neutral alleles in different populations thus causing large amounts of geographic heterogeneity in allozyme frequencies to occur. Gene flow is the cohesive force which can prevent this from happening. The measurement and effectiveness of gene flow in the face of selection and genetic drift is one of the most controversial and problematic areas in population biology.

6.1.1.1.1 Effectiveness of gene flow

Early mathematical formulae for balancing migration against drift and migration against selection were obtained by Wright (1931) and Haldane (1930) respectively. Wright's island model essentially showed that, if a fraction m of the N individuals of an island are replaced by migrants each generation from a source population with fixed gene frequencies, then the condition, $Nm > 1$, will ensure non-divergence of the island population from that of the source. Haldane showed that if, in the same situation, natural selection is operating with strength s , then the condition, $m/s > 1$, will ensure

that immigration will overcome the effect of natural selection. These two simple inequalities state that the forces of natural selection and drift can be easily counterbalanced by the opposing force of migration.

The concept of migration as a potent force has been echoed by authors such as Mayr (1954, 1963) and Stanley (1979) who claim that gene flow is commonplace and effective in uniting a species. On the other hand Ehrlich and Raven (1969) hold that gene flow is a relatively unimportant and ineffectual evolutionary force. Also Endler (1973) produced experimental evidence suggesting that gene flow is far from being a powerful force in the maintenance of a balanced polymorphism between populations. (See also Mayr (1970), Eldredge and Gould (1972) and Endler (1977) for other ideas regarding the primary source of species stability.)

6.1.1.1.2 Measurement of gene flow

Gene flow can be brought about in insects by directional migration, active dispersal⁺ or passive transport (Dobzhansky 1973). There is no evidence of such migration⁺⁺ in any drosophilids. However, active diffusion-like dispersal and passive transport

⁺ "Active dispersal involves movements usually uncorrelated in direction in different individuals of a given insect species, between places where these individuals are born and where they find shelter, food, water and oviposition sites". (Dobzhansky 1973).

⁺⁺ Migration here is used as defined by Dobzhansky (1973) as the movement "...of many individuals in more or less the same direction".

by wind or animals occurs frequently in species of *Drosophila*. Several attempts, such as the studies by Dobzhansky and Wright (1943, 1947) on *D. pseudoobscura*, were made to estimate the amount of gene flow in natural populations by measuring dispersal distances of individuals.

Although these early works suggested that migration was limited to a few kilometres, later results of Jones *et al.* (1981) for *D. pseudoobscura* and of Coyne *et al.* (1982) for *D. pseudoobscura*, *D. simulans* and *D. melanogaster* suggested that individual flies were capable of travelling large distances under both favourable and unfavourable conditions. Jones *et al.* (1981) found that flies could move at least ten kilometres within a twenty four hour period over a desert and that isolated populations separated by fifteen kilometres from other populations received large numbers of immigrants.

Indirect methods of measuring gene flow make use of allele frequency differences between subpopulations to estimate levels of gene flow. One indirect method uses Wright's (1943) fixation index, F_{ST} , statistic which is the ratio of the observed variance in allele frequency to the maximum theoretical variance for populations fixed for different alleles, and makes use of the island model of migration. For small migration rates the fixation index is approximately

$$F_{ST} \approx 1/(4Nm + 1)$$

where N is the population size and m is the proportion of exchanges among subpopulations each generation. This model is especially convenient to use because no estimates of population size are required. Nm is simply the number of migrant individuals. Another indirect method of estimating gene flow from allele frequency data was proposed by Slatkin (1981) who derived a graphical method based on the conditional average frequency of an allele, that is the mean frequency of alleles averaged over those populations in which they appeared. Slatkin (1985) extended this method and showed that the log of the average frequency of private alleles varied in an approximately linear fashion with the log of the number of immigrants Nm . This relationship could then be used to estimate gene flow. Since direct methods measure ongoing rates of gene flow and since indirect methods measure the effects of past population events, these two approaches may yield different results (Slatkin 1987).

The goal of this allozyme study as outlined in Section 1.2.1, was to examine the genetic structure of *D. melanogaster* populations in Southern Africa using electrophoretically detectable population markers. I examined geographic variation for eight polymorphic enzymes to measure the degree of differentiation at several hierarchical levels of population structure and used these results indirectly to estimate the amount of gene flow between populations.

6.2 Materials and Methods

6.2.1 Geographic Distribution of Samples

Flies were collected and isofemale lines established as described in Section 2.2.1. For this electrophoretic study certain collections were grouped together and the resultant sample distribution is shown in Fig. 6.1 and Table 6.1.

Collections were established from a number of diverse geographical areas. Sample 1, is from the edge of the Namib desert on the Atlantic coast, whereas Sample 2 is 50 km inland right inside the Namib desert. Sample 3 was made in the heart of the Fish River Canyon Park. Samples 4, 5, 6 and 7 were taken in the southern coastal region of South Africa; Sample 8 was taken in the Cape mountains and Samples 9 and 10 were collected in a semitropical region of the east coast of South Africa. The remaining samples, 11-17, were collected at locations in the interior highlands of Southern Africa. The collection from Zimbabwe (Sample 17) is the most northerly in origin (17,83°S). Samples 6 and 7 originate from the southernmost collection site, Port Elizabeth (33,96°S).

6.2.2 Electrophoretic Analysis

Horizontal starch gel electrophoretic methods followed May *et al.* (1979). Two different buffer systems were used to resolve *Drosophila* enzymes on starch gels, viz:

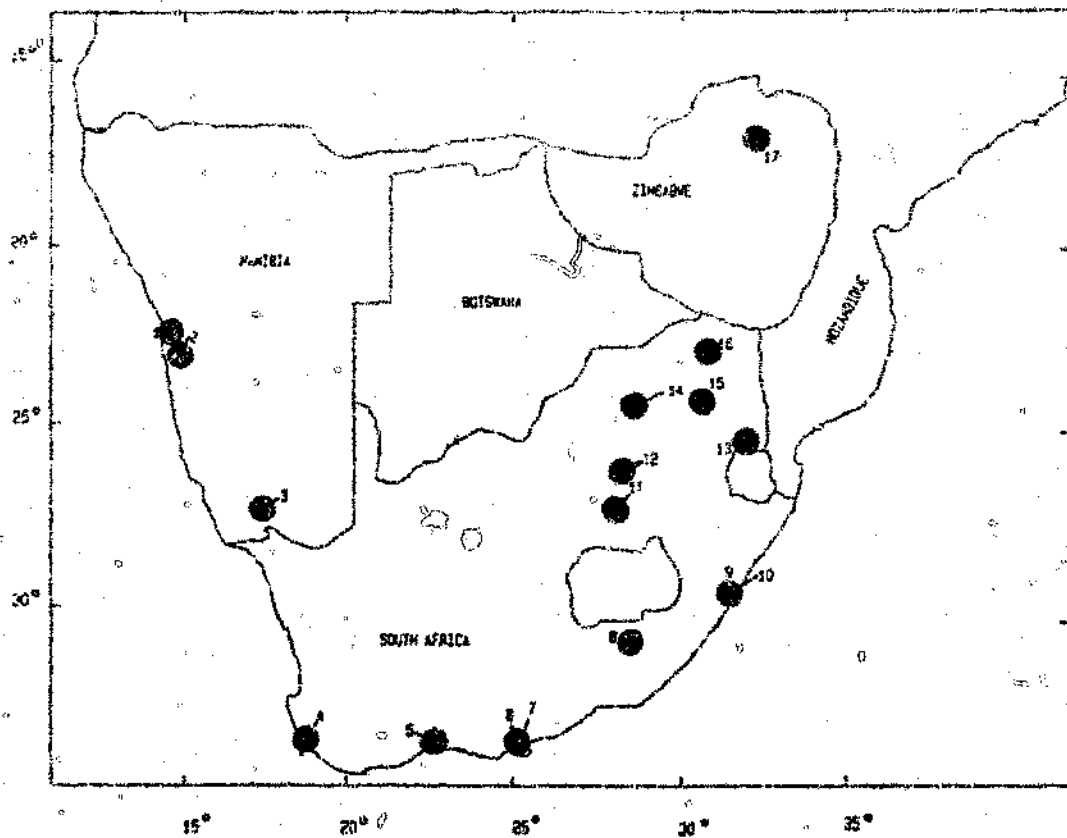


Fig. 6.1

Map of Southern Africa showing locations of *D. melanogaster* collection sites.

Colours indicate the different regions as used in the statistical analysis (Fig. 6.2).

- Namibia
- Southwestern Cape
- Southeastern Cape
- Eastern Cape/Katal
- Transvaal
- Zimbabwe

Table 6.1

Locations of samples, collection dates and numbers of isofemale lines of *D. melanogaster* for each location.

Location ¹	Location number ²	Date	Number of isofemale lines
Namibia			
Northern area			
Swakopmund	1	Oct. 1988	15
Gobabeb	2	Oct. 1988	25
Southern area			
Ai Ais	3	Oct. 1988	24
South Africa			
Southwestern Cape			
Cape Town	4	Oct. 1988	21
Southeastern Cape			
Plettenberg Bay	5	Dec. 1987	4
Port Elizabeth			
Summerstrand	6	Oct. 1988	27
Walmer	7	Oct./Nov. 1988	8
Eastern Cape/Natal			
Maclear	8	Dec. 1987	4
Durban			
Berea	9	Dec. 1988	27
Addington Beach	10	Oct. 1988	22
Transvaal			
Southern Transvaal			
Johannesburg	11	Dec. 1987/1988	33
Pretoria/Hammanskraal	12	1987	8
Eastern Transvaal			
Nelspruit	13	Nov. 1988	25
Northern Transvaal			
Pietersburg	14	Dec. 1987	4
Agatha	15	Dec. 1987	4
Northeastern Transvaal			
Thohoyandou, Venda	16	Oct. 1988	25
Zimbabwe			
Harare	17	Oct./Nov. 1988	25

¹Nested geographic subdivision used in the analysis of allele frequencies

²Location numbers in Fig. 6.1

(i) MF (Markert and Faulhaber (1965))

electrode solution : 0,18 M Tris
 (pH 8,2) 0,1 M boric acid
 0,004 M Na EDTA
 gel solution : 1:4 dilution of electrode buffer

(ii) IC (Whitt (1970))

electrode solution : 0,15 M Tris
 (pH 6,9) 0,05 M citric acid
 gel solution : 1:14 dilution of electrode buffer.

6.2.2.1 Sample Preparation

Twenty five (25) adult flies from each isofemale line were homogenized by hand using a glass rod in 0,5 ml of extraction buffer (0,1 M Tris-HCl pH 8,0) and the homogenate centrifuged at 10 000 g for 5 minutes. The clear supernatant was absorbed into filter paper wicks and inserted along slots made 3 cm into the gel. Electrophoresis was either performed immediately or, if that was not possible, homogenized samples were stored at -25°C for 24 hours or less before being analysed. (Samples stored for longer than that did not give good resolution and so, were not used).

6.2.2.2 Gel preparation, loading and running

Twelve per cent (12%) w/v starch gels were used in this study and were individually prepared with 42 g of potato starch in 350 ml of the appropriate buffer. All gels measured 22 cm x 18 cm x 1 cm. Gels were placed on 100° packs while

(i) MF (Markert and Faulhaber (1965))

electrode solution : 0,18 M Tris
 (pH 8,7) 0,1 M boric acid
 0,004 M Na EDTA
 gel solution : 1:4 dilution of electrode buffer

(ii) IC (Whitt (1970))

electrode solution : 0,15 M Tris
 (pH 6,9) 0,05 M citric acid
 gel solution : 1:14 dilution of electrode buffer.

6.2.2.1 Sample Preparation

Twenty five (25) adult flies from each isofemale line were homogenized by hand using a glass rod in 0,5 ml of extraction buffer (0,1 M Tris-HCl pH 8,0) and the homogenate centrifuged at 10 000 g for 5 minutes. The clear supernatant was absorbed into filter paper wicks and inserted along slots made 3 cm into the gel. Electrophoresis was either performed immediately or, if that was not possible, homogenized samples were stored at -25°C for 24 hours or less before being analysed. (Samples stored for longer than that did not give good resolution and so, were not used).

6.2.2.2 Gel preparation, loading and running

Twelve per cent (12%) w/v starch gels were used in this study and were individually prepared with 42 g of potato starch in 350 ml of the appropriate buffer. All gels measured 22 cm x 15 cm x 1 cm. Gels were placed on ice packs while

being loaded. Approximately 40 samples per gel were loaded. Three samples of Canton-S were run on each gel to be used as a standard marker strain. Overnight electrophoresis took place in a cold room at 4°C.

6.2.2.3 Cutting and staining of gels

After electrophoresis a gel was sliced into four, 1.5 mm slabs. Specific enzyme stains (Harris and Hopkinson 1976) were applied to each slab of the gel using an agar overlay, and the slabs incubated at 37°C in the dark. Seven enzyme systems were examined in all of the samples (Table 6.2).

6.2.2.4. Genetic interpretation of banding patterns

I was unable to obtain results from single flies and was forced to use a homogenate composed of several individual flies from the same isofemale line. Other workers who have used multiple-fly homogenates from isofemale lines for electrophoresis (e.g. Singh and Coulthart (1982)) interpreted their banding patterns as reflecting only two genomes. I, however, have concluded that for an autosomal locus the progeny from a single inseminated female used to start an isofemale line represent at least four genomes, two from each parent, and three genomes for a sex-linked locus (see O'Brien and MacIntyre 1969). Since multiple inseminations do take place (Milkman and Zeitler 1974), this interpretation may not be strictly accurate. I, however, detected no evidence of multiple paternity in the banding patterns.

Table 5.2

Enzymes electrophoretically examined, composition of staining solutions and buffer systems used.

The enzyme staining methods were adapted from Shaw and Prasad (1970) and Harris and Hopkinson (1976)

Enzyme	EC Number	Stain Composition	Buffer System*
Alcohol dehydrogenase (Adh)	1.1.1.1	10 ml Tris-8, 0.1 M 0.5 ml isopropanol 20 mg NAD 20 mg MTT 5 mg PMS 10 ml 2% Agar	MF
Dipeptidase-A (Dzp-A)	3.4.11	10 ml Tris-8, 0.5 M 20 mg leucyl-tyrosine (LT) 10 mg snake venom 10 mg peroxidase 30 mg O-dianisidine 10 ml 2% agar	MF
Glycerol-3-phosphate dehydrogenase (G3P) or (Gpdh)	1.1.1.8	10 ml Tris-8, 0.1 M 30 mg glycerol-3-phosphate 20 mg NAD 20 mg MTT 5 mg PMS 10 ml 2% agar	TC
Hexokinase (Hex 1,2)	2.7.1.1	10 ml Tris-8, 0.1 M 30 mg glucose 10 mg ATP 10 mg NAD 20 units glucose-6-P dehydrogenase 20 mg MTT 5 mg PMS 10 ml 2% agar	TC
Isocitrate dehydrogenase-NADP (Idh-NADP)	1.1.1.42	10 ml Tris-8, 0.1 M 30 mg DL-isocitrate 10 mg NADP 20 mg MTT 5 mg PMS 10 ml 2% agar	TC
6-phosphogluconate dehydrogenase (Pgd)	1.1.1.44	10 ml Tris-8, 0.1 M 30 mg 6-phosphogluconate 10 mg NADP 20 mg MTT 5 mg PMS 10 ml 2% agar	MF
Phosphoglucomutase (Pgm)	2.7.5.1	10 ml Tris-8, 0.1 M 20 mg glucose-1-phosphate with trace amounts of G-1,6-P 10 mg NAD 20 units glucose-6-P dehydrogenase 20 mg MTT 5 mg PMS 10 ml 2% agar	TC

*Buffer systems MF and TC as described in Section 6.2.2

Since most of the lines were electrophoresed within one month of establishment of the isofemale line, there was probably little allelic loss due to random drift in the cultures and all parental alleles were taken to be present. The correctness of this representation was validated by the following two observations.

(i) Asymmetric banding patterns representing a 3:1, 2:1:1 or 2:1 allelic ratio were generally found to be present.

(ii) Lines segregating for three or four alleles were not uncommon. Since it was not possible to distinguish between heterozygous and superimposed variant homozygous phenotypes at the same loci, I could not test the samples for fit to Hardy-Weinberg proportions.

6.2.2.5 Isozyme/allozyme nomenclature and homology

Loci encoding functionally similar isozymes appearing on the same gel were designated numerically beginning at the cathodal end of the gel. Synonomies with other nomenclatures using letter designations were also indicated. Alleles were designated by their electrophoretic mobilities relative to the mobility of the most common allele which was designated 100. Allelic variants migrating cathodally from the origin were prefixed with a minus sign.

Homology with allozymes/isozymes in the literature was sometimes obvious, but at other times it was difficult if not impossible to establish. Since the object of this study was to examine

genetic diversity within and between geographically separated populations this failure at times to identify homologous alleles was not of paramount importance.

6.2.3 Statistical Analysis

6.2.3.1 Hierarchical population model

Allele frequencies were organized for statistical purposes according to a hierarchical population model (Fig. 6.2). At the lowest level, comparisons were made between samples collected at different sites within the same town or district. Districts, the next level up, were then nested together within areas and finally areas clustered into regions, of which there were six namely Namibia, Southwestern Cape, Southeastern Cape, Eastern Cape and Natal, Transvaal and Zimbabwe. As far as possible, an effort was made to collect flies with this hierarchical structure in mind. However, very often practical constraints such as vast distances to be travelled, problems in attracting and collecting a sufficient number of flies in a limited period of time and finally transporting the flies back to Johannesburg, prohibited the establishment of a balanced statistical design.

REGION	AREA	DISTRICT	COLLECTION SITE
I NAMIBIA	Northern Namibia	Swakopmund Gobabeb	Swakopmund Gobabeb
	Southern Namibia	Ai Ais	Ai Ais
II SOUTH WESTERN CAPE		Cape Town	Cape Town
III SOUTH EASTERN CAPE		Plettenberg Bay Port Elizabeth	Plettenberg Bay Summerstrand Walmer
IV EASTERN CAPE/ NATAL		Maclear Durban	Maclear Addington Beach Berea
V TRANSVAAL	Southern Transvaal	Johannesburg Pretoria/ Hammanskraal	Johannesburg Pretoria Hammanskraal
	Eastern Transvaal	Nelspruit	Nelspruit
	Northern Transvaal	Pietersburg Agatha	Pietersburg Agatha
	North Eastern Transvaal	Thohoyandou	Thohoyandou
VI ZIMBABWE		Harare	Harare

Fig. 5.2 Hierarchical arrangement of Blasphilia collection sites

6.2.3.2 Allele frequency differences

The null hypothesis, H_0 , that all the samples were drawn from the same panmictic population and so would show no significant difference in allele frequencies was tested by using the likelihood ratio test statistic (G statistic) for multinomial data (Sokal and Rohlf 1969). For each comparison made (see below) G statistics and their corresponding degrees of freedom were calculated separately at each locus using the following computations:

Given n samples and m alleles let

f_{ij} = observed number of " j " alleles in the " i th" sample
($j = 1, \dots, m; i = 1, \dots, n$).

r_i = total number of alleles in the " i th" sample.

c_j = total number of " j " alleles.

N = total number of alleles in the contingency table.

$$\text{Then } G = 2 \left[\sum_{i=1}^n \sum_{j=1}^m f_{ij} (\ln f_{ij}) - \sum_{i=1}^n r_i (\ln r_i) - \sum_{j=1}^m c_j (\ln c_j) + N (\ln N) \right]$$

is an asymptotically distributed chi-square statistic (under H_0) with $(n-1)(m-1)$ degrees of freedom.

Allele frequency differences between sites within districts, between districts within areas and between areas within regions were tested in this way.

An advantage of the likelihood ratio-test over the chi-square test is that, in a chi-square test expected values in a contingency table should not be less than five. In practice, this restriction can often cause difficulties. The likelihood ratio test is thus more practical since an occasional value as low as one may be used. With loci having more than two alleles, all of the rarer alleles were pooled so that no sample had allele frequencies of zero. Some loci were fixed at certain locations. If this did not occur in too many samples then a value of one was substituted for zero in the second allelic class of the analysis of the G statistic (Smouse 1974), to facilitate computation. This procedure can be justified by assuming that the zeros resulted from finite sampling. However, with two loci, Pgd and Pgm , the number of locations with fixed loci was too great to justify such an alteration and so these two loci had to be excluded from the statistical analysis. Thus, altogether six polymorphic loci were analysed. For each comparison made, G statistics were then summed over loci and divided by the total degrees of freedom, assuming linkage equilibrium and statistical independence, to yield a summary statistic for that hierarchical level. The significance level of each test was modified to account for the increase in type I error (rejecting the null hypothesis when it is true) when multiple tests are performed on the same sample (Cooper 1968).

Thus if the required overall level of significance is 5% and if the number of simultaneous tests (which is equivalent to the number of polymorphic loci) is n then, the modified significance level for individual tests, α , must satisfy

$$(1-\alpha)^n = 0.95$$

$$\text{whence } 1-n\alpha = 0.95 \quad (\text{expanding and neglecting terms of order } \geq 2)$$

$$\text{i.e. } \alpha = 0.05/n$$

Thus in this study using six polymorphic loci the allele frequency differences for a particular comparison were considered significant at the 0.05 (0.01, 0.001) level only if the total (summed over loci) G -statistic was significant at the 0.05 (0.01, 0.001) level and if at least one locus was also significant at the 0.05/6 (0.01/6, 0.001/6) level.

6.2.3.3 Gene diversity analysis

The same model of hierarchical population structure was used to estimate gene diversities (Fig. 6.3) (Nei 1973, 1975; Chakraborty 1980; Chakraborty *et al.* 1982). The total estimated heterozygosity for each locus was first partitioned into within regions and between regions components such that

$$H_T = H_R + D_{RT} \text{ where}$$

H_T is the total Hardy Weinberg expected heterozygosity for all sampled populations assuming panmixia.

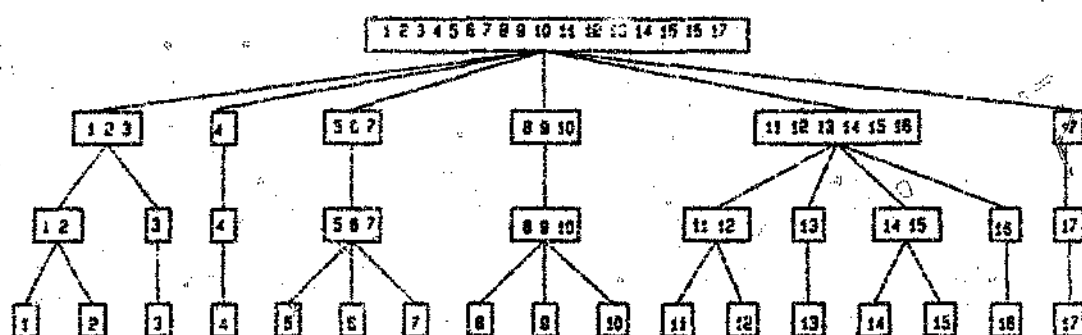


Fig. 6.3

**Hierarchical grouping as used in the
Gang Diversity Analysis**

**Numbers refer to location numbers as indicated
in Table 6.1**

H_R is the average gene diversity within regions i.e. the average over regions of H_{Rk} where H_{Rk} is the Hardy-Weinberg expected heterozygosity of region k.

D_{RT} is the absolute gene diversity (or absolute decrease in average heterozygosity) resulting from population subdivision at the between regions level.

Continuing in this way, the within regions heterozygosity (H_R) was partitioned into a within areas heterozygosity component (H_A) and a between areas diversity component (D_{AR}). Further, the within areas heterozygosity (H_A) was partitioned into within districts (H_D) and between districts (D_{DA}) components. Finally H_D was partitioned into the two components between districts gene diversity (D_{DS}) and within collection sites heterozygosity (H_S).

$$\begin{aligned}
 \text{So } H_T &= H_R + D_{RT} \\
 &= H_A + D_{AR} + D_{RT} \\
 &= H_D + D_{DA} + D_{AR} + D_{RT} \\
 &= H_S + D_{SD} + D_{DA} + D_{AR} + D_{RT}
 \end{aligned}$$

The relative gene diversity component (or proportional decrease in heterozygosity) due to variation between subpopulations, G_{XY} was obtained by dividing the absolute gene diversity component between subpopulations D_{XY} by the total heterozygosity H_T .

Thus $G_{RT} = D_{RT}/H_T$, $G_{AR} = D_{AR}/H_T$, $G_{DA} = D_{DA}/H_T$ and $G_{SD} = D_{SD}/H_T$

The heterozygosities, absolute gene diversities and relative gene diversities were initially calculated for each of the loci separately. The average relative gene diversities over all eight polymorphic loci was calculated by evaluating the ratios of average D_{XY} ($\bar{D}_{XY}$) to average total heterozygosity.

$$\text{Thus } \bar{D}_{XY} = \bar{D}_{XY}/H_T$$

The component of gene diversity due to variation between populations at collection sites within the total population, D_{ST} , is given by $D_{ST} = H_T - H_S$. Thus, the relative gene diversity, G_{ST} , is:

$$G_{ST} = D_{ST}/H_T = (H_T - H_S)/H_T = 1 - H_S/H_T$$

which is the same as Wright's fixation index F_{ST} when dealing with a locus with two alleles. For comparative purposes G_{ST} and F_{ST} can essentially be regarded as the same statistic.

6.3. Results

6.3.1. Biochemical Genetic Variation

Descriptions of the enzymes examined and their electrophoretic conditions follow.

Fig. 6.4 shows banding patterns obtained for all enzyme systems as well as photographs of the gels of *Adh*, *Gpdh* and *Idh*-NADP. The alleles for each locus as well as their composite relative frequencies, structures and genetic map positions are given in Table 6.3.

6.3.1.1 Alcohol dehydrogenase (*Adh*)

The activities of three alleles were observed for this dimeric enzyme. The most common allozyme was a cathodally migrating one (*Adh*⁻¹⁰⁰) followed in frequency by an anodally migrating allozyme (*Adh*⁰). In addition a rare cathodally migrating band (*Adh*⁻¹²⁰) was found to be present at Ai Ais, Berea (Durban) and Harare at a composite frequency of 0,007 (Table 6.3). Each band was accompanied by a faint satellite band which is "band 0" of Chambers *et al.* (1978) and is typical of the multibanded pattern of the single *Adh* gene product.

The presence of satellite bands together with the observed cathodal migration for two of these bands and a survey of the literature (Johnson and Denniston 1964; Gill *et al.* 1965; Vigue and Johnson 1973; Day *et al.* 1974; Chambers *et al.* 1978; David *et al.* 1980) suggests that the two common alleles are the same as those presented in the current literature as *Adh*^S (*Adh*⁻¹⁰⁰) and *Adh*^F (*Adh*⁰), i.e. the slow and the fast allele. Further the allele *Adh*⁻¹²⁰ could well be the ultra slow allele (*Adh*^{US}) of *Adh* which was found in the Congo (David 1978; David *et al.* 1980).

Fig. 6.4 shows banding patterns obtained for all enzyme systems as well as photographs of the gels of *Adh*, *Gpdh* and *Idh*-NADP. The alleles for each locus as well as their composite relative frequencies, structures and genetic map positions are given in Table 6.3.

6.3.1.1 Alcohol dehydrogenase (*Adh*)

The activities of three alleles were observed for this dimeric enzyme. The most common allozyme was a cathodally migrating one (*Adh*⁻¹⁰⁰) followed in frequency by an anodally migrating allozyme (*Adh*⁰). In addition a rare cathodally migrating band (*Adh*⁻¹²⁰) was found to be present at Ai Ais, Berea (Durban) and Harare at a composite frequency of 0,007 (Table 6.3). Each band was accompanied by a faint satellite band which is "band 0" of Chambers *et al.* (1978) and is typical of the multibanded pattern of the single *Adh* gene product.

The presence of satellite bands together with the observed cathodal migration for two of these bands and a survey of the literature (Johnson and Denniston 1964; Grell *et al.* 1965; Vigue and Johnson 1973; Day *et al.* 1974; Chambers *et al.* 1978; David *et al.* 1980) suggests that the two common alleles are the same as those presented in the current literature as *Adh*^S (*Adh*⁻¹⁰⁰) and *Adh*^F (*Adh*⁰), i.e. the slow and the fast allele. Further the allele *Adh*⁻¹²⁰ could well be the ultra slow allele (*Adh*^{US}) of *Adh* which was found in the Congo (David 1978; David *et al.* 1980).

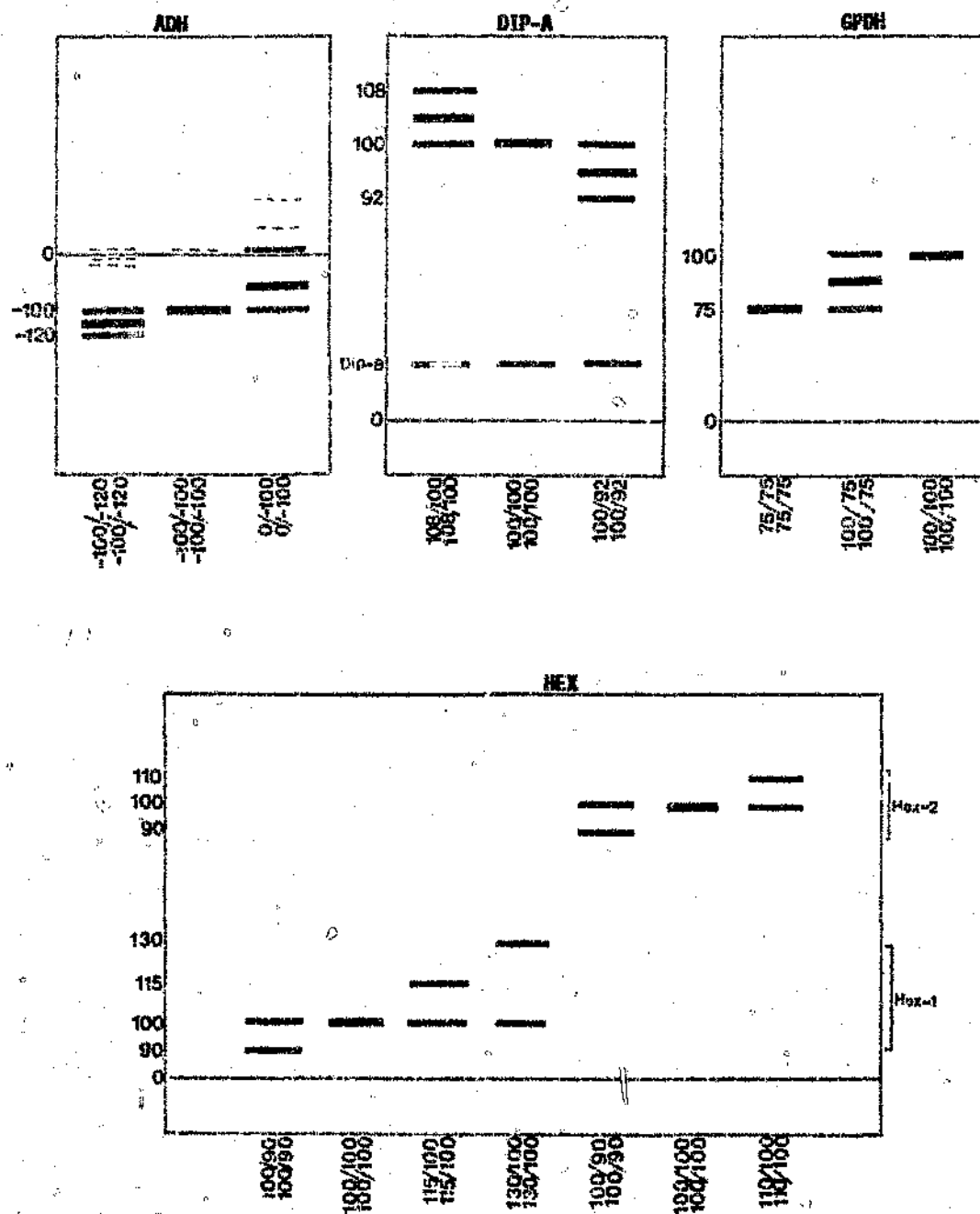


Fig. 6.4

A: Diagrammatic drawings of the gel banding patterns obtained for the alleles of *Adh*, *Dip-A*, *Gpdh*, *Hex-1*, *Hex-2*, *Idh-HADP*, *Pgd* and *Pgm*.

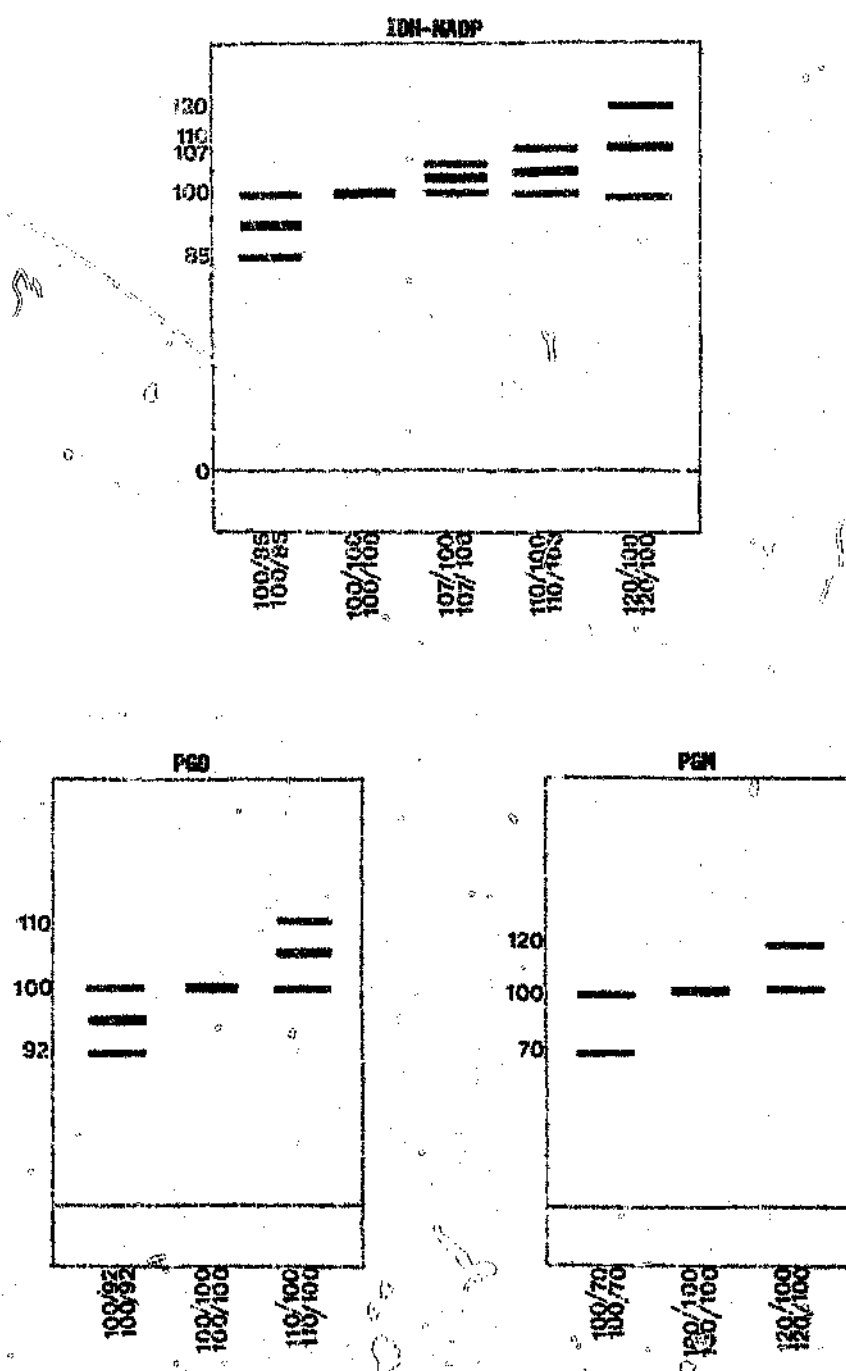


Fig. 4 A (continued)

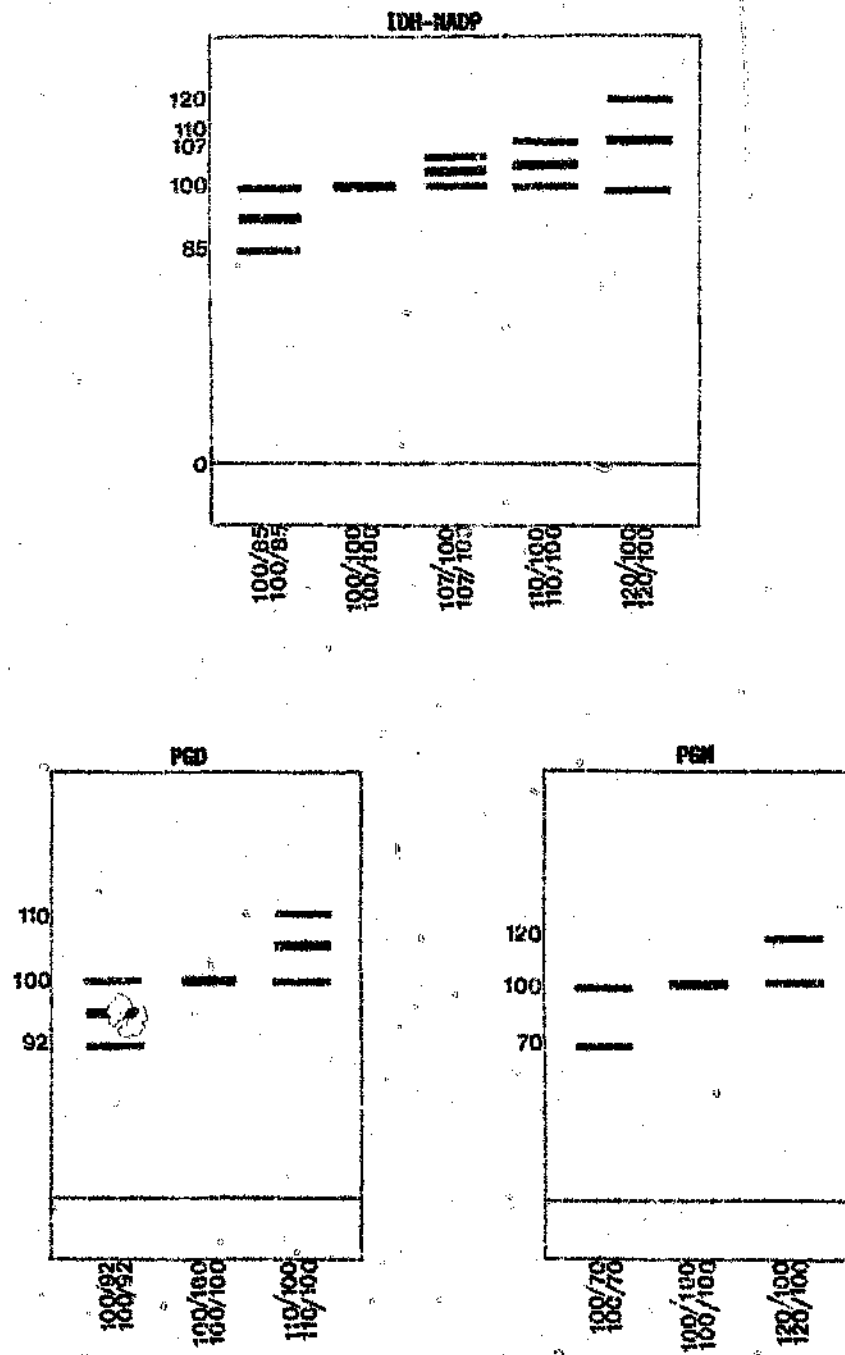


Fig. 6.4 A (continued)

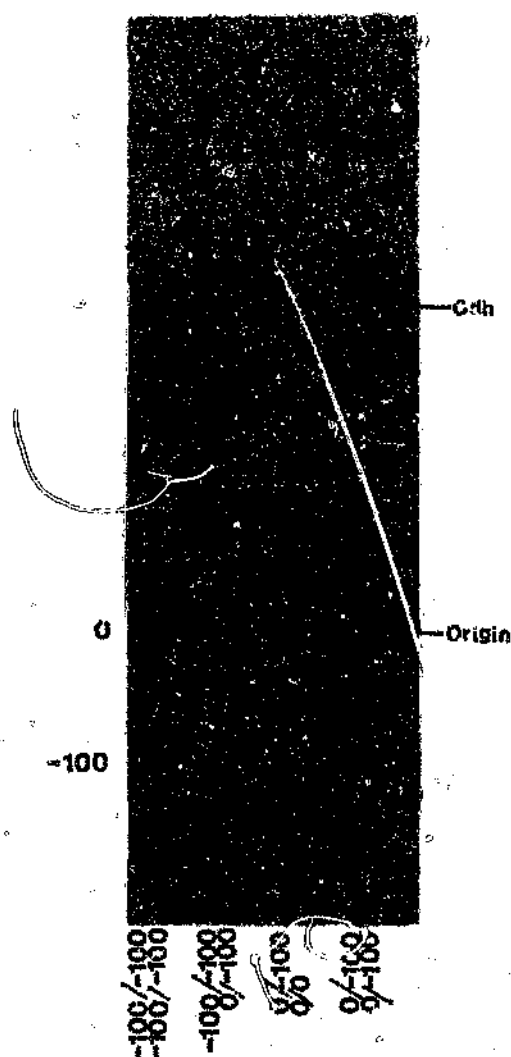


Fig. 6.4

B: (i) Electrophoretic variation of Adh.
The genotypic interpretation is given for each sample.

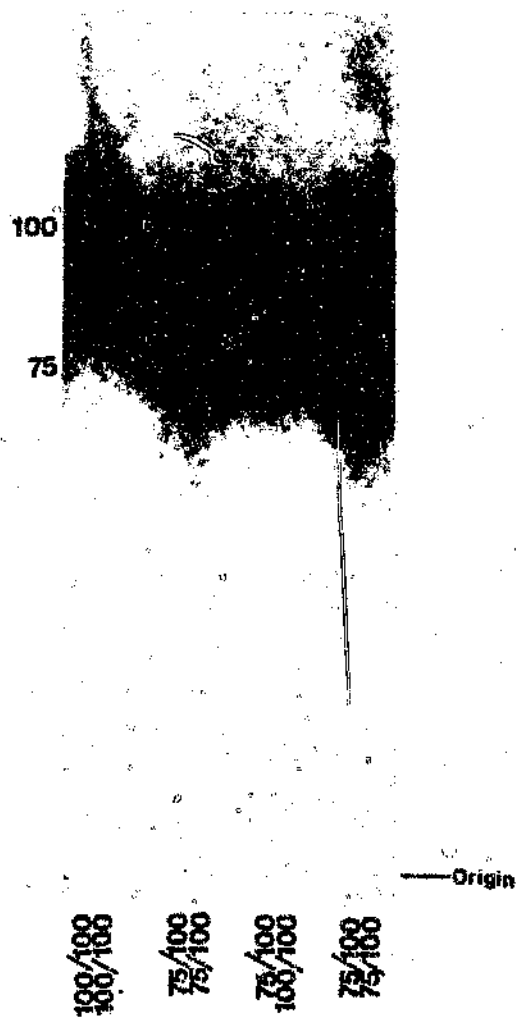


Fig. 6.4

**5: (14) Electrophoretic variation of Gdh.
The genotypic interpretation is given
for each sample.**

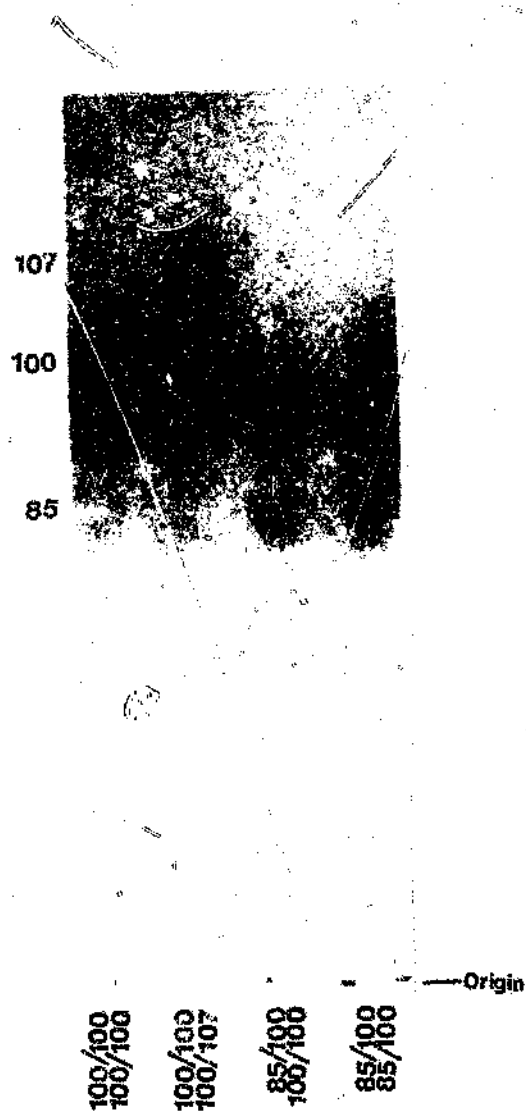


Fig. 6.4

**B: (11%) Electrophoretic variation of Idh-NADP.
The genotypic interpretation is given
for each sample.**

Table 6.3

Alleles observed at all loci, their composite relative frequencies, structures and genetic map positions

Locus	Alleles observed	Composite relative frequency	Structure	Genetic map position
<i>Aah</i>	-100 0 -120	0,678 0,315 0,007	Dimer	2- 50.1
<i>Dip A</i>	100 92 108	0,914 0,054 0,032	Dimer	2- 55.2
<i>Gpdh</i>	100 75	0,906 0,094	Dimer	2- 20.5
<i>Hex-1</i>	100 115 130 90	0,889 0,050 0,033 0,028	Monomer	2- 73.5
<i>Hex-2*</i>	100 90 110	0,948 0,041 0,011	Monomer	not mapped
<i>Idh-NADP</i>	100 85 107 110 120	0,812 0,150 0,035 0,002 0,001	Dimer	3- 27.1
<i>Pgd</i>	100 92 110	0,992 0,007 0,001	Dimer	1- 0.63
<i>Pgm</i>	100 120 70	0,942 0,040 0,018	Monomer	3- 43.4

* Scored as autosomal

5.3.1.2 Dipeptidase-A (*Dip-A*)

When starch gels are stained for dipeptidase activity, the isozyme bands which appear are dependent upon the particular dipeptide substrate used. Leucyl-tyrosine (which was used in this study) is specific for two dipeptidases, *Dip-A* and *Dip-B* (Laurie-Ahlberg 1982), the gene products of *Dip-A* being the more anodal of these two systems (Voelker and Langley 1978). *Dip-B* appeared to be monomorphic, but *Dip-A* was sometimes polymorphic with three variant phenotypes. This is in accordance with reports of Voelker and Langley (1978) and Ohnishi and Voelker (1981). Further the intermediate allele had a composite relative frequency of 0,914 (Table 6.3) which is similar to that which has been reported in other parts of the world (e.g. by Langley *et al.* (1977) in North Carolina and Gonzalez *et al.* (1982) in Tenerife and the Iberian Peninsula), indicating the possible homology of the three Southern African *Dip-A* alleles with those of the northern hemisphere.

6.3.1.3 sn-glycerol-3-phosphate dehydrogenase (*Gpdh*)

The activity of two anodally migrating allozymes was observed for this locus. There was a very common fast allele (*Gpdh*¹⁰⁰) and a slow allele (*Gpdh*⁷⁵) having composite relative frequency 0,094 (Table 6.3). The dimeric nature of this enzyme was indicated by heterozygous individuals which displayed a three-banded phenotype. The more rapidly migrating enzyme was found to be present in the Canton-S strain in two doses. In view of the ubiquity of a fast and a slow allele for *Gpdh*,

it appears that *Gpdh*¹⁰⁰ and *Gpdh*⁷⁵ are homologous with *Gpdh*^F and *Gpdh*^S — the two widespread fast and slow alleles cited in the literature (Grell 1967; Berger 1971; Sing *et al.* 1973; Langley *et al.* 1974; David 1982; Singh *et al.* 1982; Oakeshott *et al.* 1982a; Cabrera *et al.* 1982; Kusakabe and Mukai 1984; Watada *et al.* 1986; Anxolabéhère *et al.* 1988a).

6.3.1.4 Hexokinase (*Hex-1*, *Hex-2*)

Nomenclature

Three separate loci coding for Hexokinase were observed. I have followed the procedures of Ayala *et al.* (1972) and Singh *et al.* (1982) in referring to the enzyme with the least anodal migration as HEX-1, the next fastest as HEX-2 and the most anodal as HEX-3. This nomenclature is identical to that employed by Choudhary and Singh (1987). There is a certain amount of confusion regarding the nomenclature of the hexokinase enzymes and their loci. Treat-Clemons and Doane (1984) refer to the locus *Hex-1* as being the same as glucose oxidase (*Go*) and as FAD glucose dehydrogenase (*Gld*). This locus however codes for a dimeric enzyme which is not a hexokinase but was originally identified as such by Murray and Ball (1967) due to the non-specificity of the hexokinase electrophoretic assay. Cavener (1980) identified a Glucose Oxidase (*GO*) in the ejaculatory ducts of male *D. melanogaster* as being the enzyme previously misidentified as HEX-1. Furthermore, Cavener (1980) located the structural gene for this enzyme on the third chromosome. The isozyme HEX-1 referred to in this study is a monomer. It is located

on chromosome two and has been functionally classified as a fruktokinase (Jeines 1971). This locus (*Hex-1* in the present study) has variously been named *Hex-3* Singh *et al.* 1973); *Hex-C* (Mukai and Voelker 1977; Moser *et al.* 1980) and *Hex-3*, *Fk* and *Hex-C* (Treat-Clemons and Doane 1984). Singh and Rhomberg (1987b) have given *Hex-1* and *Hex-3* (nomenclature as in the present study) the genetic map designations, *Hex-C* and *Hex-A*, respectively. *Hex-2* has not been mapped.

Electrophoretic Description

Hex-3 was monomorphic in all samples.

Hex-1 is autosomal (as discussed above) and was scored as such. This monomeric enzyme exhibited activity for four different alleles, the most common being the second slowest. This number of alleles is higher than that observed in other parts of the world where three or fewer alleles have been reported. For example in North America (Singh and Rhomberg 1987a), in Japan (Kusakabe and Mukai 1984) and elsewhere in the world (Choudhary and Singh 1987). All four of these alleles were geographically widespread with five out of seventeen samples having all four alleles present.

Hex-2 — I was unable to establish the chromosomal location of this enzyme from a search of the literature. *Hex-A* and *Hex-B* (*Hex-3* and *Hex-2* in my notation) are reported by some authorities to have been mapped onto the same structural gene on the X-chromosome (Moser *et al.* (1980), and references therein). On the other hand Choudhary and Singh (1987) and Singh and Rhomberg (1987b) refer to *Hex-A* (*Hex-3* in their notation) as being on chromosome 1, but *Hex-2* (their notation again) as being unmapped.

Hex-2 exhibited activity for three different alleles. The most common allele, *Hex-2*¹⁰⁰ accounted for practically 94% of the composite frequency. Because of the preponderance of this allele, it was possible to score *Hex-2* in two different ways i.e. as an autosomal gene or as a sex-linked gene. The composite relative frequencies of the three alleles *Hex-2*¹⁰⁰, *Hex-2*⁹⁰ and *Hex-2*¹¹⁰ scored as sex-linked are 0,941, 0,045 and 0,014 respectively; whereas scored as autosomal they are 0,948, 0,041 and 0,011 respectively. Thus there is very little difference in the composite frequencies due to different scoring methods. Consequently I have assumed that *Hex-2* is an autosomal locus in the statistical analysis. In Appendix D I have recalculated the relevant statistics assuming that *Hex-2* is sex-linked, but the differences are negligible.

No collection site had all eight loci polymorphic on the 95% criterion for polymorphism. However Nelspruit had seven and Pretoria/Hammanskraal had six polymorphic loci. These collection sites are both in the Transvaal and have moderate climates with no extremes of temperature.

Pretoria and Hammanskraal have a great deal of human traffic, and so a high rate of polymorphism here is not surprising. However the high rate of polymorphism at Nelspruit is remarkable since the collection was actually made on a farm a few kilometres from the town.

6.3.2.1 Allele frequency ranges and clinal patterns

Figs 6.5 A - H show allele frequencies and approximate 95% confidence intervals for the most common allele of each of the eight enzymes. The figures have been arranged in such a way that the frequencies of geographically adjacent collection sites are in juxtaposition to each other. This was done in order to highlight any clinal pattern which might have emerged due to proximal migration. A few such patterns were apparent. The cline which covered the largest geographic area was from Gobabeb to Thohoyandou for *Adh*⁻¹⁰⁰ which varied from 0,480 at Gobabeb to 1,000 at Thohoyandou (Fig. 6.5A). Smaller clinal patterns were exhibited in the Transvaal in a northern direction from Johannesburg to Pietersburg for both *Hex*-1¹⁰⁰ and *Hex*-2¹⁰⁰. *Hex*-1¹⁰⁰ varied from 0,947 to 0,563 while *Hex*-2¹⁰⁰ varied from 0,962 to 0,750 (Figs 6.5D, E).

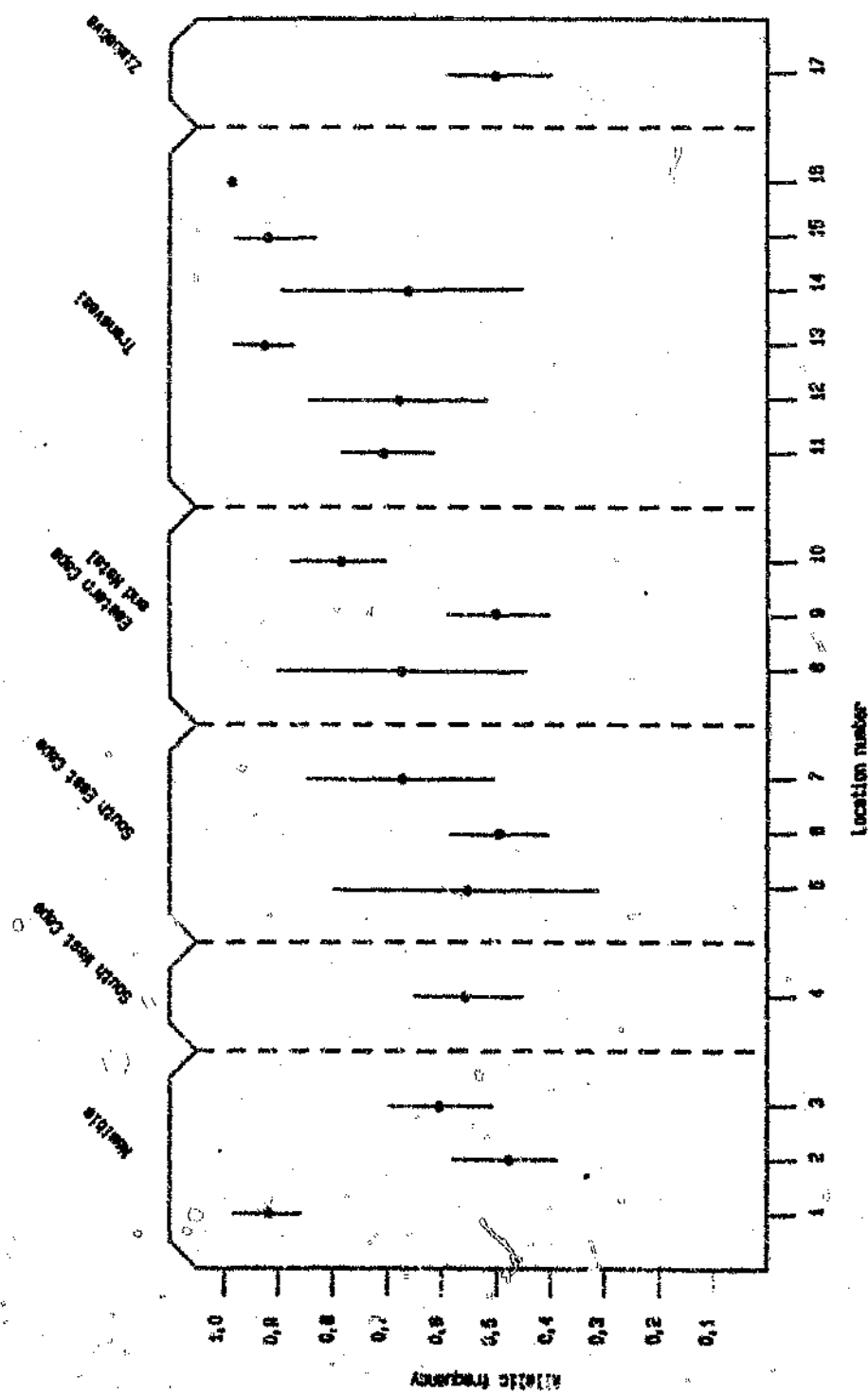


Fig. 6.5 A: Adh-100

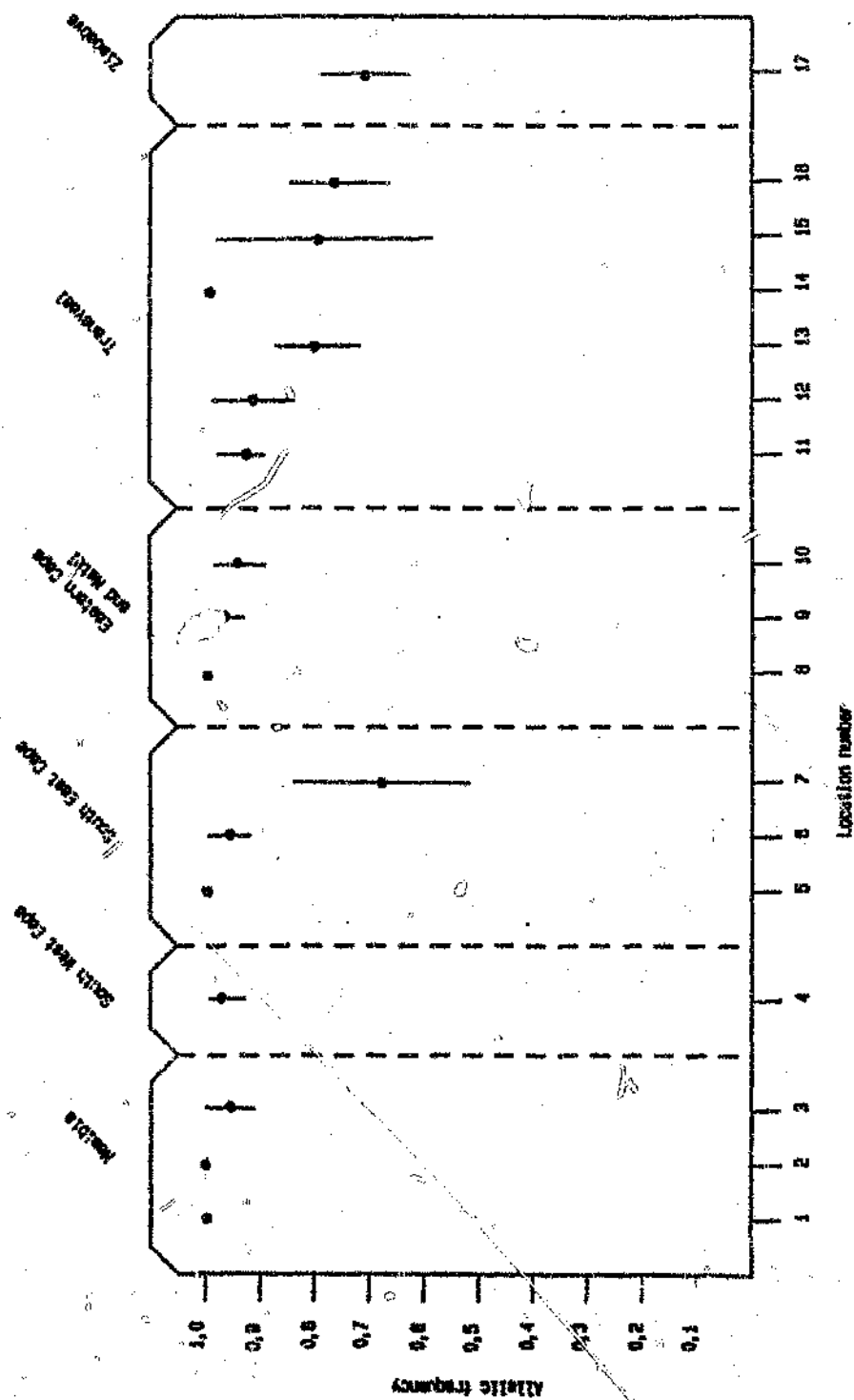


Fig. 6.5 B: Dip-A100

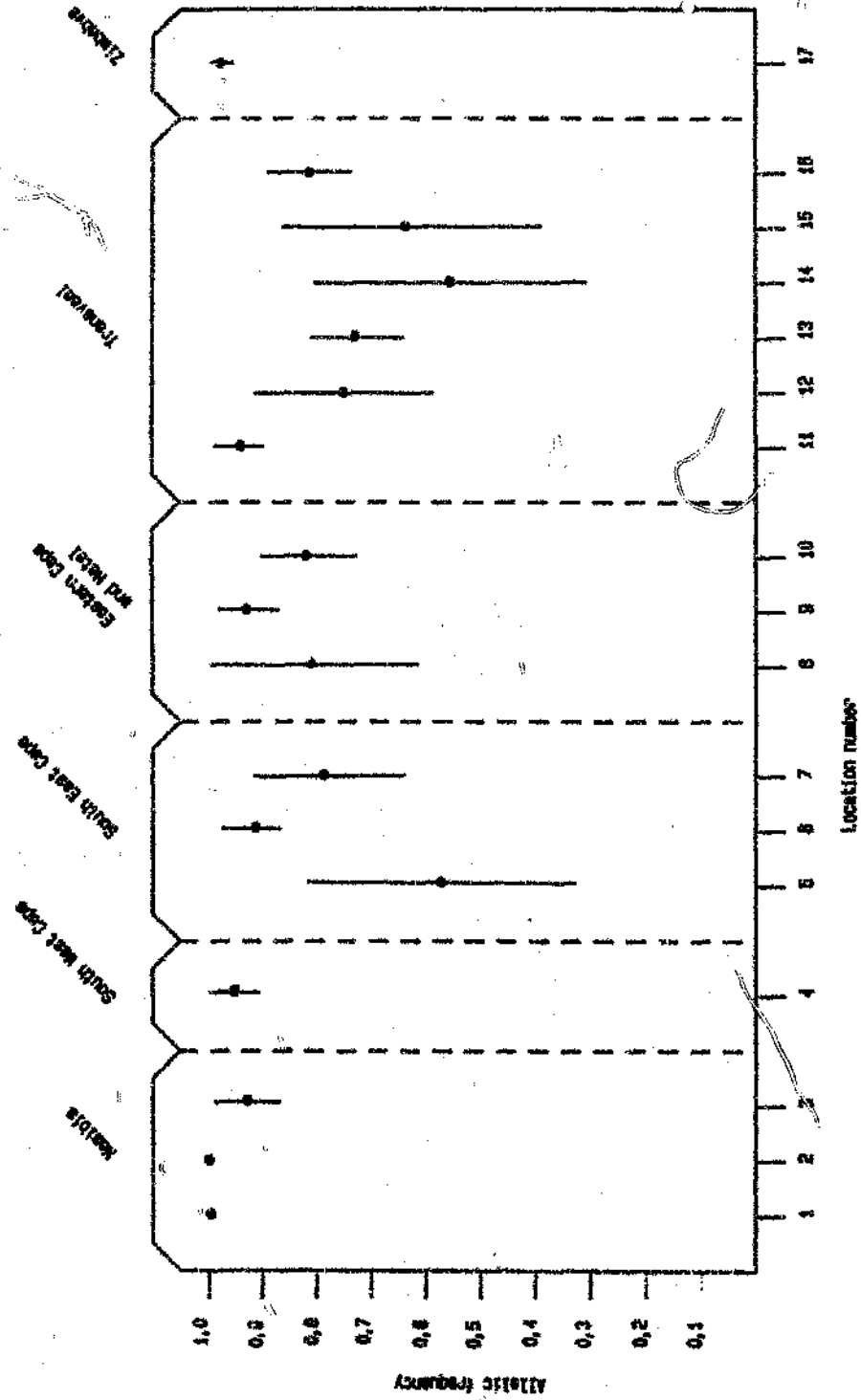


Fig. 6.5 D: Hex-100

6.3.1.5 Isocitrate Dehydrogenase - NADP (*Idh*-NADP)

I observed a single locus with five different alleles producing single- and triple-banded phenotypes typical of a dimeric enzyme. This number of alleles is greater than that reported in studies of non-African populations of *D. melanogaster*. For example, González *et al.* (1982) reported two alleles in the Iberian Peninsula and Tenerife and Choudhary and Singh (1987) report the same number (two) in a worldwide study. On the other hand three *Idh*-NADP alleles have been observed by Oakeshott (1979) for an Australian population and Band (1975) reported the presence of four *Idh*-NADP alleles in a population in South Amherst, Mass.

Two of the alleles, *Idh*-NADP¹¹⁰ and *Idh*-NADP¹²⁰, had very low composite frequencies (0.002 and 0.001 respectively (Table 6.3)) and each was found in only one location. *Idh*-NADP¹¹⁰ was found in the Pretoria/Hammanskraal district and *Idh*-NADP¹²⁰ in Nelspruit. The most common of the other three alleles was the intermediate allele *Idh*-NADP¹⁰⁰ with composite relative frequency of 0.812 (Table 6.3). This was followed in frequency by the slow allele *Idh*-NADP⁸⁵ with composite relative frequency of 0.150 and finally the fastest of these three alleles *Idh*-NADP¹⁰⁷ had composite relative frequency of 0.035 (Table 6.3). These frequencies are in line with the frequencies of the three alleles *Idh*-NADP^F, *Idh*-NADP^S and *Idh*-NADP^{SS} found in a *D. melanogaster* wild type population in Adelaide (Oakeshott 1979).

6.3.1.6 6-Phosphogluconate dehydrogenase (*Pgd*)

The activities of the products of three different alleles were observed for this dimeric sex-linked enzyme. All three allozymes were anodally migrating. On the 95% polymorphism criterion, *Pgd* was monomorphic for the intermediate allele *Pgd*¹⁰⁰ in all collection sites. A very rare fast allele *Pgd*¹¹⁰ (composite relative frequency = 0,001 (Table 6.3)) is likely the same as the rare *Pgd*^{UF} (ultra fast) allele found to be present in Australian populations (Oakeshott *et al.*, 1983a). The slow allele was also fairly rare with composite relative frequency = 0,007 (Table 6.3). It seems likely that the intermediate allele *Pgd*¹⁰⁰ is homologous with that referred to as the allele "3" or the fast allele by Singh *et al.* (1982), Oakeshott *et al.* (1983a) and Singh and Rhomberg (1987a).

6.3.1.7 Phosphoglucomutase (*Pgm*)

Three of the seven known alleles (Trippa *et al.* 1978; Oakeshott *et al.* 1981) for this monomeric anodally migrating enzyme appeared to be present in Southern African populations. In view of the extensive world wide reportage of the preponderance of three *Pgm* alleles of which the intermediate one is present in high frequencies (Trippa *et al.* 1970; Sing *et al.* 1973; Langley *et al.* 1974; Oakeshott *et al.* 1981; Singh *et al.* 1982; Anderson and Oakeshott 1984; Singh and Rhomberg 1987a) and since just such a case exists in this study, it would appear that the Southern African alleles *Pgm*¹⁰⁰, *Pgm*¹²⁰ and

*Pgm*⁷⁰ are homologous to those present in other geographic areas such as Australasia, North America and Europe/Asia (Oakeshott *et al.* (1981), and references therein).

6.3.2 Allele Frequencies

Allele frequencies of the eight loci examined are presented in Appendix E. Tables 6.4 and 6.5. summarize the number of alleles for each locus, the number of collection sites at which the locus was polymorphic as well as the number of polymorphic loci for each collection site. The number of alleles per locus varied from two for *Gpdh* to five for *Idh-NADP* (Table 6.4). Based on the 95% criterion for polymorphism, it can be seen that *Adh* was polymorphic in 16 out of the 17 collection sites whereas *Pgd* was monomorphic throughout (Table 6.4).

Further, based on the 95% criterion for polymorphism, the collection sites exhibiting the least number of polymorphic loci were Swakopmund and Gobabeb (Table 6.5). Both Swakopmund and Gobabeb are fairly inaccessible. The latter is an isolated community in the middle of the Namib desert and the lack of polymorphism in this case is easily explained by the founder effect. This could also account for the lack of polymorphic loci at Swakopmund which is a coastal town on the border of the Namib desert and consequently can be reached only by sea or via the desert. Thus, inhospitable terrain, which must be crossed in order to reach this town would provide a natural barrier to migration.

Table 6.4 Number of alleles and number of samples polymorphic for each of eight loci examined in Southern African populations.

Locus	Number of Alleles	Number (percentage) of polymorphic samples	
		95% criterion	99% criterion
<i>Adh</i>	3	16 (94%)	16 (94%)
<i>Dip-A</i>	3	7 (41%)	12 (70%)
<i>Gpdh</i>	2	7 (41%)	9 (53%)
<i>Hex-1</i>	4	13 (76%)	15 (88%)
<i>Hex-2</i>	3	6* (37%)	11* (69%)
<i>Idh-NADP</i>	5	15 (88%)	16 (94%)
<i>Pgd</i>	3	0 (0%)	4 (23%)
<i>Pgm</i>	3	6 (35%)	9 (53%)

*A total of 17 samples was examined for seven of the enzymes. As *Hex-2* was not scored in Thohoyandou, there was a total of 16 samples surveyed for this locus.

Table 6.5 Number of loci polymorphic in each of the Southern African samples

	Collection Site	Number (percentage of loci polymorphic)	
		95% criterion	99% criterion
1	Swakopmund	2 (25%)	3 (37%)
2	Gobabeb	2 (25%)	2 (25%)
3	Ai Ais	4 (50%)	8 (100%)
4	Cape Town	4 (50%)	7 (87%)
5	Plettenberg Bay	4 (50%)	4 (50%)
	Port Elizabeth		
6	Summerstrand	5 (62%)	7 (87%)
7	Walmer	5 (62%)	6 (75%)
8	Maclear	3 (37%)	3 (37%)
	Durban		
9	Berea	4 (50%)	7 (87%)
10	Addington Beach	4 (50%)	5 (62%)
11	Johannesburg	4 (50%)	7 (87%)
12	Pretoria/Hammanskraal	6 (75%)	7 (87%)
13	Nelspruit	7 (87%)	8 (100%)
14	Pietersburg	4 (50%)	4 (50%)
15	Agatha	5 (62%)	5 (62%)
16	Thohoyandou	3* (43%)	4* (57%)
17	Harare	4 (50%)	6 (75%)

*As Hex-2 was not scored in Thohoyandou, there was a total of seven loci examined at this collection site. At all other collection sites, 8 loci were analysed.

Fig. 6.5

Allele frequencies (approximate 95% confidence intervals) for the most common allele of each of the eight enzyme loci:

A: Adh
 B: Dip-A
 C: Gdh
 D: Hex-1
 E: Hex-2
 F: Idh-NADP
 G: Pod
 H: Pgm

Location numbers are:

- 1 Swakopmund
- 2 Gobabeb
- 3 At Ais
- 4 Cape Town
- 5 Plettenberg Bay
- 6 Summerstrand (Port Elizabeth)
- 7 Walker (Port Elizabeth)
- 8 Maclear
- 9 Edeza (Durban)
- 10 Addington Beach (Durban)
- 11 Johannesburg
- 12 Pretoria/Hammanskraal
- 13 Nelspruit
- 14 Pietersburg
- 15 Agatha
- 16 Thohoyandou
- 17 Harare

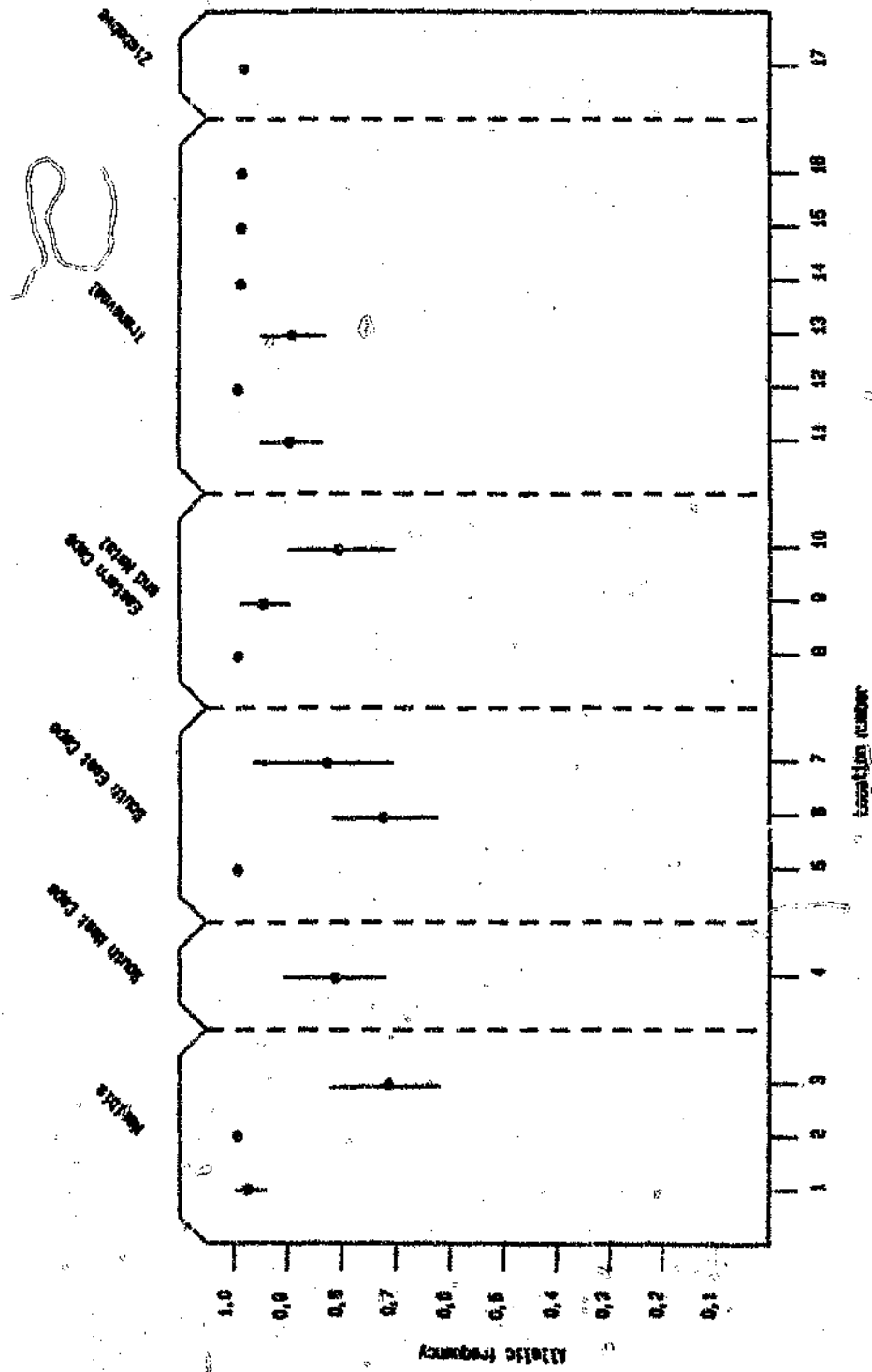


Fig. 6.5 C: Gpdh 100

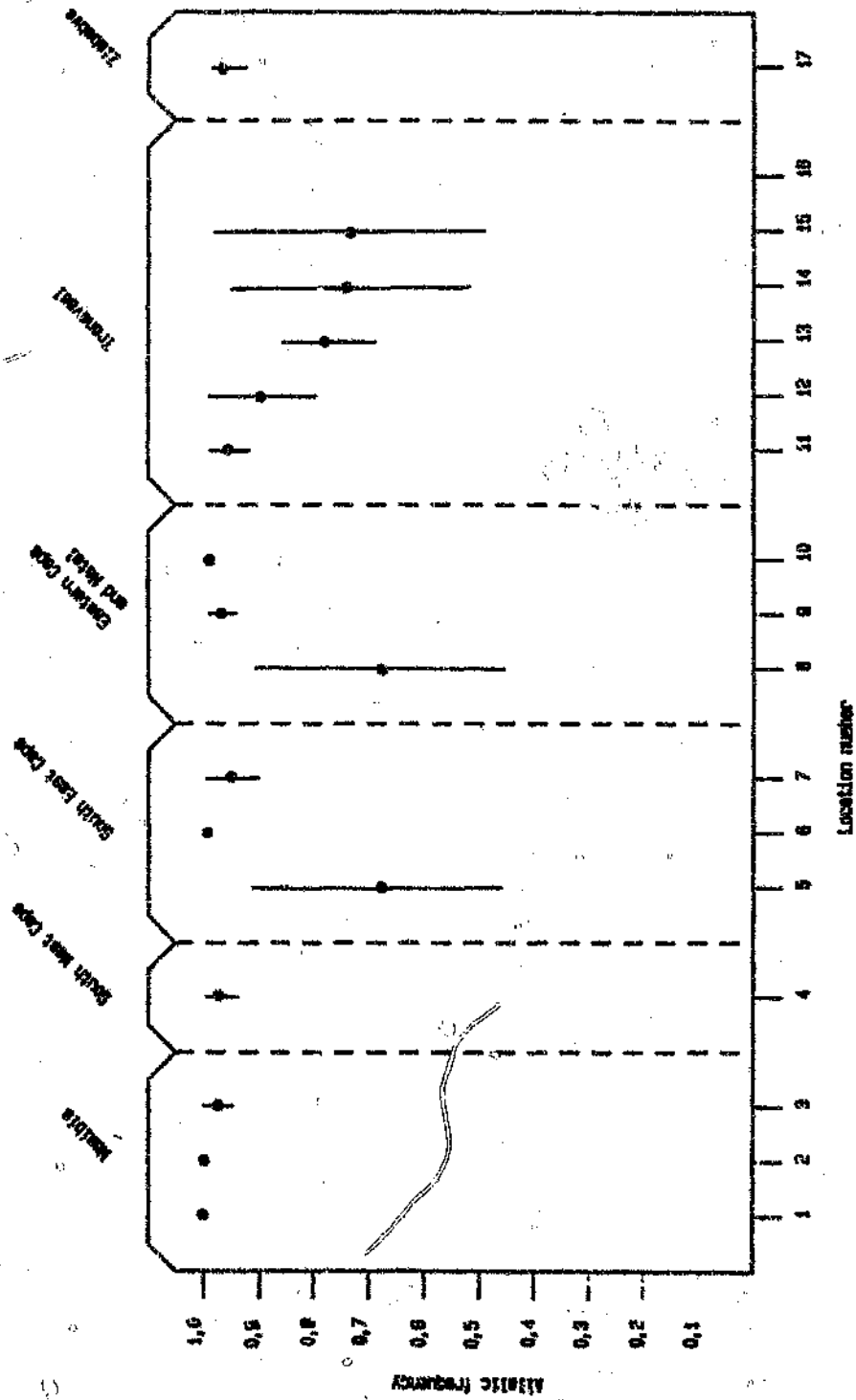


Fig 6.5 E: Max-2100

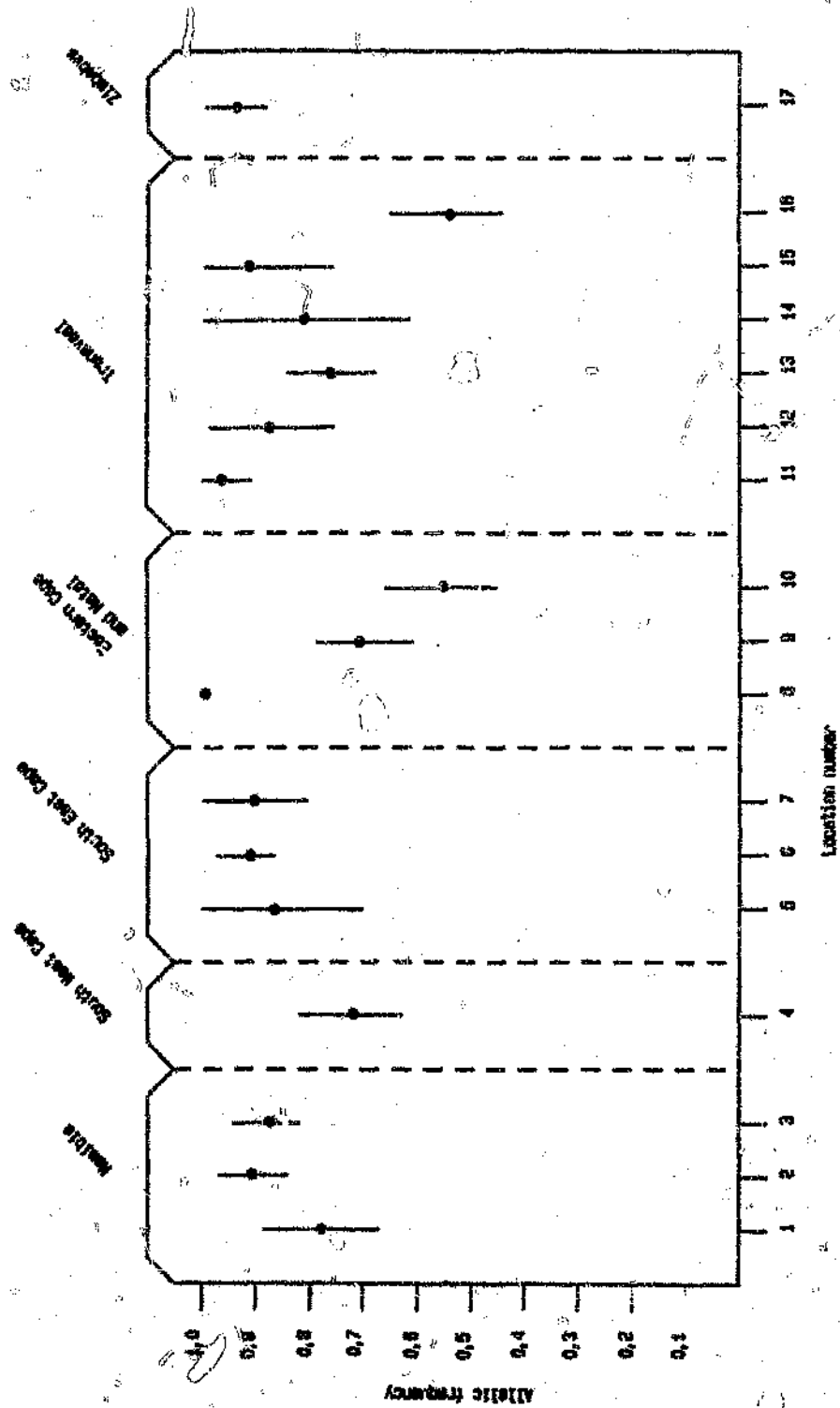


Fig. 6.5 F: Idh-HA3100

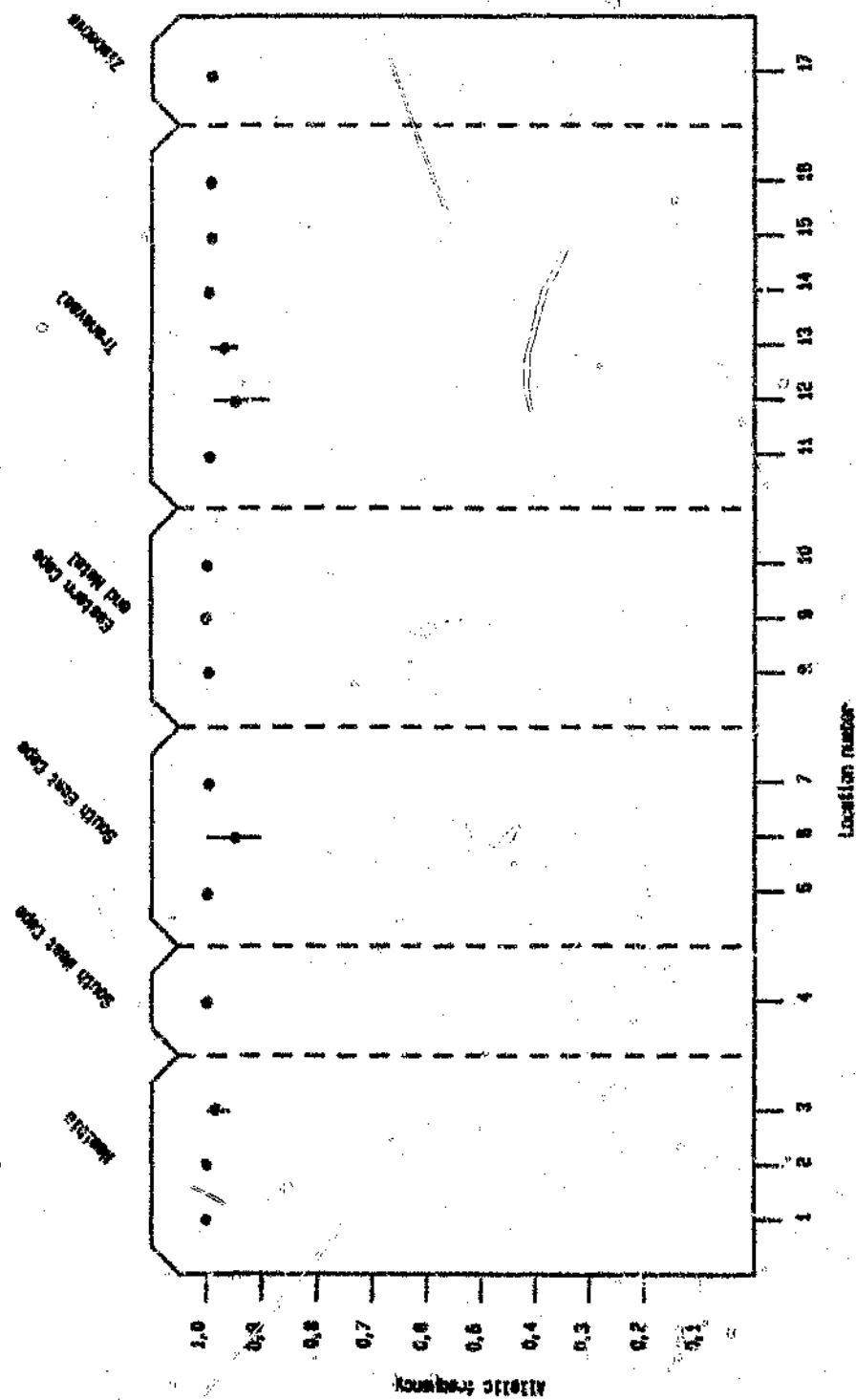


Fig. 6.5 6: Pod100

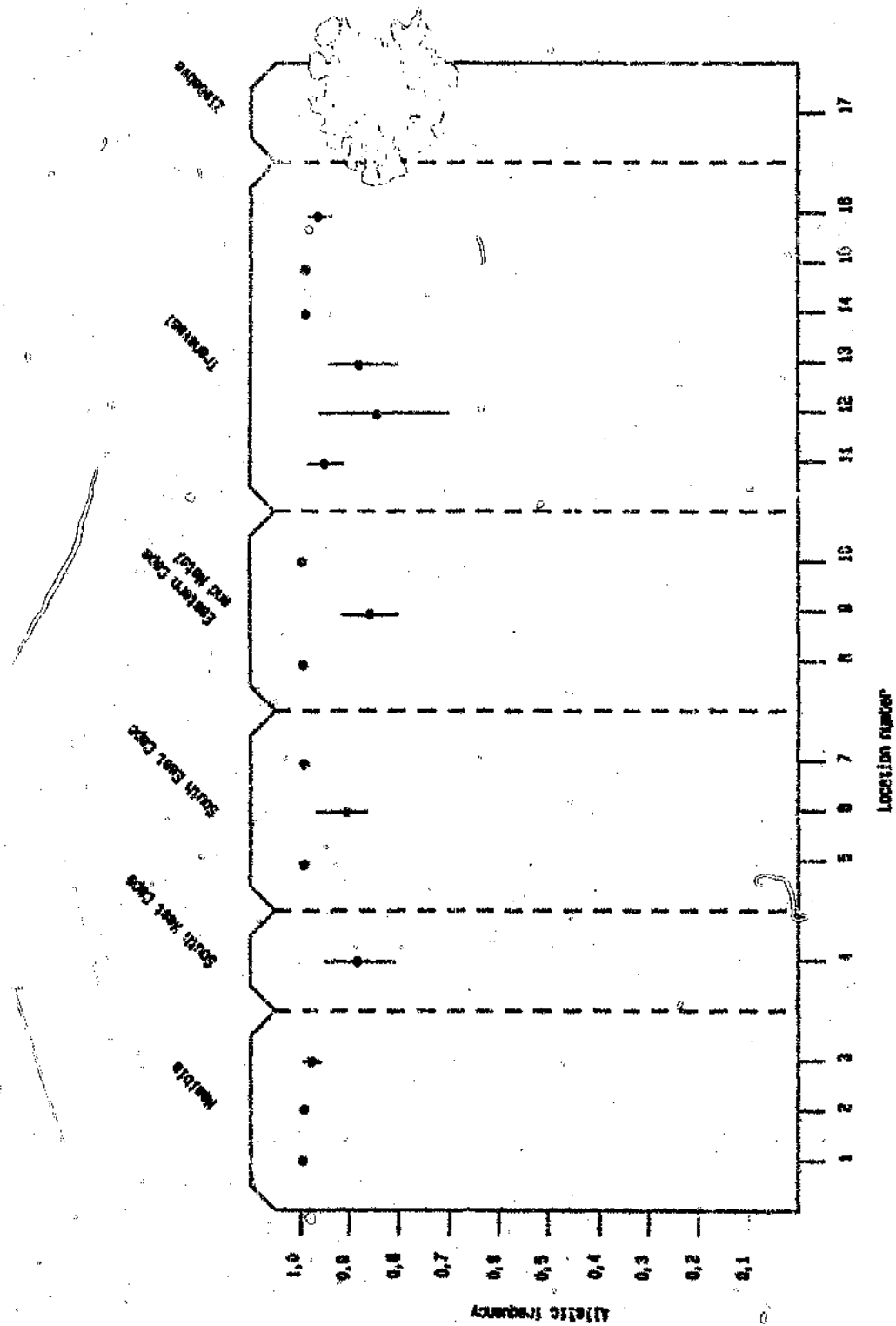


Fig. 6.5 H: Page 100

Finally *Idh*-NADP (Fig. 6.5F) exhibited several small clinal patterns. This allele increased in frequency from 0,738 in Cape Town to 0,926 in Port Elizabeth. The direction of this cline was reversed in the Transvaal. Frequencies of *Idh*-NADP¹⁰⁰ decreased from 0,961 in Johannesburg to 0,760 in Nelspruit and from 0,917 in Agatha to 0,540 in Thohoyandou.

Allele frequencies were generally confined to relatively narrow ranges. The frequency of *Adh*¹⁰⁰ was the most variable ranging from a low of 0,480 in Gobabeb to a high of 1,000 in Thohoyandou (Fig. 6.5A). *Idh*-NADP¹⁰⁰ had a slightly narrower frequency range than *Adh*¹⁰⁰, varying from 0,540 in Thohoyandou to 1,000 in Maclear (Fig. 6.5F). The narrowest range was that of *Gpd*¹⁰⁰ from 0,951 in Summerstrand to 1,000 in most other places (Fig. 6.5G).

6.3.3 Nested Contingency Table Analysis (G-tests)

The total heterogeneity in allelic frequencies of six polymorphic loci, *Adh*, *Pip-A*, *Gpdh*, *Hes-1*, *Hes-2* and *Idh*-NADP, was partitioned in the manner described in 6.2.3.1 and allele frequency differences among samples were compared using a nested contingency table analysis. This analysis of variation of these six loci is given in Table 6.6. Sections 6.3.3.1 - 6.3.3.4 refer to values obtained in this table.

Table 6.6

Nested contingency-table analysis of allele frequencies in Southern African populations of *D. melanogaster*. The G-test (log likelihood ratio test) was used to detect significant frequency differences.

Source of variation	Locus												Total	df
	<i>Adh</i>	df	<i>Dip-A</i>	df	<i>Gpdh</i>	df	<i>Hex-1</i>	df	<i>Hex-2</i>	df	<i>Idh-NADP</i>	df		
Among Regions	93.43***	5	70.06***	5	47.08***	5	61.09***	5	38.47***	5	68.83***	5	378.95 ^c	30
Namibia														
Between areas	0.54	1	1.06	1	38.10***	1	6.31	1	0.26	1	0.18	1	46.44 ^c	6
Between districts	39.32***	1	0.13	1	0.13	1	0.13	1	0.13	1	5.41	1	45.25 ^c	6
Southeastern Cape														
Between districts	0.02	1	0.26	1	3.61	1	9.05*	1	11.35**	1	0.36	1	24.66 ^c	6
Between sites	3.58	1	17.06***	1	2.09	1	3.93	1	0.07	1	0.13	1	26.86 ^c	6
Eastern Cape/Natal														
Between districts	0.20	1	0.25	1	0.28	1	0.60	1	19.32***	1	7.57*	1	28.22 ^c	6
Between sites	18.86***	1	0.44	1	8.23*	1	6.46	1	0.02	1	4.53	1	38.53 ^c	6
Transvaal														
Among areas	52.89***	3	18.23**	3	9.84*	3	24.01***	3	19.43***	2	62.36***	3	186.76 ^c	17
Between districts														
So. Transvaal	0.13	1	0.04	1	1.53	1	9.61*	1	1.48	1	2.95	1	15.73 ^a	6
No. Transvaal	3.53	1	1.19	1	0.00	1	0.13	1	0.00	1	0.64	1	5.49	6
Total	212.50***	16	108.71***	15	110.88***	16	121.32***	16	90.63***	16	152.95***	16	796.80 ^c	95

* P<0.05/6
 ** P<0.01/6
 *** P<0.001/6
 a P<0.05
 b P<0.01
 c P<0.001

6.3.3.1 Heterogeneity between sites within districts

Overall allele frequency differences between sites within districts were highly significant in both the Southeastern Cape (Port Elizabeth) and Natal (Durban). In Port Elizabeth, a single locus, *Dip-A* ($G = 17,06$; $P < 0,001/6$) was responsible for this heterogeneity. In Durban, two loci, *Adh* ($G = 18,86$; $P < 0,001/6$) and *Gpdh* ($G = 8,23$; $P < 0,05/6$) were the main contributors to this lack of homogeneity. These allele frequency differences appeared between samples 10-20 kilometres apart.

6.3.3.2 Heterogeneity between districts within areas

Significant allele frequency differences were found among districts in northern Namibia for *Adh* ($G = 39,32$; $P < 0,001/6$), in the Southeastern Cape for *Hex-1* ($G = 9,05$; $P < 0,05/6$), *Hex-2* ($G = 11,35$; $P < 0,01/6$), in the Eastern Cape/Natal for *Hex-2* ($G = 19,32$; $P < 0,001/6$) and *Idh-NADP* ($G = 7,57$; $P < 0,05/6$), and in the Southern Transvaal for *Hex-1* ($G = 9,61$; $P < 0,05/6$). These among/~~between~~ district comparisons represented geographic heterogeneity in allele frequencies on the order of 50 to 200 kilometres.

6.3.3.3 Heterogeneity between areas within regions

Between the northern and southern parts of Namibia a single locus *Gpdh* had significantly different allele frequencies ($G = 38,10$; $P < 0,001/6$). The high degree of heterogeneity exhibited in the allele frequencies within the Transvaal was present to a greater or lesser extent in all six loci analysed.

6.3.3.4 Heterogeneity between regions

The total frequency differences as well as the frequency differences at individual loci were all highly significant for the comparison between regions.

6.3.3.5 Comparison of heterogeneity between levels

The above results indicated a significant amount of allele frequency heterogeneity between population samples within Southern Africa. This genetic difference increased with increasing distance between sample populations as is evidenced by the increase in the standardised measure which reached a maximum value of 12.63 at the between-regions level. These differences in allele frequencies between sites, districts, areas and regions could not be ascribed to any particular enzyme. There appeared to be a random assortment of loci responsible for this genetic variation.

Because of the lack of homogeneity manifested at the lowest level, the observations of significant heterogeneity at higher levels must be viewed with caution. This is because it is statistically unsound to pool heterogeneous allele frequency data for higher level testing since an indication of higher level heterogeneity could simply be due to low level variation being passed along. For this reason, approximate F ratios were computed by summing the likelihood (G) statistics over loci, dividing this sum by its associated degrees of freedom and comparing this standardised measure with the standardised measure at the next lower level (Smouse and Kojima 1972).

The F statistic tests the null hypothesis that the degree of allele frequency heterogeneity at the upper level is no greater than that at the next lower level. These results are presented in Table 6.7 which shows that the amount of heterogeneity between districts was not significantly greater than the heterogeneity present between sites within districts. At higher levels, there was a significantly greater degree of heterogeneity among areas than among districts ($F_{23,30} = 2.55$; $P < 0.01$). There was not, however, a greater degree of heterogeneity in the regions to areas comparison. When the standardized measure of heterogeneity among districts is used as an error term, there was a highly significant degree of heterogeneity among regions ($F_{30,30} = 3.17$; $P < 0.005$).

These F statistics are unbalanced, that is the distribution of samples did not allow for a nested set of comparisons for each area, and I had to assume that the degree of heterogeneity over a given distance in one area would be similar to that in another area. Since Southern Africa is unevenly populated, suitable habitats for *Drosophila* are probably not evenly distributed and my assumption may not be valid. I therefore made another set of comparisons using only those areas with a more complete set of nested samples. Neither of the two districts/sites F ratios were significant, as was the case for the unbalanced tests. At the next higher nested level, the areas/districts ratio was significant for the samples from the Transvaal ($F_{17,12} = 6.21$; $P < 0.005$), but not for the Namibian samples. Thus it appears that there is more differentiation among areas in some regions than in other regions.

Table 6.7 Approximate F ratios

A Unbalanced F ratios: The total standardized measure of one level is compared with that of a lower level.

B Balanced F ratios: The standardized measure obtained between particular geographic areas or districts is compared with the standardized measure obtained between geographic districts or collection sites (at a lower level) that are, physically contained within the (higher level) areas or districts.

A

	Unbalanced Design
<u>Regions</u>	$\frac{12,63_{10}}{10,14_{23}} = 1,25$
<u>Areas</u>	
<u>Areas</u>	$\frac{10,14_{13}}{3,98_{30}} = 2,55^*$
<u>Districts</u>	
<u>Districts</u>	$\frac{3,98_{30}}{5,45_{12}} = 0,73$
<u>Sites within Districts</u>	
<u>Regions</u>	$\frac{12,63_{30}}{3,98_{30}} = 3,17^{**}$
<u>Districts</u>	

B

	Balanced Design
<u>Namibia</u>	$\frac{7,74_6}{7,54_6} = 1,03$
<u>N.Namibia</u>	
<u>Transvaal</u>	$\frac{10,99_{17}}{1,77_{12}} = 6,21^{**}$
<u>S.Tvl+N.Tvl</u>	
<u>S.E.Cape</u>	$\frac{4,11_6}{4,48_6} = 0,92$
<u>Port Elizabeth</u>	
<u>E. Cape+Natal</u>	$\frac{4,70_6}{6,48_6} = 0,73$
<u>Durban</u>	

* P<0,01

** P<0,005

In summation, the following three facts emerged from the likelihood analysis:

1. There was a significant amount of heterogeneity between locations. Even locally, within the same city, two collection sites ten kilometres apart produced significantly different allele frequencies.
2. This heterogeneity increased as the distance between collection sites increased.
3. The difference in allele frequencies between collections increased dramatically when progressing from the between district level to the between areas level. Heterogeneity continued to increase to reach a maximum at the between regions level.

The F -statistics obtained in this study tended to compare the amount of genetic heterogeneity between samples taken from collection sites further apart to that between samples taken from collection sites nearer together. Unfortunately the geographic distribution of the samples precluded a balanced nested set of comparisons for each district, area and region. Thus unbalanced statistics were obtained. However, when these results are viewed in conjunction with the gene diversity analysis (Section 6.3.4) a clear picture of the Southern African population structure leading to genetic differentiation emerges.

6.3.4 Gene Diversity Analysis

Heterozygosity estimates revealed that an average of 87,75% of the total genetic variation observed in the Southern African populations could be ascribed to variation between individuals within the same collection site ($\bar{H}_S/\bar{H}_T = 0,8775$, Table 6.8). Population subdivision and genetic differentiation therefore accounted for 12,25% of the total genetic diversity in the *D. melanogaster* populations sampled in this study.

6.3.4.1 Between collection sites within districts

The proportional decrease in total heterozygosity resulting from allele frequency differences between sites within districts ranged from 0,04% (*Hex-2*) to 2,87% (*Dip-A*) and averaged 1,1% (Table 6.8). When measured as a percentage of the total genetic diversity among population samples, gene diversity estimates for collection site differences ranged from 0,32% ($D_{SD} = 0,00008$) at the *Hex-2* locus to 32,61% ($D_{SD} = 0,00015$) at the *Pgd* locus, and averaged 9,26% ($\bar{D}_{SD} = 0,00228$) for all loci combined (Table 6.9).

Table 6.8 Estimates of heterozygosity and gene differentiation coefficients for Southern African populations of *D. melanogaster*

Locus	Heterozygosity ¹					Relative gene diversity components (x10) ²				
	H_T	H_R	H_A	H_D	H_S	G_{RT}	G_{AR}	G_{DA}	G_{SD}	H_S/H_T
<i>Adh</i>	0,4274	0,4043	0,3949	0,3789	0,3723	0,5403	0,2197	0,3739	0,1550	0,8711
<i>Dip-A</i>	0,1584	0,1486	0,1454	0,1409	0,1363	0,6196	0,2011	0,2876	0,2875	0,8604
<i>Gpdh</i>	0,1449	0,1397	0,1330	0,1277	0,1257	0,3579	0,4614	0,3647	0,1394	0,8676
<i>Hox-1</i>	0,2994	0,2847	0,2773	0,2686	0,2662	0,4927	0,2455	0,2919	0,0786	0,8891
<i>Hox-2</i>	0,1808	0,1728	0,1679	0,1562	0,1561	0,4425	0,2712	0,6472	0,0044	0,8634
<i>Idh-NADP</i>	0,2908	0,2809	0,2700	0,2572	0,2560	0,3383	0,3780	0,4366	0,0429	0,8804
<i>Pgd</i>	0,0140	0,0139	0,0138	0,0137	0,0136	0,0471	0,0658	0,1089	0,1049	0,9673
<i>Pgm</i>	0,0912	0,0876	0,0861	0,0850	0,0838	0,3944	0,1618	0,1178	0,1360	0,9190
Average	0,2009	0,1916	0,1861	0,1785	0,1763	0,4629	0,2738	0,3783	0,1095	0,8775

¹ H_T estimates the total heterozygosity present under the assumption of panmixia among sampled populations. H_R , H_A , H_D , and H_S are the average heterozygosities within regions, areas, districts and collection sites, respectively

² G_{RT} , G_{AR} , G_{DA} and G_{SD} measure the proportional decrease in total heterozygosity due to population sub-division and genetic differentiation among regions, areas within regions, districts within areas, and sites within districts, respectively. G_{XY} is the relative gene diversity component given by $G_{XY} = (H_Y - H_X)/H_T$ where $XY = RT, AR, DA$ or SD . H_S/H_T measures the proportion of the total variation that is present within sample sites.

Table 6.9 Absolute gene diversity estimates at the region, area, district and collection site levels.

Locus	Absolute decrease in heterozygosity ¹					Proportional decrease in heterozygosity (%)			
	$D_{RT} \times 10$	$D_{AR} \times 10$	$D_{DA} \times 10$	$D_{SD} \times 10$	Total $\times 10$	% D_{RT}	% D_{AR}	% D_{DA}	% D_{SD}
<i>Adh</i>	0,2309	0,0939	0,1598	0,0663	0,5509	41,92	17,04	29,01	12,03
<i>Dip-A</i>	0,0982	0,0319	0,0456	0,0456	0,2213	44,39	14,41	20,60	20,60
<i>Gpdh</i>	0,3519	0,0669	0,0528	0,0202	0,4918	27,06	34,88	27,53	10,53
<i>Hes-1</i>	0,1475	0,0735	0,0874	0,0235	0,3319	44,45	22,14	26,32	7,09
<i>Hes-2</i>	0,3800	0,0490	0,1170	0,0008	0,5468	32,41	19,86	47,41	0,32
<i>Idh-NADP</i>	0,0984	0,1099	0,1270	0,0125	0,3478	28,29	31,61	36,51	3,59
<i>Pgd</i>	0,0007	0,0009	0,0015	0,0015	0,0046	15,22	19,56	32,61	32,61
<i>Pgm</i>	0,0360	0,0147	0,0107	0,0124	0,0738	48,69	19,97	14,54	16,80
Average	0,0930	0,0551	0,0752	0,0228	0,2461	37,79	22,39	30,56	9,26

¹ D_{RT} , D_{AR} , D_{DA} and D_{SD} measure the absolute decrease in heterozygosity resulting from population subdivision and genetic differentiation at the region, area, district and collection site level, respectively. The corresponding proportional decreases in genetic diversity are shown as percentages i.e. % $D_{XY} = (D_{XY} \times 100) / (D_{RT} + D_{AR} + D_{DA} + D_{SD})$, for $XY = RT, AR, DA$ or SD .

6.3.4.2 Between districts within areas

The average gene diversity between districts within areas ($\bar{D}_{DA}$) accounted for 30,56% of the total gene diversity among population samples (Table 6.9). The contribution of this between districts component varied from a low of 14,54% ($D_{DA} = 0,00107$) for *Pgm* to a high of 47,41% ($D_{DA} = 0,01170$) for *Hex-2* (Table 6.9). Approximately 3,78% ($G_{DA} = 0,03783$) of the proportional decrease in heterozygosity was due to allele frequency differences between districts within areas (Table 6.8). Comparison of the two average absolute gene diversities i.e. between collection sites within districts and between districts within areas ($\bar{D}_{SD}$ and $\bar{D}_{DA}$) showed that the latter was more than three times that of the former (Table 6.9), thus indicating the relative importance of spatial differences at these levels. Two of the eight enzymes showed a maximal decrease in heterozygosity due to subdivision into districts within areas. They were *Hex-2* and *Idh-NADP*.

6.3.4.3 Between areas within regions

Population structuring and genetic differentiation at the between areas level accounted for 2,74% of the proportional decrease in total heterozygosity (Table 6.8). This proportional decrease ranged from 0,66% for *Pgd* to 4,61% for *Gpdh* (Table 6.8). The proportion of total genetic diversity among population samples due to differences between areas ($\bar{D}_{AR}$) averaged 22,39% ($\bar{D}_{AR} = 0,00551$) with a low of 14,41% ($D_{AR} = 0,00319$) for *Dip-A* and a high of 34,88% ($D_{AR} = 0,00669$) for *Gpdh* (Table 6.9). The only locus which showed a maximum decrease in heterozygosity due to area subdivision within regions was *Gpdh*.

6.3.4.4 Between regions

Differences in heterozygosity between the six regions: Namibia, Southeastern Cape, Southwestern Cape, Eastern Cape/Natal, Transvaal and Zimbabwe, accounted for the single greatest contribution to both the relative and absolute gene diversity components due to variation between subpopulations. Half of the loci exhibited a maximal decrease in heterozygosity due to regional subdivision, viz: *Adh*, *Dip-A*, *Hex-1* and *Pgm*. The contribution to absolute gene diversity due to regional subdivision was greater than 1,5 times that due to area subdivision within regions. ($\bar{D}_{AR} = 0,00551$, $\bar{D}_{RT} = 0,00930$, Table 6.9).

The decrease in heterozygosity due to regional subdivision averaged 4,63%. The minimum value was 0,47% for *Pgd* and the maximum value was 6,20% for *Dip-A* (Table 6.8). When measured as a proportion of the total genetic diversity among population samples, gene diversity estimates for regional differences ranged from 15,22% ($D_{RT} = 0,00007$) at the *Pgd* locus to 48,69% ($D_{RT} = 0,0036$) at the *Pgm* locus, and averaged 37,79% ($\bar{D}_{RT} = 0,00930$) for all loci combined (Table 6.9).

6.3.4.5 Comparison of gene diversity among population levels

The relative decreases in heterozygosity are illustrated graphically in the pie diagram (Fig. 5.6) which indicates the proportional contribution to total gene diversity due to population subdivision. Since 87,75% ($\bar{H}_S/\bar{H}_T = 0,8775$) of the total heterozygosity was due to differences between individuals within population samples, there remained 12,25% which was apportioned to the different levels as follow (Table 6.8):

- 4,63% ... to separation of 500-3000 km
- 2,74% due to separation of 400-800 km
- 3,78% due to separation of 50-300 km
- 1,10% due to separation of approximately 10 km

The major decrease in heterozygosity was due to regional subdivision and the second most important contribution to the decrease in heterozygosity was due to district subdivision within areas. Thus one would expect a significant difference in allele frequency distribution due to district subdivision within areas and this is exactly what was obtained in the heterogeneity results, i.e. the heterogeneity in samples between areas was found to be significantly greater than that between districts (Section 6.3.3.5).

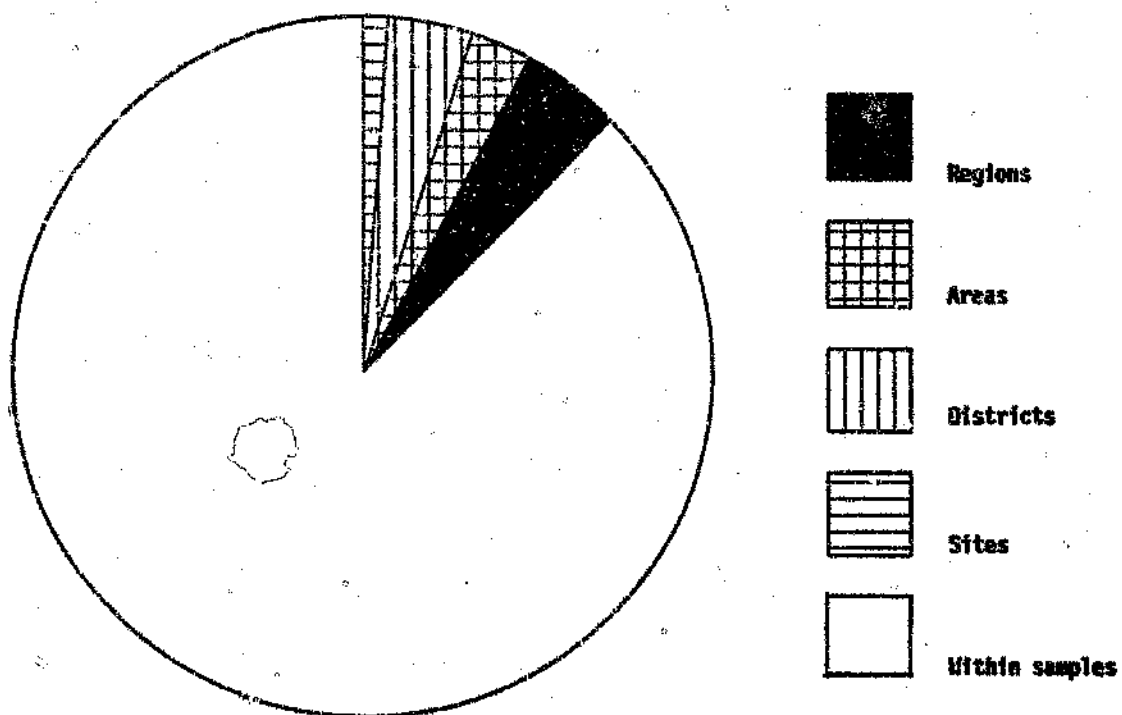


Fig. 6.6

Illustration of the proportional contribution to total gene diversity due to population subdivision.

6.3.5 Estimation of Gene Flow

Slatkin (1981), using the assumptions of the infinite alleles model for a simulation study, showed that the conditional average frequency of an allele (see below for definition) was approximately independent of selection intensity and mutation rate, but depended strongly on the overall level of gene flow. The conditional average frequency of an allele is obtained as follows:

For any allele, the occupancy number, i , is defined as the number of demes in which the allele is present. If n is the number of alleles having occupancy number i , and if $\overline{p}_r(i)$ ($r = 1, \dots, n$) is the frequency (averaged over i) of each of these n alleles, then the conditional average frequency $\overline{p}(i)$, is defined by:

$$\overline{p}(i) = \sum_{r=1}^n \overline{p}_r(i) / n$$

Unfortunately, as Slatkin (1981) commented, his results do not furnish one with an analytic theory that would allow the estimation of the true or effective migration rate using a statistic derived from $\overline{p}(i)$. Further, it is not clear

why $\bar{P}(i)$ should depend only on the migration rate, although it is intuitively clear that the occupancy number, i , should vary with rate of migration.

For any population an incoming migrant with an allele not already in that population would increase the value of i . Further, for rare alleles particularly, migration out could remove that allele entirely from the deme and thus decrease i . The relationship between conditional average frequency and migration rate is an inverse one in that, the higher the migration rate, the lower the values of $\bar{P}(i)$.

Fig. 6.7 shows the conditional average frequencies of alleles in the 17 populations considered in this study as a function of the proportion of populations in which they were found. If the results of Slatkin (1981) for other species are used as a standard measure, the graph in Fig. 6.7 indicates that the Southern African *D. melanogaster* races would be characterized as having high gene flow.

Slatkin (1985) obtained a method for estimating Nm (the average number of migrant exchanges between populations) by means of private alleles, (i.e. alleles appearing in only a single population). Use of this method and correcting for sample size gave a value of Nm of 6.9. However, as only eight loci were evaluated in this study, and as only three private alleles were obtained, this estimate of Nm must be viewed with caution.

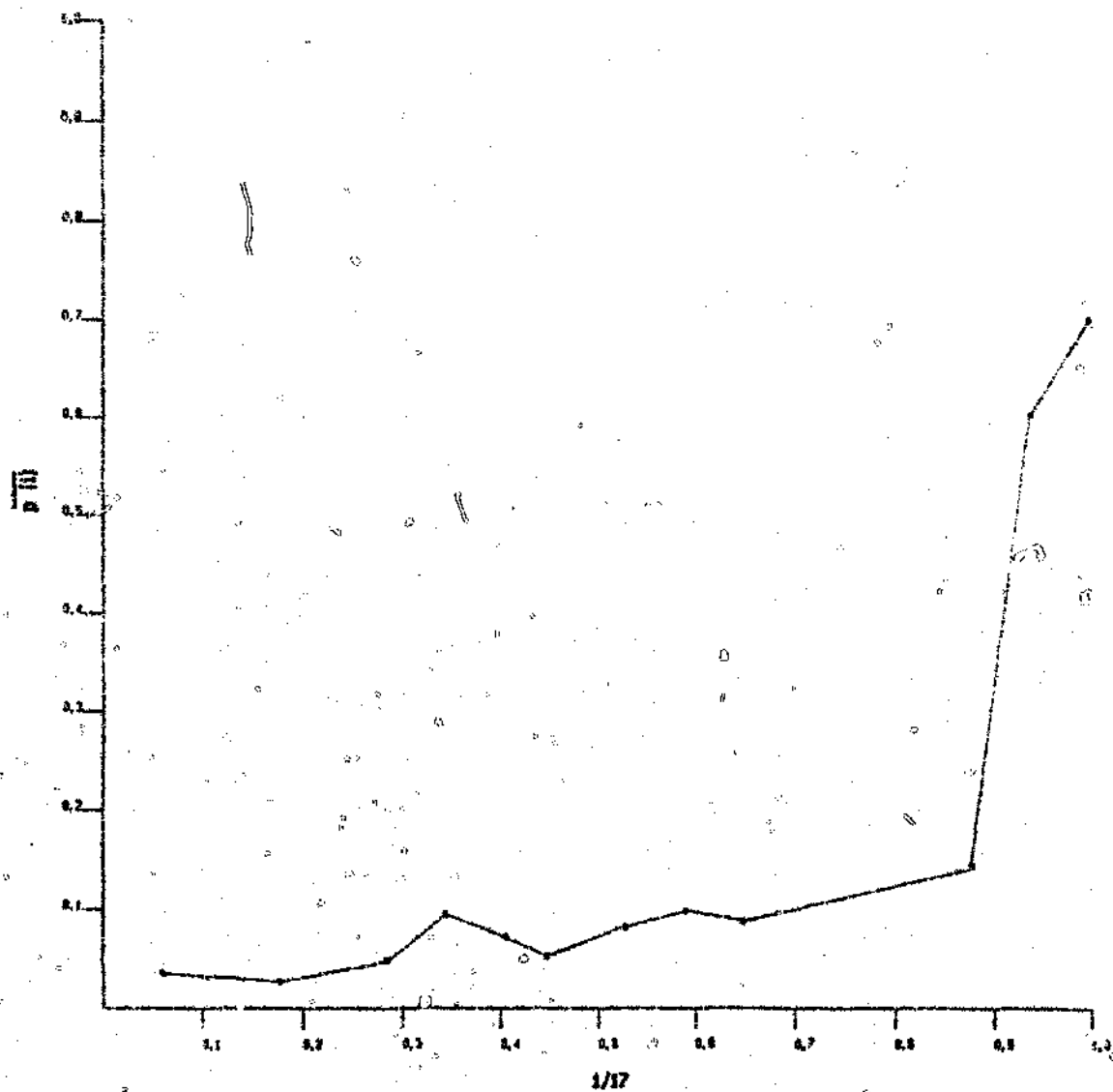


Fig. 6.7

Conditional average frequency $p(i)$, plotted against the occupancy rate (i/d) , where d is the total number of clines studied and i is the number of clines occupied by an allele.

Migration can also be estimated from allele frequency data using an estimate of the fixation index F_{ST} and the island model of migration (Wright 1943).

$$F_{ST} = 1/(1 + 4Nm)$$

where N is the effective population size assumed to be equal in all populations, and m is the proportion of migrants between each subpopulation. I have used G_{ST} ($1 - H_S/H_T$) instead of F_{ST} since G_{ST} is an extension of F_{ST} for loci with more than two alleles. Using an estimate of 0.1225 for G_{ST} (Table 6.8) and solving for Nm gives a migration rate of 1.8 individuals per generation.

Both of these methods of estimating gene flow suggest that gene flow among Southern African populations is relatively high, and this is in accordance with estimates obtained by Singh and Rhomberg (1987a) who estimated average migration rates of 1-2 individuals for populations of *D. melanogaster* worldwide.

6.3.6 Comparison of Southern African Distributions with Those in Other Parts of the World

Table 6.10 lists some comparisons between G_{ST} values obtained in Southern African populations and F_{ST} values of other authors[†].

6.3.6.1 Alcohol dehydrogenase (*Adh*)

The G_{ST} value of 0,1289 for *Adh* in Southern African was similar to that observed for this locus in tropical Africa and in the tropical parts of Australia (David *et al.* 1989) where values of 0,134 and 0,123 respectively were obtained. On the other hand, a value of 0,5 was generally observed for global comparisons between regions (Singh *et al.* 1982; Choudhary and Singh 1987; David *et al.* 1989). This is not surprising in view of the large degree of differentiation in the frequency of the *Adh*^F allele worldwide (David *et al.* 1989) which is in keeping with the putative adaptive properties of the ADH enzyme and its consequent latitudinal variation, particularly in the northern hemisphere.

[†]Wright's fixation index, F_{ST} , is given in terms of a two allelic system for a single locus whereas the evaluation of G_{ST} incorporated more than one locus with more than two alleles. So, strictly speaking, a comparison of G_{ST} with F_{ST} (even for a single locus) is mathematically incorrect, unless the weighted average F_{ST} is used (Wright 1978) or unless the given locus is diallelic. However Table 6.10 uses results from Singh *et al.* (1982) and from Choudhary and Singh (1987). Their method of computing average heterozygosity is that of Lewontin and Hubby (1966) in which the formula

$$H = 1 - \sum_{i=1}^n p_i^2$$

(p_i is the frequency of the i th allele; n) is used. In this case F_{ST} and G_{ST} are identical.

Table 6.10

Comparison between G_{ST} values obtained in Southern African populations of *D. melanogaster* and F_{ST} values obtained for *D. melanogaster* populations in other parts of the world.

Locus	G_{ST} (1)	F_{ST} (2)	F_{ST} (3)
<i>Adh</i>	0,1289	0,4906	0,503
<i>Dip-A</i>	0,1396	-	0,054
<i>Gpdh</i>	0,1323	0,1080	0,107
<i>Hex-1</i>	0,1109	-	0,242
<i>Hex-2</i>	0,1365	-	-
<i>Idh-NADP</i>	0,1196	-	0,052
<i>Pgd</i>	0,0327	0,2479	0,249
<i>Pgm</i>	0,0810	0,0762	0,077

(1) Southern Africa, this study ($G_{ST} = 1 - H_S/H_T$).

(2) F_{ST} values for populations in Canada, USA, France, West Africa, Taiwan, Vietnam and Australia (Singh et al. 1982).

(3) F_{ST} values for populations in Canada, Argentina, Sweden, Ukraine, USSR, France, West Africa, Central Africa, Korea, Taiwan, Vietnam and Australia (Choudhary and Singh 1981).

The Southern African populations were drawn from a relatively narrow geographic range as opposed to those used in the comparison studies (Table 6.10). Singh *et al.* (1982) have taken allele frequency statistics from places as diverse as West Africa (6,3°N), to France (46,1 °N), and Choudhary and Singh (1987) obtained allele frequencies from populations in 15 different areas from all over the world.

Examination of the distribution of the *Adh* alleles in various parts of the world reveals, in general, the predominance of a single allele over a continent or over a continuous geographic range within a continent. The fast allele of *Adh* is the predominant allele in Australia (Oakeshott *et al.* 1982a; Singh *et al.* 1982), in Europe (David 1982; Charles-Palabost *et al.* 1985; Anxolabéhère *et al.* 1988a), in Japan on the main islands (Watada *et al.* 1986), in the Peoples' Republic of China (Jiang *et al.* 1989), in Vietnam and Taiwan (Singh *et al.* 1982) and in North Africa (Capy *et al.* 1983). On the other hand, this study[†] together with those of David (1982), Singh *et al.* (1982), and David *et al.* (1986), suggest that the predominant *Adh* allele on the African continent, south of the Sahara, is the slow allele. The slow allele also predominates in the French West Indies (Capy *et al.* 1986). No clear cut pattern of a predominant allele has emerged in North America. Although the slow *Adh* allele dominates

[†]I am assuming that the *Adh*⁻¹⁰⁰ and *Adh*⁰ alleles in this study correspond to the common *Adh*^S and *Adh*^F alleles that occur worldwide.

in east coast populations (Vigue and Johnson 1973), this has not, as yet proved to be the case on the west coast where an even distribution of fast and slow alleles was shown to exist (Singh and Rhomberg 1987a).

David *et al.* (1981) found, using data from various parts of the world that, overall, the *Adh^F* allele increased in frequency with increasing latitude. This was however not a linear increase.

The authors divided sample populations into three latitudinal groups: tropical "Mediterranean" and cool temperate. Whereas, tropical populations (i.e. between 30°N and 30°S) were generally similar, having a low frequency of *Adh^F* (average 15%) and a smooth increase with latitude, the "Mediterranean" collections (from between 30° and 42° latitude) were found to have a much higher average frequency of *Adh^F* (70%) and a broader range of variability for a given latitude. Further north (from 42°N to 49°N) *Adh^F* increased to an overall frequency of 95% and a greater homogeneity over a broad geographic area was observed. These authors plotted a graph of latitude versus *Adh^F* percentage and, by joining the peripheral points with straight line segments, demarcated a region containing all of their sample points. (Fig. 1 - their paper).

The Southern African populations fall into two of the latitudinal regions considered by David *et al.* (1989) to be the tropical and the "Mediterranean" (collections were taken from 17,83°S to 33,96°S). Although most of these populations had Adh^F frequencies falling within the region delineated by David *et al.* (1989), there were six "tropical" populations that fell outside this area. Two of these populations, Gobabeb and Harare, had frequencies of Adh^F which lay above the demarcated region (0,520 and 0,460 respectively) and four of these populations, Swakopmund, Nelspruit, Agatha and Thohoyandou had frequencies of Adh^F which lay below the demarcated region (0,067, 0,060, 0,063 and 0,0 respectively).

6.3.6.2 6-Phosphogluconate dehydrogenase (Pgd)

Table 6.10 gives an F_{ST} value of 0,25 for Pgd in the two studies covering populations throughout the world (Singh *et al.* 1982; Choudhary and Singh 1987). The value of G_{ST} for Pgd in Southern Africa is 0,0327. Out of the 17 populations investigated, 13 were monomorphic for the intermediate allele Pgd^{100} on the 99% criterion and all 17 were monomorphic on the 95% criterion (Table 6.4).

This locus also generally showed a lower level of polymorphism in Central and West African populations (Singh and Rhomberg 1987a). On the other hand Singh and Rhomberg (1987a) found this locus to be highly polymorphic in North America. In the same vein Oakeshott *et al.* (1983a) found frequencies of the fast allele of *Pgd* to range from 0.012 to 1.000 in Australasia. So *Pgd* is interesting in that it appears to be either virtually monomorphic or highly polymorphic over large geographic areas. Possible selective action on *Pgd* has been noted by Oakeshott *et al.* (1983a) in that these authors found the regression of the frequency of the fast allele on latitude to be highly significant. Further Singh and Rhomberg (1987a) found that *Pgd* allele frequencies showed a significant correlation with maximum precipitation but not with latitude or temperature in west coast populations in North America. (Due to the monomorphism of *Pgd* at most collection sites, I could not test this enzyme for correlation with any environmental variable in Southern Africa.) This study has revealed that *Pgd* is virtually monomorphic in the southern tip of the African continent. This, together with its lack of polymorphism in equatorial Africa, suggests that one allele might predominate throughout sub-Saharan Africa. This monomorphism of *Pgd* in African as well as in European populations was commented by David and Capi (1988).

6.3.6.3 sn-Glycerol-3-phosphate dehydrogenase (*Gpdh*)

The G_{ST} value obtained for *Gpdh* in this study was 0.1323 which is close to 0.11, the F_{ST} value obtained for this enzyme in other parts of the world (Table 6.10). These low values indicate that this is not a highly differentiated enzyme, and that this pattern of differentiation is similar in Southern Africa to that in other parts of the world. The interesting aspect in this case is that the high preponderance of the fast allele (assuming homology of *Gpdh*¹⁰⁰ in this study with *Gpdh*^F of other researchers) in Southern African population (frequency = 0.906) is once more repeated in tropical Africa where frequencies of this allele are even higher (eg 0.988 and 1.000 as obtained by David (1982) and Singh *et al.* (1982) respectively). Although high frequencies of the fast allele have been reported in other parts of the world, they are not as high as those in Africa and have been found to be as low as 0.380 and 0.474 (Charles-Palabost *et al.* 1985; Singh *et al.* 1982) in France.

The slow allele (assuming homology of *Gpdh*⁷⁵ with *Gpdh*^S of other researchers) was found to be totally absent in a high proportion of African populations (8 out of 17 in this study, one out of one in West Africa (Singh *et al.* 1982) and two out of six in tropical Africa (David 1982)). However, a high degree of polymorphism for *Gpdh* has been found elsewhere in the world, e.g. in Australia, France, Europe, Canada,

USA, Taiwan, Vietnam and Japan, (information for above gathered from Oakeshott *et al.* 1982a, Girard *et al.* 1977, David 1982, Singh *et al.* 1982, Charles-Palabost *et al.* 1985 and Watada *et al.* 1986).

Thus this study has confirmed that the African continent is characterized by high frequencies of the fast allele of *Gpdh* and a large proportion of monomorphic populations.

6.3.6.4 Other enzymes

Idh-NADP and *Dip*-A were found to have higher G_{ST} values in Southern African populations than elsewhere in the world. G_{ST} values of these two enzymes were more than twice as high as the F_{ST} values obtained by Choudhary and Singh (1987) in populations from around the globe (Table 6.10). This suggests that Southern African populations are more highly differentiated at these two loci than are other populations.

The G_{ST} value of *Hex*-1 was approximately half the F_{ST} value obtained for this locus in the world-wide survey (Choudhary and Singh 1987), while the G_{ST} value for *Pgm* was virtually identical in all three studies (Table 6.10). The *Pgm* locus is interesting in that a survey of the literature indicates that world-wide most populations are polymorphic for three alleles: slow, intermediate and fast. Moreover, the intermediate allele is present in very high frequencies, from 70% to 90% (Kojima *et al.* 1970, Oakeshott 1979; Capi *et al.* 1983, 1986; Singh and Rhomberg 1987a). In particular, the frequency of the intermediate *Pgm* allele in Africa had been given variously

as 0,893 (Capy *et al.* 1986), 0,709 (Singh *et al.* 1982) and 0,944 (Capy *et al.* 1983). In this study, the intermediate *Pgm* allele had a frequency of 0,942 and the *Pgm* locus was polymorphic in 9 out of 17 samples on the 99% criterion and in 6 out of 17 on the 95% criterion.

6.4 Discussion

I have made use of allele frequency statistics to investigate the degree of heterogeneity and the gene diversity pattern within Southern African populations of *D. melanogaster*. The results have also been analysed in relation to gene flow and spatial structure and a clear picture of the Southern African populations of *D. melanogaster* has emerged. The findings and conclusions reached are discussed in detail below.

6.4.1 Microgeographic Variation

Significantly different allele frequencies were detected in two sets of samples that were collected in coastal cities. In each of these cities one of the pair of samples was taken on the beach front and the other approximately 10 km inland. The inland sites were 70 metres (Walmer, Port Elizabeth) and 170 metres (Berea, Durban) above sea level and may represent quite different environments from the sea level sites. It is not clear whether there is a high degree of gene flow over these distances and the observed differences are due to differential selection in the different habitats, or whether the differences are due to random drift in the face of limited gene flow.

The two loci which are mainly responsible for this significant heterogeneity (*Dip-A* in Port Elizabeth and *Adh* in Durban) both showed some variation with an environmental factor (Section 7.3) indicating that some form of selection is possibly in operation. There was no consistent difference however, between inland and coastal samples as might be expected if for example humidity from the sea were acting as a selective agent on these loci over this geographic scale. On the other hand, strong allele frequency differences may appear over a few kilometres because of a high rate of local extinction and recolonization. Any allele frequency differences may be due to founder effects preceding annual population explosions during the summer.

A significant degree of heterogeneity was also detected over loci among/between districts separated by 50 to 300 km in each of the five areas except Northern Transvaal. Small sample sizes may have been responsible for failing to detect any allele frequency heterogeneity in this latter area. The habitats in each of these areas appeared to be relatively homogeneous, except for northern Namibia in which one sample was collected near the sea at Swakopmund and the other about 50 kilometres inland in the Namib Desert at Gobabeb. Swakopmund is cooled by the cold upwelled waters of the Benguela Current which produces a fog bank along the coast, especially in summer. Gobabeb, in contrast, lies in the Namib Desert and experiences extreme diurnal fluctuations in temperature.

The only locus which showed strong differences between the locations was *Adh*. This locus is apparently under selection in many areas (see reviews by van Delden 1983 and Chambers 1988), and exhibits a high degree of polymorphism worldwide. Apart from *Adh*, the samples from both Swakopmund and Gobabeb showed reduced levels of genetic variation in comparison with other locations and this may be the result of frequent bottle necks in population size in response to the harsh environment.

6.4.2 Macrogeographic Variation

On a macrogeographic level the electrophoretic results obtained in this study were analysed for patterns of gene flow and selection gradients as well as for heterogeneity and gene diversity. A pronounced increase in genetic heterogeneity between locations separated by distances greater than 50 kilometres was observed. Both the D_{xy} statistics of the gene diversity analysis and the standardized measures of heterogeneity of the contingency-table analyses were roughly similar and indicated that the greatest divergence occurred between regions on a geographic scale of 500 to 3000 kilometres. The degrees of divergence among areas on a scale of 400 to 800 kilometres were only slightly less than that among regions. I shall examine this pattern of differentiation in terms of various models of gene flow and population structure.

6.4.2.1 Gene flow measurements

Measurements of gene flow in many of the *Drosophila* species have been consistently high. Slatkin (1985) estimated values of Nm based on the research of Ayala *et al.* (1974) for *D. willistoni* and of Singh (1979) for *D. pseudoobscura* as 9.9 and 1.0 respectively. Singh and Rhomberg (1987a) estimated Nm values of *D. melanogaster* populations in the USA, East Asia and globally to vary between 1.1 and 2.7. Migration estimates ranged from 1.5 to 6.9 individuals per generation and the graphic gene-flow analysis of Slatkin (1981) also indicated a high rate of migration.

6.4.2.2 Sampling techniques

The high rate of gene flow obtained does not prevent a significant amount of heterogeneity arising between subpopulations (examples are David 1982, Singh *et al.* 1982, Kusakabe and Mukai 1984, Charles-Palabost *et al.* 1985, Watada *et al.* 1986, Choudhary and Singh 1987, Singh and Rhomberg 1987a,b and this study). Yet it is generally noted that migration between populations of the order of one individual per generation is sufficient gene flow to ensure that only a minimal amount of differentiation exists between populations. Spieth (1974) showed that, if $Nm = 1$, then the genetic identity of two genes in different populations will be similar to that of two genes in the same population unless "mutation is so large as to produce substantial differentiation within the individual populations".

Many authors (e.g. Wright 1969) are aware of this apparent dichotomy i.e. statistically significant patterns of divergence and yet low F_{ST} values and high migration rates. The problem of inferring rates of genetic exchange among subpopulations on the basis of allele frequency data has been addressed by Allendorf and Phelps (1981). They noted, using simulation studies, that significant divergence among subpopulations was detected in populations with an mN value of 25. They showed that this effect (of significant heterogeneity with very little or no isolation) was particularly exaggerated when the progeny of a small number of (non-isolated) reproducing adults was sampled. In their study, Allendorf and Phelps (1981) were referring to fish populations where the samples are generally fry which may have been spawned in a stream separated from the parental mating area such as a lake. Nonetheless, the observations made can be applied to *Drosophila* populations in that these populations have a dynamic metapopulation structure that is characterized by a large potential for high rates of population growth, high rates of local population extinction and frequent recolonization. Thus flies sampled may well emanate from a few transient progenitors in much the same way as the fry which are sampled, are the progeny of a few spawning adults. These founder populations may be quite different from ancestral populations as well as from adjacent populations arising from the same ancestral population. This would lead to a genetic mosaic of populations on a relatively small spatial scale.

In this study I showed that collection sites ten kilometres apart displayed considerable heterogeneity. The results of other studies have shown that significant differences can appear over distances of only a few hundred metres (David and Capi (1988), and references therein).

The usual inference upon finding significantly different allele frequencies is reproductive isolation. In this case, the differences may arise as a consequence of metapopulation structure and not because of divergence following reproductive isolation. The general problem in studying genetic structure of *Drosophila* populations arises in choosing the appropriate spatial scale for sampling. The sampling of *Drosophila* populations for genetic analysis is labour intensive and usually only a few samples are collected which are then taken to be representative of the populations in an area.

In order to prevent erroneous conclusions being made about the dynamics of a constantly changing system, I suggest that the best way to sample *Drosophila* populations is to use a fine scale level of sampling i.e. to collect samples of *Drosophila* at five to ten kilometre intervals over the required area, and to repeat these collections over a period of time in order to stabilize spatial and temporal differences.

6.4.2.3 Modelling considerations

The estimates of Nm obtained using either Wright's fixation index or the private alleles model of Slatkin, both assume Wright's island model conditions i.e. that immigrants arrive randomly from the rest of the population and that the rate

of gene flow is uniform throughout the area under consideration. This is biologically unrealistic, particularly in places such as Africa where human populations tend to be distributed in patches with large distances between them. In view of the commensal nature of *D. melanogaster*, and of the accompanying assumption that the greater the human traffic the greater the movement of *D. melanogaster*, one would not expect an even migration flow between samples which are being taken from so large a geographic areas as in this study. Since immigrants are more likely to come from subpopulations close by, a better model to use would be the stepping stone model of Kimura and Weiss (1964). Slatkin (1985) found that, when simulating a one dimensional stepping stone model of migration, a higher value of $\overline{p(i)}$ (the conditional average frequency of private alleles) was obtained $\overline{p(i)} = 0.109$ for the stepping stone model as opposed to 0.054 for the island model - all other conditions were constant). This implies that the stepping stone model would produce a lower estimate of gene flow than the island model.

An assumption used in many theoretical models is that the sample populations are in a state of equilibrium both genetically and demographically. The theoretical effect on gene flow of demographically unstable populations has been investigated by Slatkin (1977) who used a "metapopulation" infinite alleles model in which migrants come, not from a single external source but from within the collection of populations. He

in fact showed, that extinction and recolonization serve to augment the level of gene flow in that the amount of differentiation in local populations was reduced. This appears to contradict the findings of Allendorf and Phelps (1981) (although the latter did not concern themselves with extinctions and recolonizations *per se*). Data obtained by means of a fine scale level of sampling could be used to test a metapopulation model in which extinctions due to demographic and environmental stochasticity together with catastrophe and subsequent recolonization are an ongoing process.

6.4.3 Origin of Southern African Populations of *D. melanogaster*

The origin of *D. melanogaster* in Southern Africa is not certain. The *melanogaster-simulans* evolutionary split is thought to have occurred about 2-3 million years ago (Stephens and Nei 1985; Cariou 1987; Lachaise *et al.* 1987) in equatorial Africa at the time of major tectonic activity in the Rift and the subsequent aridification of East Africa (Lachaise *et al.* 1987). Modern world populations of *D. melanogaster* have probably dispersed from these ancestral populations in West and Central Africa. It is hypothesized (e.g. Lachaise *et al.* 1987) that, during the late Pleistocene or Holocene periods, *D. melanogaster* was able to spread northward across the Sahara due to the favourable wet climatic conditions which occurred repeatedly during this period.

It has been suggested by David and Capi (1983) that due to the "continuity of favourable habitats, Southern African populations might also be included (with these ancestral populations)." Natural dispersal would apparently only be possible along the eastern edge of the subcontinent as the Namib and Kalahari Deserts in Southwestern and South Central Africa formed a barrier to migration as did the Sahara Desert in North Africa.

From the tropical origin of most South African butterflies, some migration southward through the middle part of the Zambezi Valley, appears to have taken place at some time during the Pleistocene (van Zinderen Bakker 1967). Consequently, Lachaise *et al.* (1987) hypothesized that "a southward migration of *D. melanogaster* populations from equatorial Africa towards South Africa might have occurred during temporal windows of the late Pleistocene". Alternatively present-day populations may have originated at a later time by dispersal from West and Central Africa or possibly by colonization from Europe and Asia through human immigration.

The first conclusion is that Southern African populations of *D. melanogaster* represent a single taxonomic unit. There did not appear to be large population differences among Southern African populations that would suggest the existence of, for instance, both European and Central African stocks in Southern Africa. Since the survey included only eight loci

and since I was uncertain about the homologies of alleles with other studies, I was not able to calculate genetic distances with other populations. The conclusion is based, therefore, on allele-frequency similarities and comparisons of levels of polymorphism with other geographic areas. Nevertheless, the tests for allele-frequency heterogeneity and the G_{ST} values, which averaged 12.2%, indicated that there was a significant degree of genetic differentiation amongst Southern African populations. This level of differentiation has also been found among populations within other areas in Africa and elsewhere in the world and I do not interpret it as reflecting differences among populations with different origins.

The second conclusion is that Southern African populations are closely related to West and Central African populations. Although there tends to be a preponderance of a single allele for some loci worldwide, there are several loci which show differentiation between continents. In section 6.3.6.1 mention is made of the fact that for alcohol dehydrogenase the "slow" allele (Adh^{100}) was found to be the most common in sub-Saharan Africa whereas the "fast" allele predominates elsewhere in the world. Further Adh^{120} appears to be homologous with the rare ultra slow Adh^{US} allele found in the Congo (David 1978). The similarity between the Southern African populations and those of tropical Africa insofar as the high frequency of the "fast" $Gpdh$ allele is concerned was noted in section 6.3.6.3. The closeness of these allele frequencies and levels of genetic variation of Southern African population with those in the rest of sub-Saharan Africa indicate a close historical affinity between the two.

CHAPTER SEVEN

7. Clinal Patterns and Other Correlations

7.1. Introduction

A strong argument in favour of adaptive selection is the occurrence of clinal variation in genetic traits such as allozyme (or inversion) frequencies or in more complex characteristics such as quantitative morphological variations or behavioural patterns. When these clines occur they are often latitudinal, since environmental gradients which may act as a forcing mechanism are associated with decreasing pole-ward solar insolation which causes variation in parameters such as temperature, rainfall, length of day etc. The argument that selection is responsible for these latitudinal clines is strengthened when the same variation with latitude is noted on different continents in populations with different evolutionary histories.

Worldwide latitudinal clines in allele frequencies have been reported for various metabolic enzymes in *Drosophila melanogaster*. For instance, clines have been found for *Adh*, Aldehyde oxidase (*AO*), Carbonic anhydrase (*Ca-2*), *Eat-C*, *Eat-6*, Leucine aminopeptidase (*Lap-D*), *Gpdh*, Octonol dehydrogenase (*Odh*), Glucose-6-phosphate dehydrogenase (*G6pd*) and *Pgd* (Johnson and Schaffer 1973; Vigue and Johnson 1973; Smith *et al.* 1978; Oakeshott *et al.* 1981, 1982a, 1983a,b; Singh *et al.* 1982; Singh and Rhomberg 1987a) to name a few examples. Other environmental gradients associated with longitude, altitude, rainfall, temperature and food availability have also been correlated with allozyme variation.

7.1.1 Alcohol Dehydrogenase

In trying to detect some correlation between environmental variables and allelic patterns in individual genes, it makes sense to examine those enzymes which act directly on an external substrate and whose allozymes differ biochemically. The variable being tested is then the substrate.

Further, enzymes which use substrates originating from the environment appear to be more variable because of qualitative and quantitative variation of natural substrates, than enzymes which use substrates arising as products of enzymes in a metabolic pathway (Gillespie and Kojima 1968; Kojima *et al.* 1970). In this respect Alcohol dehydrogenase has been one of the most extensively investigated enzymes.

7.1.1.1 Ethanol tolerance

Drosophila melanogaster is nutritionally dependent on yeasts and bacteria (Sang 1956) and breeds in wineries as well as in orchards and rubbish dumps, where microflora are present in decaying fruits and vegetables. Since ethanol occurs naturally in fruits and is also a product of fermentation, this primary alcohol is found (often in high concentrations) in *Drosophila* breeding and feeding environments. ADH catalyzes the oxidation of alcohols to aldehydes and ketones, and so it has generally been assumed that its main function in *D. melanogaster* is to aid in the detoxification of environmental alcohols, particularly ethanol. Direct evidence for this biological function was obtained

when flies homozygous for *Adh* null alleles (i.e. lacking ADH activity) died within a few hours on media containing 8% ethanol (Grell *et al.* 1968). The association between the polymorphic *Adh* gene and ethanol tolerance has consequently been extensively investigated.

Homozygotes of the *Adh^F* allele have greater *in vitro* ADH activity than homozygotes of *Adh^S*, the other common allele, while heterozygotes often show intermediate levels of activity (Gibson 1970; Vigue and Johnson 1973; Day *et al.* 1974; Lewis and Gibson 1978; McDonald *et al.* 1980; Anderson and Gibson 1985; Barbancho *et al.* 1987). These results seemed to indicate that the fast allele would be selected for in a high ethanol environment and that *Adh^F* homozygotes would have an advantage in the detoxification of alcohols.

Indeed, in some experiments an increase in the frequency of the *Adh^F* allele was observed when polymorphic laboratory populations were maintained on 6 to 20% ethanol-supplemented media for several generations (Gibson 1970; Bijlsma-Meeles and van Delden 1974; Briscoe *et al.* 1975; van Delden *et al.* 1978; Cavener and Clegg 1978; Vigue *et al.* 1982). Other researchers, however, found no evidence of an increase in the fast allele with an increase in alcohol concentration (Gibson *et al.* 1979; Oakeshott *et al.* 1984). Barbancho *et al.* (1987) examined ethanol tolerance in laboratory populations and concluded that the adaptation of *D. melanogaster* to ethanol-containing food could be accomplished without significant changes of the ADH activity in adult flies.

Conflicting results have also been obtained in comparisons of winery and non-winery populations. Higher frequencies of the fast allele were present in wineries in Spain (Briscoe *et al.* 1975; Munoz-Serrano *et al.* 1985) and in Canada (Hickey and McLean 1980). On the other hand, the frequencies of the *Adh^F* allele in populations inside wineries in California (Marks *et al.* 1980), and in Australia (McKenzie and Parsons 1974; McKenzie and McKechnie 1978; Gibson *et al.* 1981; Gibson and Wilks 1988) were not significantly different from those in samples taken from outside of the wine cellars. Comparisons were made on populations from inside and outside wineries because of the assumption (implicit or explicit) that the ethanol content inside was higher than that outside. However Gibson *et al.* (1981) found that ethanol levels in *Drosophila* breeding sites in seepages of unfortified wine inside wineries were similar to those in other breeding places, such as decaying fruits and vegetables. Observed frequency differences inside and outside of wineries could therefore be due to some factor other than variation in ethanol concentration.

The difference between the *in vitro* results and those obtained for wild flies demonstrates the fallacy of extending the results of *in vitro* experiments to laboratory or natural populations. Middleton and Kacser (1983) showed that the difference in ethanol flux between fast and slow *Adh* homozygotes was not significant. They concluded that control of ethanol flux is shared among all of the relevant enzymes in the metabolic pathway and they warned against predicting *in vivo* consequences of genotypic

change on the basis of *in vitro* enzymatic activities. (On the other hand, Heinstra *et al.* (1987) showed that differences in ADH activity in larvae subjected to different feeding conditions could cause flux changes.)

Another point is that laboratory studies which use levels of ethanol greater than about 5% may not have any relevance to natural populations. There was no difference observed in the mortality rates of flies homozygous for the *Adh^S* and *Adh^F* alleles maintained on media containing 3% ethanol (Ainsley and Kitto 1975). This is the probable level of ethanol present in the environment of non-winery flies. The ethanol concentration in fermenting bananas varies from 0.03 to 4% (Ainsley and Kitto 1975; Gibson *et al.* 1981) and it is probable that *D. melanogaster* encounters a maximum of 4% ethanol at its breeding sites (McKechnie and Morgan 1982; Oakeshott *et al.* 1982b). These studies which showed an increase in the frequency of the *Adh^F* allele in laboratory populations, were looking at unnaturally high ethanol levels in the range of 6 to 20%. It is therefore unlikely that the concentration of environmental ethanol acts as a strong selective force in natural populations.

7.1.1.2 Other Alcohols

Although ethanol is the predominant alcohol in habitats frequented by *D. melanogaster*, it is not the only one. The kinds and concentrations of alcohols in nature vary with the type of decomposing materials and with the native microflora whose