



National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service

Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

PATTERNS OF MOLECULAR EVOLUTION AT THE
AMYLASE LOCUS IN DROSOPHILA.

by

Sumaia Abukashawa

A thesis

presented to the University of Ottawa
in partial fulfillment of the
requirement for the degree of
Doctor of Philosophy
in the Department of Biology





National Library
of Canada

Bibliothèque nationale
du Canada

Canadian Theses Service Service des thèses canadiennes

Ottawa, Canada
K1A 0N4

The author has granted an irrevocable non-exclusive licence allowing the National Library of Canada to reproduce, loan, distribute or sell copies of his/her thesis by any means and in any form or format, making this thesis available to interested persons.

The author retains ownership of the copyright in his/her thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without his/her permission.

L'auteur a accordé une licence irrévocable et non exclusive permettant à la Bibliothèque nationale du Canada de reproduire, prêter, distribuer ou vendre des copies de sa thèse de quelque manière et sous quelque forme que ce soit pour mettre des exemplaires de cette thèse à la disposition des personnes intéressées.

L'auteur conserve la propriété du droit d'auteur qui protège sa thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

ISBN 0-315-60081-0

Canada



UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ACKNOWLEDGEMENTS

I would like to acknowledge with gratitude the support, encouragement and the invaluable supervision of Dr. Donal Hickey. He corrected the draft of this thesis in an untiring manner and with enthusiasm. His academic remarks and rich experience were a great help.

I would also like to thank members of my committee, Dr. L. Bonen, Dr. G. Carmody, Dr. P. Moreau and Dr. P. Lau for their valuable comments, advice and help.

Thanks are due to Dr. C. Langely and Dr. M. Kreitman who gave me the opportunity to learn the details of the 4-cutter technique at NIEH, N. Carolina, USA.

I am especially grateful to my fellow graduate students, Allen Good, Bernhard Benkel, Ada Loverre, Babi Magoulas, Veronique Payant and Pam Simons for contributing to the rich intellectual experience I have enjoyed in Dr. Hickey's lab.

I am indebted to Mrs. J. Leibovitz for helping with the Harvard Graphics, Mr. P. Brunon for the fine photography and Mr. Yves Genest for the technical help.

I reserve a word of appreciation to my friend Kaarina Benkel and my husband Abdelrahman, whose patience, concern and full support made it easy for me to complete this work

through, and to my children Ibrahim and Atika: with them it was a major operation to write this thesis; without them it could have been an impossible task.

I am also indebted to the University of Khartoum, Sudan, for granting me a Scholarship and for the University of Ottawa, Canada, for granting me the chance to to do this work.

I am indebted to many others whose names do not appear in this acknowledgement.

ABSTRACT

Genes encoding the starch-degrading enzyme, alpha-amylase, are found in all major groups of animals, plants and microbes. In this thesis, amylase-coding sequences have been chosen as a model system to investigate the patterns of molecular evolution in an enzyme-coding gene. Previous phylogenetic comparisons of amylase-coding sequences have shown high levels of primary sequence conservation over long evolutionary periods. The studies described here have concentrated on the evolution of these genes within the genus Drosophila. I have studied patterns of genetic variation within populations of a single species, Drosophila melanogaster, using two techniques: (i) allozyme variation and (ii) restriction fragment length polymorphisms (RFLPs). In order to get more precise information on the intraspecific patterns of variation, I have also isolated and partially sequenced the amylase genes from a wildtype strain of D. melanogaster. This sequence was compared to the known sequences which have already been described for laboratory strains. The work was extended to interspecific comparisons by studying amylase sequences from species that were closely-related to D. melanogaster (in this case D. erecta), and also a distantly-related species (D. virilis). This involved the

isolation and sequencing of the amylase gene from a D. virilis genomic library. The results revealed several interesting patterns in the evolution of: (i) the primary gene sequences (e.g., gene conversion and codon bias), (ii) gene structure (e.g., changes in intron frequency and location) and (iii) gene organization (e.g., variation in the number of gene copies). These results, concerning both short-term and long-term patterns of molecular evolution within the genus Drosophila, are discussed in the context of what is currently known about patterns of molecular evolution in general, and about the molecular evolution of amylases in particular.

RESUME

Les gènes qui codent pour l'alpha-amylase, l'enzyme responsable de la dégradation de l'amidon, se retrouvent dans plusieurs groupes d'animaux, de plantes, et de microbes. Dans cette thèse, les séquences codant l'amylase ont servi de modèle dans l'examen des patterns de l'évolution moléculaire d'un gène codant pour une enzyme. Les comparaisons phylogénétiques antérieures entre des séquences codant pour l'amylase ont révélé une conservation marquée des séquences primaires pendant des longues périodes au cours de l'évolution. Les recherches décrites dans le présent ouvrage ont porté essentiellement sur l'évolution de ces gènes à l'intérieur du genus Drosophila. Nous avons examiné la variation génétique à l'intérieur de diverses populations appartenant à une espèce unique, Drosophila melanogaster. Pour arriver à cette fin, deux techniques ont été utilisées: (i) la variation des allozymes et (ii) le polymorphisme des fragments de restriction (RFLP). Afin d'obtenir des données plus précises au sujet des patterns de variation à l'intérieur d'une espèce, le gène de l'amylase d'une souche sauvage de D. melanogaster fut isolé et séquencé partiellement. De plus, cette dernière séquence a été comparée à des séquences connues de souches de laboratoire. La portée de cette analyse a été élargie en incluant des

comparaisons entre espèces; des séquences d'amylase appartenant à des espèces qui sont étroitement apparentées avec D. melanogaster (dans ce cas, D. erecta) ont été examinées, ainsi que des séquences appartenant à une espèce qui est relativement peu apparentée (D. virilis). A cette fin, on a isolé et séquencé le gène pour l'amylase qu'on a obtenu d'une librairie génomique de D. virilis. Les résultats obtenus ont révélé plusieurs patterns intéressants dans l'évolution de: (i) les séquences primaires des gènes (e.g. la conversion du gène et le choix de codon); (ii) la structure du gène (e.g., changement dans les fréquences des introns et dans leur localisation); et (iii) l'organisation du gène (e.g., la variation du nombre de copies du gène). Les résultats obtenus, concernant les patterns à court et à long terme de l'évolution moléculaire du genus Drosophila, sont discutés en fonction des connaissances actuelles des patterns généraux de l'évolution moléculaire et plus précisément, de ce qui est connu de l'évolution moléculaire des amylases.

LIST OF FIGURES

	Page
Figure 2.1 Amylase enzyme electrophoretic patterns of several <u>D.melanogaster</u> strains	31
Figure 2.2 Amylase enzyme activity from isogenic <u>D.melanogaster</u> strains with different enzyme and activity patterns	32
Figure 2.3 Restriction digest of total fly DNA probed with pOR-m7	38
Figure 2.4 Restriction digest of total fly DNA probed with pCS-1.4 showing restriction site polymorphisms	39
Figure 2.5 Hind III digest of total fly DNA from different <u>D.melanogaster</u> strains with length polymorphisms	40
Figure 2.6 General organization of the autosomal amylase locus in 50 strains with RFLPs	43
Figure 2.7 RFLPs due to insertions/deletions in the amylase region of 50 <u>D.melanogaster</u> strains	44
Figure 3.1 Lambda genomic clones from makokou ^{4,6} probed with pCS1.4	65
Figure 3.2 Genomic clones from Makokou ^{4,6} probed with pIR-m7	66

Figure 3.3	Restriction map of Makokou ^{4,6} amylase genes.	67
Figure 3.4	Nucleotide sequence of Makokou ^{4,6} proximal and distal amylase genes compared to the proximal amylase gene of OR-R	71
Figure 4.1	Comparison of duplicated amylase coding sequences in <u>D.melanogaster</u>	80
Figure 4.2	Comparison of alpha-amylase sequences of <u>D.erecta</u> with the proximal of <u>D.melanogaster</u>	83
Figure 4.3	Models of evolution of mammalian salivary amylase genes	85
Figure 5.1	Screening of <u>D.virilis</u> genomic library with <u>D.melanogaster</u> cDNA sequences	96
Figure 5.2	Restriction map for <u>D.virilis</u> amylase-containing sequences	97
Figure 5.3	<u>D.virilis</u> genomic clones containing amylase gene	98
Figure 5.4	Restriction pattern of total fly DNA probed for amylase sequences	99
Figure 5.5	Sequencing strategy for <u>D.virilis</u> amylase gene	100
Figure 5.6	Complete sequence of the chromosomal segment encompassing the amylase gene of <u>D.virilis</u>	101
Figure 5.7	Comparison of amylase nucleotide sequence of <u>D.virilis</u> to that of <u>D.pseudoobscura</u> and <u>D.melanogaster</u>	101

Figure 5.8	Matrix analysis of interspecies sequence conservation at the amylase locus	112
Figure 5.9	Comparison between the introns of <u>D.virilis</u> and <u>D.pseudoobscura</u>	114
Figure 5.10	Amino acid sequence comparison of <u>D.virilis</u> , <u>D.pseudoobscura</u> and <u>D.melanogaster</u>	121

LIST OF TABLES

	Page
Table 1.1 Amino acid sequence of conserved regions of amylase protein in different organisms	3
Table 2.1 Enzyme polymorphisms and restriction polymorphisms in isogenic strains of <u>D.melanogaster</u> from different localities	33
Table 2.2 Polymorphic restriction sites scored in the amylase region of <u>D.melanogaster</u>	45
Table 2.3 Non-random association among amylase region restriction sites and enzyme polymorphisms	47
Table 5.1 Frequency of nucleotides in the first, second and third position codons used by amylase genes of <u>Drosophila</u>	118
Table 5.2 Distribution of codons in the amylase genes of <u>Drosophila</u>	127

CONTENTS

ACKNOWLEDGEMENTS	i
ABSTRACT	ii
RESUME	iv
LIST OF FIGURES	vi
LIST OF TABLES	ix

CHAPTER	PAGE
1. LITERATURE REVIEW	
Introduction	1
Prokaryotic alpha-amylase	4
Amylase evolution in the vertebrates	6
Insect amylases	8
The <u>Drosophila</u> amylases	10
Amylase polymorphisms in <u>Drosophila</u>	10
Linkage Studies and Chromosome Assignments	12
Amylase gene duplication	14
Regulation of amylase synthesis.....	17
Conclusion	22
2. ALLOZYME AND RESTRICTION FRAGMENT LENGTH POLYMORPHISMS	
Introduction	24
Materials and Methods	26

Results	30
Discussion	50
3. INTRASPECIFIC DNA SEQUENCE COMPARISON	
Introduction	55
Materials and Methods	57
Results	64
Discussion	74
4. CONCERTED EVOLUTION OF THE DUPLICATED AMYLASE CODING SEQUENCES	
Introduction	77
Materials and Methods	78
Results	79
Discussion	81
5. LONG TERM INTERSPECIFIC SEQUENCE COMPARISONS	
Introduction	88
Materials and Methods	92
Results	94
Discussion	128
6. SUMMARY AND CONCLUSIONS.....	
	137

REFERENCES	143
APPENDIX 1.1.....	172
APPENDIX 2.1	172
APPENDIX 2.2	179
APPENDIX 2.3.....	182
APPENDIX 5.1.....	183

CHAPTER 1

LITERATURE REVIEW

Introduction:

Alpha-amylase genes are expressed within all the major groups of prokaryotic and eukaryotic organisms. Discovered by Leuchs (1831) and originally called "diastase", alpha-amylase was defined by Kuhn (1925) as a carbohydrate-degrading protein which yields saccharide products with the alpha configuration. The enzyme is an endoamylase. It does not require free ends of amylose chains for activity, but rather is capable of hydrolysing internal alpha-1,4-glucoside bonds. Alpha amylase (alpha-1,4-glucan, 4-glucanohydrolase; EC 3.2.1.1) is ubiquitous among higher animals which, to one extent or another, process exogenous starch and/or glycogen as part of their diets (for a detailed review see Lowenstein 1987).

The early discovery of the amylase enzyme with a major role in the digestive process, coupled with the relative ease

with which the amylase can be purified and assayed, led to its use in the original studies of enzyme kinetics as recorded by Henri (1902) and Michaelis and Pechstein (1914). The ready availability of specimens with amylase activity has prompted extensive analyses in a wide variety of organisms: for example, bacteria (Robyt and Ackerman 1973), plants (Dube and Nordin 1961; Stein et al. 1960; Thoma et al. 1970), small mammals (Berk et al. 1963; Kaplan et al. 1973; Malaciniski and Rutter 1969; Nielson 1969; Nielson and Sick 1975), cattle, horses, swine (Hasselholt et al. 1966), primates (Duane et al. 1971; 1972), and humans (Karn et al. 1975).

The comparison of amylase amino acid sequences from animal, plant and bacterial origin identified two highly conserved regions (Toda et al. 1982; Meisler and Gumucio 1986). The fact that these sequences are conserved in amylases of animals, plants and bacteria, and are located at comparable positions within all of these proteins, is strongly suggestive of a role in catalysis or substrate binding. It is somewhat surprising that the sequences of the mammalian active sites are more similar to the bacterial sequence than to the plant and fungal sequences (Table 1.1).

Table 1.1. Amino acid sequences of two highly conserved portions of the amylase protein in different organisms (from Meisler and Gumucio 1986).

Source	Sequence	Residues	homology
Conserved sequence I			
1. Mouse pancreas	V D A V I N H	110-116	7/7
2. Mouse parotid	V D A V I N H	110-116	7/7
3. Rat Pancreas	V D A V I N H	105-111	7/7
4. Porcine pancreas	V D A V I N H	95-102	7/7
5. Human pancreas	V D A V I N H	110-116	7/7
6. Human parotid	V D A V I N H	110-116	7/7
7. <u>B.subtilis</u>	V D A V I N H	137-143	7/7
8. Barley	A D I V I N H	111-117	5/7
9. Aspergillus	A D V V A N H	116-122	5/7
Conserved sequence II			
1. Mouse pancreas	G F R L D A A K H M	205-214	10/10
2. Mouse parotid	G F R L D A S K H M	208-217	9/10
3. Rat pancreas	G F R L D A A K H M	200-209	10/10
4. Porcine pancreas	G F R I D A S K H M	193-202	8/10
5. Human pancreas	G F R L D A S K H M	208-217	10/10
6. Human parotid	G F R I D A S K H M	208-217	10/10
7. <u>B.subtilis</u>	G F R F D A A K H I	213-222	8/10
8. Barley	D G A L D W G P H M	138-147	4/10
9. Aspergillus	G L R I D T V K H V	202-211	5/10

A=Alanine; D=Aspartic Acid; F=Phenylalanine; G=Glycine;
H=histidine
I=Isoleucine; K=Lysine; L=Leucine; M=Methionine; N=Asparagine;
P=Proline; R=Arginine; S=Serine; T=Threonine; V=Valine;
W=Tryptophan; Y=Tyrosine
* Homology is with relation to mouse pancreatic amylase.

These conserved sequences suggest that bacterial, plant and mammalian amylases are direct descendants of a common progenitor molecule whose active site has been retained during the course of evolution, and the similarities of the molecular weight of essentially all amylases that have been studied (Karn and Malacinski 1978) is consistent with this idea. The alternative explanation would involve evolutionary convergence of different ancestral sequences to acquire these short conserved elements. Recent molecular analyses provide strong evidence in favor of the monophyletic origin of all amylases; an example of such evidence is the cross-hybridization of the mouse amylase cDNA with the amylase gene of Drosophila (Levy et al. 1986; Gemmill et al. 1986; Benkel et al., 1987). The widespread distribution of amylase in unicellular and multicellular organisms provides evidence that the ability to digest exogenous carbohydrates has been important throughout evolution.

PROKARYOTIC ALPHA-AMYLASES

Previous comparisons of alpha-amylase amino acid sequences (Nakajima et al. 1986; Rogers 1985) recognized four conserved regions positioned at similar intervals along the length of each protein, two of which are shown above (Table 1.1). It has been suggested that at least three of these are

involved in substrate binding or contribute components of the active site. Long et al. (1987) compared amino acid sequences from different organisms with their *Streptomyces* sequence. While little overall similarity could be detected between the streptomyces (*Streptomyces limosus*) enzyme and alpha-amylases derived from *Bacillus subtilis* (Yang et al. 1983), *Aspergillus oryzae* (Toda et al. 1983), *B.amyloliquefaciens* (Takkinem et al. 1983) and barley (Rogers and Milliman 1983), substantial similarity of the streptomyces sequence was observed to the alpha-amylases of *Mus musculus* (mouse pancreatic alpha-amylase, Hagenbuechle et al. 1980) and *Drosophila melanogaster* (Boer and Hickey 1986).

Why the *Streptomyces limosus* enzyme should show a high degree of amino acid sequence identity to alpha amylases of mammalian and invertebrate origin, while the remaining alpha-amylases differ much more extensively, is not clear. The identity observed between the streptomyces enzyme and the alpha-amylase of *D.melanogaster* extends more or less throughout the sequence, and it would seem unlikely, given the divergence of the other sequences, that functional constraints alone (e.g., residues at the catalytic site) are responsible for this degree of extensive sequence conservation. Long et al. (1987) suggested a common ancestry for streptomyces, mammalian and invertebrate alpha-amylases and separate evolutionary origins for the differing alpha-

amylases. In other words, they propose that the streptomycete, mammalian and invertebrate alpha-amylases may have originated from a common and more recent evolutionary branch point than this group and the *Bacillus* amylases. This argument implies that amylase genes had already duplicated and begun to diverge in their sequences before the divergence of prokaryotes and eukaryotes. As will be discussed below, there are also examples of much more recent duplications in amylase genes.

AMYLASE EVOLUTION IN THE VERTEBRATES

Pancreatic amylase is ubiquitous among vertebrate species, salivary amylase is of more limited distribution occurring chiefly among mammalian groups, and is therefore thought to be of more recent evolutionary origin (Meisler and Gumucio, 1986). Differences in antigenicity and in tryptic peptide maps provided the first indication that salivary and pancreatic amylases are actually encoded by distinct genes. Mammalian pancreatic amylase appears to be descended from the enzyme present in the dispersed secretory cells of the intestine of protochordates and primitive vertebrates. In teleost fish these cells are collected into a discrete organ, the pancreas, which persists as the major amylase-containing organ in all higher vertebrates (Barrington 1957). In

contrast, salivary glands first appear with the terrestrial amphibians and reptiles (Romer 1962). Amylase is present in the salivary secretions of modern amphibians and reptiles (Karn and Malacinski 1987). The structure of the mammalian salivary amylase gene suggests that it evolved by duplication of preexisting pancreatic genes, coupled with the acquisition of a new distal promoter with tissue specificity for the parotid gland (Meisler and Gumucio 1986).

In some species, salivary glands also produce other digestive enzymes commonly found in the pancreas, including ribonuclease and phospholipase A2. The dual expression of these digestive enzymes in pancreas and salivary gland raises the intriguing possibility that a distal parotid promoter, related to that of the amylase gene, may be a common element in these genes.

In the mouse, low levels of the "salivary" amylase are also expressed in the liver (Schibler et al., 1983). There is no evidence that the distal liver promoter associated with the mouse salivary amylase gene was acquired earlier than the parotid promoter. No species has been identified in which amylase is expressed in liver but not in a salivary gland. Nor is there any species in which liver amylase production is greater than the very low level seen in mouse liver. Thus, the expression of amylase in mouse liver is probably of recent evolutionary origin and, because of the very low

levels of expression, may be of no physiological significance.

Another intriguing aspect of the evolution of the amylase gene family in vertebrates is the species variation in the type of salivary gland which produces amylase. In mice and voles, parotid glands produce amylase but submandibular and submaxillary glands do not (Hjorth 1979). However, guinea pig submandibular glands and rat submaxillary glands do produce amylase (Bloom *et al.* 1975; Shear *et al.* 1973). In man, amylase has been localized histochemically in parotid, submaxillary and submandibular glands (Korsurd and Brandtzaeg 1982). The genetic and molecular basis for this variation is as yet unexplained. Among mammalian species, production of amylase by salivary glands has only been observed within primates, rodents and lagomorphs (Karn and Malacinski 1978). Within these orders, there are a few species lacking salivary amylase, including Chinese hamster (Dawson and Huang 1981) and squirrel monkey (McGeachin and Akin 1982).

INSECT AMYLASES

Although the majority of the studies involving multicellular animals have concentrated on mammalian amylases, a broader comparative study of alpha amylases within the metazoa is being initiated. Amylase-coding

sequences have been cloned and sequenced from Drosophila (Order: Diptera) (Gemmill et al. 1985; Benkel et al. 1987, Boer and Hickey 1986, Brown et al 1989), from the flour beetle, Tribolium castaneum (Order: Coleoptera) (Hickey et al. 1987, Hickey, Genest, Abukashawa and Benkel in preparation), and from the cricket Achaeta domesticus (Order: Orthoptera) (Hickey et al. unpublished results). The predicted amino acid sequences of the protein products are found to be identical at more than 60% of the residue positions among the three distantly related insect species, an indication of homology of the coding genes of these insects.

A study of the adaptive changes that have occurred in the pH dependence of amylase enzyme activity was done by Hickey et al. (1989) using amylase genes from different insect groups. They found that the pH optima for activity of Tribolium, Drosophila and the silk worm Bombyx amylases is 5.0, 7.5 and 10.5 respectively. This wide variation in pH optima within the Class Insecta is explained by the correlation with a difference in the pH of the lumen of the midgut. The relative acidity of the midgut of Tribolium compared with the basic lumen of Bombyx is thought to reflect the divergence in feeding habits and habitats between the two insect groups (Hickey et al. 1989).

THE DROSOPHILA AMYLASES

Drosophila alpha-amylase is a monomeric enzyme with a molecular weight of 54,500 (Doane et al. 1975), and isoelectric points in the range of 4.6-5.2. The enzyme accounts for up to 0.1% of the total soluble protein in Drosophila and can be purified relatively easily by precipitation as an enzyme-substrate complex in ethanol. The amino acid composition of Drosophila amylase resembles that of the mammalian enzymes (Boer and Hickey 1986) although no immunological cross reaction is detected between the two alpha-amylases. In contrast to the large differences in pH optima between insect groups, the amylases of Drosophila and mammals have similar biochemical properties (Hickey et al. 1989); this is attributed to the fact that both the mammalian and Dipteran amylases act under similar physiological conditions.

Amylase Polymorphisms in Drosophila

Over the past twenty years, the measurement of genetic variation has been made almost exclusively by one technique, i.e., native-gel electrophoresis of soluble proteins. This technique distinguishes protein variants by differential

rates of migration in either polyacrylamide or agar gels under a uniform electric field.

These efforts have made major contributions to our knowledge about the distribution of genetic polymorphism between species, populations and loci. In many species of plants and animals, and especially in Drosophila, electrophoresis has revealed extensive protein polymorphism for soluble enzyme loci. For example, the average level of heterozygosity in D.pseudobscura, which is typical of many invertebrates, is estimated to be 0.125 (Powell 1975) based on electrophoretic studies of 27 loci.

Pioneering work in the amylase locus of Drosophila melanogaster has been done by Kikkawa and Abe (1960). In establishing the electrophoretic phenotypes of this organism, Kikkawa (1964) designated the common, single-banded zymogram pattern as the ancestral phenotype and also described several less common, double-banded variants. Since that work, genetic variability of D. melanogaster amylase has been reported on the basis of isozyme mobility, heat denaturation, specific activity, and combinations of these parameters (Doane et al. 1975; Hickey 1979). Eight electrophoretic variants of amylase have been found in D. melanogaster. In their survey of protein polymorphisms in Drosophila, Singh et al. (1982) found amylase polymorphism in all populations surveyed, although some are much more polymorphic than others. The

amylase variation in West African populations is remarkably high (Hickey 1979) with a total of 23 amylase phenotypes, and the most common phenotypes accounting for only 17% of the total frequency (Singh et al. 1982).

This finding of genetic polymorphism is not unique to amylases and may not have significance with regard to natural selection (Kimura and Ohta 1974). Johnson (1974), however, in discussing the observations of Gillespie and Kojima (1968), has suggested that enzymes whose substrates originate from the external environment are more variable than enzymes involved in intermediary metabolism: "the relationships observed suggest that enzyme polymorphisms increase fitness by providing a means of metabolically compensating for a varying environment." de Jong and Sharloo (1976) discussed their observations on selection on D.melanogaster amylase phenotypes with regard to Haldane's (1949) suggestion that chance fluctuations in the environment may have a strongly determinative effect on allelic frequencies found at some later time.

Linkage studies and chromosome assignments

The earliest report of amylase linkage analysis resulted from Matsumura's investigations of the silkworm in 1934. Kikkawa (1960) subsequently mapped the D.melanogaster amylase

gene between positions 77 and 80 on the genetic map of chromosome 2 by using strains with heritable variations in amylase enzyme activity and various well-defined gene markers. In later descriptions of electrophoretic phenotypes, it was suggested that amylase in D.melanogaster might be controlled by two closely linked genes (Kikkawa 1964).

Direct genetic evidence was obtained for two amylase loci in D.melanogaster by Doane (1967) and Eahn(1967). Bahn's study consisted of an intensive screening effort to obtain recombinants between c (curved wings) and wt (welt eyes), two closely linked markers which bracket the amylase loci. He obtained 5,039 recombinants out of a total of 77,523 offspring. Electrophoretic phenotypes showed that 6 of the recombinants had occurred between the amylase loci, yielding a map distance of 0.008 Centimorgan.

Approximate cytological localization of the amylase region on chromosome 2 was obtained by Doane (1969a) from recombinational analysis involving a duplication inserted at section 52F on chromosome 2R. She showed that the amylase region mapped to the right of, and relatively close to, section 52F. In a later study involving translocation mapping, Bahn (1971a) localized the amylase genes between segments 54B and 55. Hybridization of cloned sequences of the amylase genes to polytene chromosomes made the localization of the genes within section 54A1-B1 (Gemmill et al. 1985).

Amylase Gene Duplication

The amylase genes in Drosophila melanogaster provide a good experimental model of gene duplication. An important step in the creation of multiple gene copies is the duplication of a gene to produce two copies. Gene duplication is probably the most important mechanism for generating new genes and new biochemical processes that have facilitated the evolution of complex organisms from primitive ones. It is also important for generating new genes of the same function and hence enables the production of a large quantity of RNAs or proteins (Ohno 1970). The fact that gene duplication has played a vital role in evolution can be seen from the comparison of the DNA contents of various organisms. Perhaps by being able to make more of a particular protein, organisms with two copies of a particular gene have an advantage over organisms with only one copy. Two copies can duplicate again to produce four, and so on. Once two copies of a gene exist, it is possible that, through the accumulation of different base substitutions, the sequence of the two genes could diverge over hundreds of thousands of generations. Consequently, the two copies would no longer be identical. Thus, gene duplication may provide a basis for the origin of adaptive novelties. Furthermore, one would expect that the

longer the multiple copies exist, the greater the sequence difference between the genes will be.

Whether or not the amylase gene is duplicated in strains characterized by a single amylase isozyme has been, until recently, questionable. The molecular cloning and characterization of the amylase locus in D.melanogaster provided direct evidence that all strains have two copies of the amylase structural gene including the amylase null strain which produces no electrophoretic band (Hickey et al., 1988). The restriction map shows that the duplicated genes are present in inverted orientation (Levy et al. 1985, Benkel et al. 1987), and comparisons of nucleotides from the two genes indicate that they are divergently transcribed (Boer and Hickey 1986). The two copies, each 1.4 kb in length are separated by approximately 4.8 Kb of non-coding sequence.

The number and pattern of electrophoretic variants within each species, and in cross-species hybrids (Dainou et al. 1987), coupled with the recent analysis of restriction maps (Payant et al. 1988), have shown that the duplication is not restricted to D.melanogaster but is also present in the seven other species of the melanogaster species subgroup and two members of the obscura subgroup (Brown et al. 1989, Doane et al. 1987).

When the duplicated coding sequences within D.melanogaster were compared (Boer and Hickey, 1986), they

seemed to be very similar, pointing towards a very recent duplication event, whereas the amount of sequence divergence is much greater if flanking upstream sequences are considered, hence leading to longer estimates of the time since the sequences had diverged. From genomic southern blot analysis of seven related Drosophila species, it appeared that the duplication is not a recent one and that it predated the speciation events within the species subgroup (Payant et al. 1988). This suggested the possibility of concerted evolution of the duplicated structural genes; this process would explain the similarity in the coding sequences, despite the considerable age of the duplication.

Gene conversion is the major type of recombination between repeated genes during mitosis and meiosis in fungi (Scherer & Davis 1980; Jackson & Fink 1981; Klein 1984) and between repeated genes transfected into mammalian cell lines (Liskay & Stachelek 1983; Subramani & Runitz 1985; Brenner et al. 1985). While direct evidence for conversion is generally lacking, many instances of recombinational processes similar to gene conversion have been also detected in multigene families whose members are clustered within the genome (Slightom et al. 1980; Lieberhaber et al. 1981; Cohen et al. 1982; Pease et al. 1983; Weiss et al. 1983; Krawinkel et al. 1983). Short conversions between the members of these families, i.e., conversions that do not encompass the entire

gene, have been shown to give rise to new sequence variants. However, the long term effect of such sequence transfer between members of a multigene family is not clear, although it has been suggested that it would lead to concerted evolution of the genes (Ohta 1980; Baltimore 1981; Dover 1982; Nagylaki & Petes 1983; Arnheim 1983; Nagylaki 1984a,b).

We now have more direct evidence for rapid gene conversion and concerted evolution of Drosophila amylase sequences (see Chapter 4). Similar conversion events have also been described recently for mammalian amylase sequences (Gumucio et al. 1989); but, again, the evidence is not as conclusive as in the Drosophila amylase example.

REGULATION OF ALPHA AMYLASE SYNTHESIS IN DROSOPHILA

DEVELOPMENTAL AND TISSUE-SPECIFIC CONTROL

The alpha amylase genes of *Drosophila* are typical of higher eukaryotic genes in that their expression shows definite patterns of tissue-specificity and developmental stage specificity; they are unusual in also having a dietary-regulated pattern of gene expression that is more typical of prokaryotic genes.

The developmental control of amylase gene expression is illustrated by the variations in amylase concentrations that have been detected during the life span of Drosophila melanogaster (Doane, 1969). For instance, it has been shown that amylase levels rise during successive moults of Drosophila larvae and then drop following formation of the puparium (Doane 1969b). In emerging adult fruit flies, the level of amylase increases approximately 10-fold and remains at a relatively high level. Doane (1969b) also reported a shift in the proportion of the total activity found in isozymes of an Amy^{3,6} strain. Nearly all the activity is found in the number 3 band in newly hatched larvae; but by the end of the instar, 60% appeared in the number 6 band and only 40% remained in the number 3 band. At the time of emergence of the adult, almost all the activity had reverted to the number 3 band but again shifted to the number 6 as the adult aged.

Furthermore, the relative activity of different isoamylases differs in different tissues of the same individual. Third instar larvae of Amy^{3,6} have more Amy³ activity than Amy⁶ activity in the haemolymph yet more Amy⁶ than Amy³ activity in the midgut. One week after emergence, however, flies show more Amy⁶ activity in the haemolymph, and almost only Amy³ activity in the midgut. Most amylase activity in larvae and adults of D.melanogaster, is found in the midgut and haemolymph with very low levels in the

salivary glands and fat body (Doane, 1969b). In D.melanogaster, the tissue specific and temporal expression of amylase in the posterior midgut of the adult is, in part, under genetic control by a separate, trans-acting locus, termed map (mid gut activity pattern; Abraham and Doane 1978; Doane et al. 1983). It appears that the map gene regulates the level of translatable amylase mRNA (Buchberg 1983; Doane et al. 1983). Although map regulates the level of amylase expressed in the adult posterior midgut, it seems to have little or no effect on amylase expression in the anterior midgut. There is at least one other controlling element located in chromosome 2R that is genetically separable from both map and amy and that regulates amylase levels in the anterior midgut (Doane 1980; Doane et al. 1983). Similar patterns of mid-gut expression have been found in D.pseudoobscura (Powell and Litchtenfels 1979).

Effects of Dietary Carbohydrate

Dietary conditions affect amylase activity levels in D.melanogaster (Abe 1958; Doane 1969b; Hickey 1977; Hoorn and Scharloo 1978; Yamazaki and Matsuo 1984). Experiments by Hickey and Benkel (1982) demonstrated that amylase activity is repressed, rather than enhanced, by simple dietary sugars such as glucose, maltose and sucrose. The repressing effect

of the disaccharides is due to changes in the level of amylase enzyme quantity i.e. reduction in the number of amylase molecules rather than a change in the catalytic efficiency of the enzyme (Hickey and Benkel 1982). In more recent studies Benkel and Hickey (1985; 1986a) showed that larvae of the Oregon-R strain of D.melanogaster gave a hundred fold repression of amylase in response to the addition of dietary glucose, and this change correlates with a similar reduction in the levels of translatable amylase mRNA assayed in micro-injected Xenopus oocytes. Moreover, they showed that RNA from repressed and derepressed larvae probed with amylase sequences show comparable changes in the abundance of amylase mRNA; both results provide evidence that the change in enzyme quantity occurs at a pre-translational level (Benkel and Hickey 1987). Dietary repression of amylase, like the strain-specific differences in activity, is controlled by regulatory factors which probably affect the levels of transcription at the amylase locus.

The glucose repression of the amylase genes in D.melanogaster, like repression of bacterial and yeast catabolite systems, can be reversed by the use of exogenous cyclic AMP (Benkel and Hickey 1986a, Hickey et al. 1989; Magoulas and Hickey, 1989). *Drosophila* populations from tropical West Africa are found to be less responsive to dietary glucose (Benkel and Hickey, 1986b). Sibling species

of D.melanogaster also show glucose repression of amylase
(Payant et al. 1988).

SUMMARY

Since amylase genes have a wide phylogenetic distribution and show significant evolutionary conservation of their amino acid sequences, they have received considerable attention, especially the prokaryotic and mammalian enzymes. Amylase genes in Drosophila are of special interest because they combine features of eukaryotic genes, such as tissue-specific and developmental stage-specific expression, with glucose regulated expression, a feature typical of prokaryotic genes.

The experimental work presented in the coming chapters is an attempt to quantitatively investigate both intra-specific and interspecific DNA sequence variation at the amylase locus within the genus Drosophila. These results should help link short-term microevolutionary changes to the longer-term macro-evolutionary patterns of amylase gene evolution. Chapter 2 deals with surveys of genetic variation within populations of D. melanogaster, using both enzyme polymorphisms and restriction fragment length polymorphisms; Chapter 3 also deals with genotypic differences between strains, but at the level of DNA sequence; Chapter 4 extends the analysis to include a closely-related species, D. erecta; and Chapter 5 examines the structure and sequence of the amylase gene in D. virilis, a species which diverged from

the melanogaster group approximately sixty million years ago and which is therefore more distantly related to species within the melanogaster subgroup.

CHAPTER 2

ALLOZYME AND RESTRICTION FRAGMENT LENGTH POLYMORPHISMS

INTRODUCTION

In the past, electrophoretically-detected enzyme polymorphisms have been used extensively to study both the extent and patterns of genetic variation in natural populations (reviewed by Lewontin 1974). Until recently, however, there has been little information available on the extent of DNA-sequence variation in natural populations. With the advent of new DNA-based technology, there has been a shift towards the study of mutational variation directly at the level of DNA. Some of this recent work has involved the sequencing of extensive stretches of DNA from many genotypes. The pioneering work of Kreitman (1983) showed that there is an unexpectedly large fraction of the total variation which is not expressed in the protein product. Most of this work, however, is indirect in that it uses restriction fragment length polymorphisms (RFLPs) as an index of the underlying DNA sequence variation (Avis et al. 1979; Jeffreys 1979;

Brown 1980; Wyman and White 1980; Langley et al. 1982; Birely 1984; Aquadro et al. 1986; Cross and Birely 1986; Kreitman and Aguade 1988; Schaeffer et al. 1987; Aguade 1988; Langley et al. 1988). Restriction endonuclease mapping allows a relatively quick survey of large fragments of DNA in many strains or individuals. As well as identifying restriction site or individual DNA base variations, the technique also identifies large insertions and deletions of DNA relative to a standard sequence in a selected section of DNA.

Given the existence of a very large body of information on allozyme polymorphisms (for many species and genes) along with the potential for gathering an equally large data set on DNA polymorphisms in the future, it is of interest to correlate the electrophoretically-detected enzyme polymorphism with the variation at the level of DNA, at a given gene locus. To do this I have chosen the amylase locus in Drosophila melanogaster; this gene-enzyme system is known to show high levels of enzyme polymorphism in nature (Hickey, 1979a; Singh et al. 1982) and it has recently been shown to display significant levels of restriction fragment length variation (Gemmell et al. 1985; Langley et al. 1988).

The aim of the work presented here is to look at DNA variation at a locus which is already very polymorphic at the protein level. Rather than doing random sampling of genotypes from natural populations, a set of genotypes that represent

a variety of enzyme variants and geographic origins was chosen. In this chapter, I will describe D.melanogaster populations from different localities in three different continents (America, Europe and Africa) as well as some isolated oceanic island populations (Fiji, Mauritius and Seychelles). This study deals with variations between 50 strains in the restriction map of a 22 Kb region of DNA which contains the coding sequences for the duplicated amylase genes.

MATERIALS AND METHODS

Fly stocks

Drosophila melanogaster strains from Ottawa, Canada were collected by Dr. D. Hickey at Ottawa university and Dr. G. Carmody at Carleton University, the strains from Fiji Islands were collected by V. Payant, and strains from Mauritius, Seychelles and Egypt were a gift from Dr. J. David at Laboratoire de Génétique et Biologie Evolutive, France. The West African strains are also provided by J.David. The null strain is obtained from D. Hickey and strains from Italy and Nigeria are provided by Ada Loverre.

Drosophila stocks were maintained as isofemale lines. Isochromosomal lines were established by the standard Curly chromosome extraction method, using SM5 as extractor chromosome (Lindsley and Grell 1986) as in Appendix 2.2. Amylase electrophoretic phenotypes were determined for all strains using 5% polyacrylamide gels (19:1 acrylamide:bisacrylamide), pH 8.9 according to the method described by Hickey (1981). Single adult flies, approximately one week old, which had been maintained on an "instant" Drosophila medium (Carolina Biologicals) were used for electrophoresis both before and after chromosome extraction (Appendix 2.1 and Table 2.1). Single flies were crushed and the crude homogenates were electrophoresed at 200 V. Gels were incubated in a 3% starch solution for one hour then stained with Iodine- Potassium Iodide solution (1:2:10) to visualize the bands resulting from the activity of the amylase on the starch (Hickey 1981; Benkel and Hickey 1986).

DNA Preparation And Electrophoresis

Total high molecular weight DNA was extracted from one gram of flies from each strain according to the method of Davis and Davidson (1984). Fly homogenates were lysed, treated with RNAase and proteinase K, and extracted with phenol/chloroform and ethanol precipitated. Restriction

enzyme digestions of 5 µg DNA were performed for 4 to 6 hours using 2 to 3 units of enzyme per microgram of DNA, under the conditions recommended by the supplier. Six-cutter digests (enzymes with a hexanucleotide recognition sequence) including Bam HI, Bgl II, Eco RI, Hind III, Pst I, PvuII, Sal I and Xho I, were separated in 0.8%-1% agarose at low voltage gradients in Tris Acetate and 2.5mM EDTA. Four-cutter digests of Alu I, Ban I, Dde I, Msp I, Hae III, Hinf I, Hinp I, Hha I, Sau 3A, Sau 961 and TaqI were ethanol precipitated, washed, dried under vacuum and resuspended in 3 µl of formamide loading buffer (94% formamide/0.05% xylene cyanol/0.05% bromophenol blue/ 10mM Na₂EDTA, pH 7.2. After denaturing samples at 90°C for 5 minutes, 2 µl of each sample was loaded onto a standard 30 cm X 40 cm X 0.4mm 5% polyacrylamide/7M urea buffer gradient sequencing gels (Kreitman and Aguade 1986). Gels were run at 1200-1300V. Electrophoretic transfer of DNA from the gel to New England Nuclear/Dupont GeneScreen and subsequent UV cross-linking was performed as described by Church and Gilbert (1984).

Probe Preparation

Membrane-bound DNA was hybridized with [³²P] labelled molecular probes. Two different probes were used (i) pORM7, a cDNA probe which hybridizes with both of the duplicated

amylase coding sequences and (ii) pCS-1.4, a 1.4 kb Eco RI fragment which is specific for the region upstream of the proximal amylase coding region (Benkel et al. 1987). Probes were prepared based on a method that uses priming of second strand DNA synthesis using random primers in the presence of klenow, dNTPs and ^{32}P -dCTP. 25 ng of gel purified DNA fragments were labelled using 3000 Ci/mmol of $[\alpha\text{-}^{32}]\text{dCTP}$ to a specific activity of $1\text{-}5 \times 10^9 \text{dpm}/\mu\text{g}$ of DNA (Hellman and Peterson 1987).

Hybridization and Wash

Prehybridization and hybridization was at 65°C using bags or polycarbonate (lexan) tube sealed at one end in solutions of 5X Denhardt's, 5X SSC, 0.5% SDS, 0.01 EDTA and 500 $\mu\text{g}/\text{ml}$ denatured salmon sperm DNA. The membranes were washed three times at room temperature using 5mM Na_2PO_4 , 1mM EDTA, and 0.2% SDS. A stringent wash was done at 65°C in 0.1 SSC and 0.1% SDS.

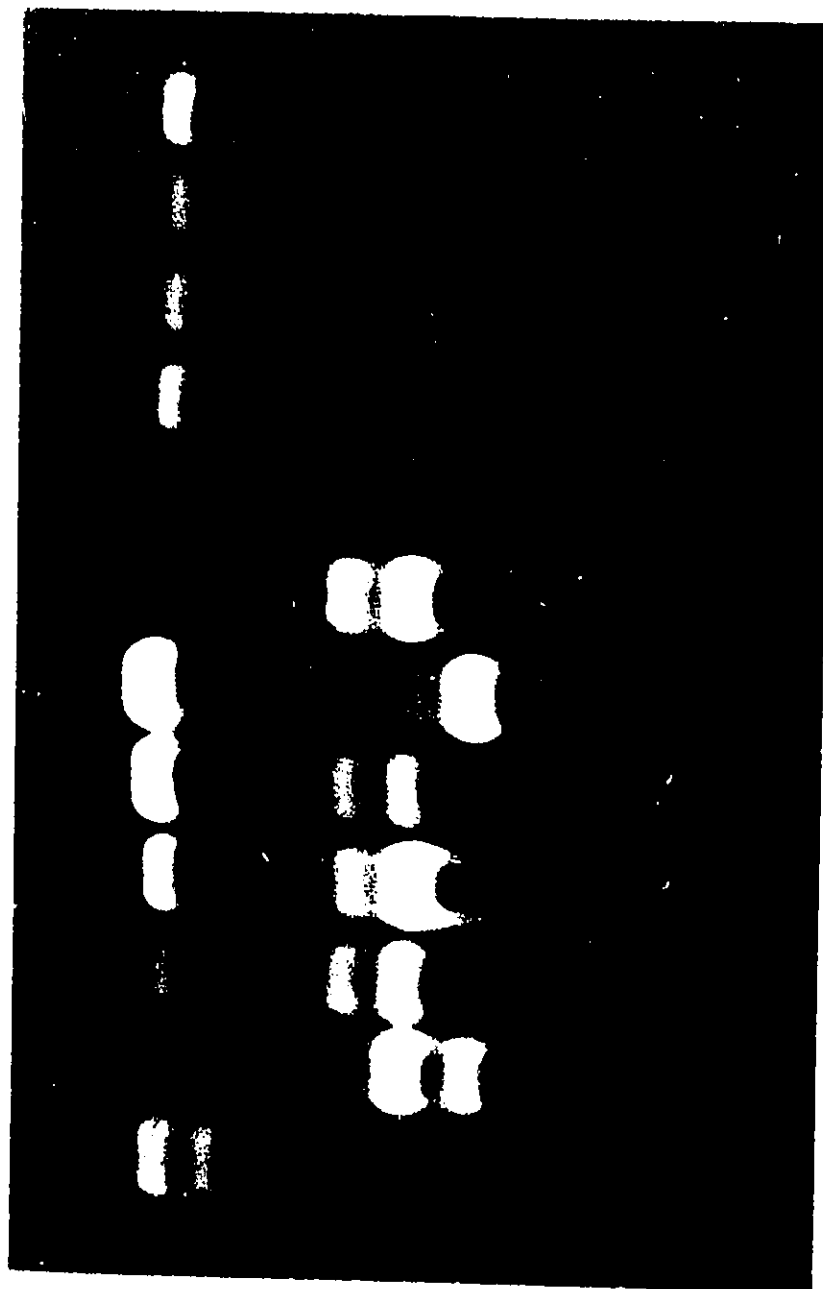
Probes were removed by boiling the membranes for 5 minutes in 5mM EDTA and 0.5% SDS for rehybridization.

RESULTS

Electrophoretic Phenotypes of D.melanogaster

An initial survey of amylase phenotypes in more than 200 strains of D. melanogaster was made (see Appendix 2.1). An electrophoretic gel showing the amylase phenotypes for some of the strains is shown in Figure 2.1. First, I used mass homogenates of 5 flies from each strain; this provided the genotype for the single-banded amylase variants. In order to test whether the samples with multiple-banded patterns were homozygous or heterozygous, five individual flies from each of these strains were assayed; lines which were polymorphic showed differences in the banding pattern between individuals within a strain. After this initial survey, fifty lines were chosen for further study, based on the fact that they were representative of the full range of allozyme types and geographical origins. Second chromosomes (which contain the amylase locus) were extracted from these fifty lines using the crossing scheme shown in Appendix 2.2.

Electrophoretic phenotypes for these isochromosomal lines were scored in starch-iodine- stained gels (Hickey, 1981) and the results are shown in Table 2.1 and Figure 2.2. The patterns shown in Figure 2.2 illustrate the range of single and double-banded amylase phenotypes. From the Figure, it



Italy 108

Italy 62

Italy 61

Italy 8

Benin 3

Benin 37

Benin 31 A

Benin 30 A

Benin 30

Benin 27 A

marker

Amy 7
Amy 6
Amy 5
Amy 4
Amy 3
Amy 2
Amy 1

Figure 2.1. Amylase enzyme electrophoretic patterns of several Drosophila melanogaster strains from Italy and West Africa, before chromosome extraction.

Strain designations are shown at the top; amylase mobilities are marked on the side. Each homogenate contained two larval equivalents.

Figure 2.2. Negative print of a gel stained for amylase activity, showing isogenic strains with different enzyme and activity patterns.

Lanes 1 to 3 contain homogenates from three different strains, each of which produces amylase with the same electrophoretic mobility (Amy¹), but with different levels of activity. Lane 4 shows a strain with the Amy³ phenotype, lane 5 has Amy⁴, and lane 10 contains Amy⁵. Lanes 6 through 9 contain homogenates of genotypes which give double-banded amylase patterns (Amy^{1,3}, Amy^{1,3}, Amy^{4,6}, and Amy^{5,6}, in that order).

1 2 3 4 5 6 7 8 9 10



— AMY 6
— AMY 5
— AMY 4
— AMY 3
— AMY 1

Table 2.1. Enzyme polymorphisms and restriction polymorphisms in isogenic strains of D. melanogaster from different geographic locations.

Strain	Allozyme Pattern	Restriction Fragments *	
		<u>Eco RI</u>	<u>Sal I</u>
1. Oregon R	1	5.2, 5.0 Kb	6Kb
2. Italy 108	1	5.2, 3.8	6Kb
3. Benin 16R	1	5.2, 5.0	6Kb
4. Ottawa 101	1	5.2, 3.8	6Kb
5. Ottawa 107c	1	5.2, 5.0	6Kb
6. Ottawa 107d	1	5.2, 5.0	6Kb
7. Ottawa 202	1	5.2, 3.8	6Kb
8. Ottawa 210	1	5.2, 3.8	6Kb
9. Ottawa 34	1	5.2, 3.8	6Kb
10. Maurice	1	5.2, 3.8	6Kb
11. Seychelles	1	5.2, 3.8	6Kb
12. Tunisia 224	1	5.2, 5.0	6Kb
13. Tunisia 227	1	5.2, 5.0	6Kb
14. Tunisia 250	1	5.2, 5.0	6Kb
15. Fiji 301	1	5.2, 5.0	6Kb
16. Fiji 302	1	5.2, 5.0	6Kb
17. Fiji 303	1	5.2, 5.0	6Kb
18. Fiji 306	1	5.2, 5.0	6Kb
19. Benin 7cA	1	5.2, 5.0	6Kb
20. Benin 9cA	1	5.2, 5.0	6Kb
21. Benin 23c	1	5.2, 5.0	6Kb
22. Benin 23cA	1	5.2, 5.0	6Kb
23. Benin 22	3	5.2, 5.0	6Kb
24. Ipassa	3	5.2, 3.8	6Kb
25. Liberaville	3	5.2, 5.0	6Kb
26. Kumba	4	5.2, 3.8	6Kb
27. Ottawa 31	4	5.2, 3.8	6Kb
28. Makokou	1,2	5.2, 6.3	6Kb
29. Italy 19	1,3	5.2, 6.3	6Kb
30. Ottawa 37	1,3	5.2, 6.3	6Kb
31. Canton S	1,3	5.2, 6.3	6Kb
32. Brownsville	1,3	5.2, 6.3	6Kb
33. Italy 69	1,3	5.2, 6.3	6Kb
34. Benin 5a	1,5	5.2, 6.3	6Kb
35. Benin 5b	1,5	5.2, 6.3	6Kb
36. Benin 25a	1,5	5.2, 6.3	6Kb
37. Benin 25b	1,5	5.2, 6.3	6Kb
38. Ottawa 235	1,6	5.2, 6.3	6Kb

Table 2.1. cont.

39. Benin 30	1,6	5.2, 6.3	6Kb
40. Ottawa 107b	2,6	5.2, 6.3	6Kb
41. Makokou	4,6	5.2, 6.3	6Kb
42. Cotonou	5,6	5.2, 6.3	6kb
43. Benin 16c	5,6	5.2, 6.3	6Kb
44. Lagos 49	5,6	5.2, 6.3	6Kb
45. Benin 35a	5,6	5.2, 6.3	6Kb
46. Benin 35b	5,6	5.2, 6.3	6kb
47. Benin 1a	1,4,5	5.2, 6.3	6kb
48. David's	5,6,7	5.2, 4.1	6Kb
49. Cairo	1,3,	5.2, 6.3	6kb
50. Cairo	1,5		

* Eco and Sal are genomic DNA fragments, resulting from cleavage with Eco RI and Sal I restriction enzymes; blots were probed with pOR-m7, which is specific for the amylase coding region. Fragment sizes were determined relative to the electrophoretic mobility of Eco RI-HindIII double-digested lambda DNA.

can also be seen that enzyme band intensity varies significantly between strains, including strains which are characterized by the same electrophoretic mobility. It also varies between the two bands within a double-banded phenotype. For instance, if we compare lanes 6 and 7 in Figure 2.2, it can be seen that both genotypes have the electrophoretic pattern Amy^{1,3}; in sample 6, however, the faster migrating band is the most intense, whereas in sample 7 the relative band intensities are reversed.

These genetically-determined differences in amylase gene expression have been described previously (Doane 1969; Hickey 1981), and they have recently been shown to reflect differential sensitivity of the duplicated amylase genes to dietary glucose (Benkel and Hickey 1986b). The fast-migrating, single-banded amylase phenotype (lanes 1 to 3, Figure 2.2 and lanes 1 to 4, Figure 2.1) is typical of Drosophila strains originating in Europe and North America, whereas the more slowly-migrating, double-banded patterns are much more common in Equatorial Africa (see Appendix 2.1). Figure 2.1 gives an example of the populational differences in allozyme pattern that one observes between European and African populations, even using very small samples such as the one shown in the Figure. The European lines are generally homozygous for the Amy¹ allele, whereas isofemale lines from

West Africa are very variable in their electrophoretic phenotype. This variation occurs not only between lines, but also within lines, many of which are heterozygous and segregating for different single-banded and double-banded electromorphs. This segregation, which is evidenced by triple-banded amylase phenotypes, is not seen in the derived isogenic strains (see, for example, Figure 2.2).

It was already known from earlier studies that the amylase locus is highly polymorphic for electrophoretic alleles in D.melanogaster, but that temperate populations are less polymorphic than tropical or sub-tropical populations (Hickey 1979a; Singh et al. 1982). This high degree of amylase structural variation is also clear in African populations of other closely related species (Dainou et al., 1988). This peculiar pattern of electrophoretic variation which correlates with geographic localization lead Hickey (1979) to postulate that either D.melanogaster originated in Africa or that environmental conditions in that continent favour amylase variants which are rare or absent elsewhere. The focus of this study was not just to confirm these allozyme patterns, but to correlate the allozyme variation with the restriction enzyme data (described below).

Restriction Site Polymorphisms

The basic restriction map presented in Figure 2.6 was generated from the hybridization patterns of total genomic DNA; it is consistent with those generated from the study of cDNA or genomic clones (Gemmill et al. 1986; Benkel et al. 1987). To facilitate comparison with the work of Langley et al. (1988), the center between the two transcriptional units has been given the co-ordinate 0.0. Restriction sites within the probed region will be referred to in relation to this co-ordinate. The analysis of DNA from the fifty strains of D.melanogaster indicates that the duplicated gene structure is conserved in all strains, regardless of their amylase phenotype or geographic origin. This is consistent with other molecular results using the amylase region of D.melanogaster (Gemmill et al. 1986; Langley et al. 1988) and in related species (Payant et al. 1988). All strains contain two divergently-transcribed amylase coding sequences, located approximately 4Kb apart on the second chromosome.

When cutting with enzymes such as Sal I and Bam HI, which cleave within the coding region of the two genes, a single pattern is seen for all strains (Figure 2.3); however, cleaving with enzymes like EcoR I or Hind III that cut within the flanking non conserved sequences, frequent polymorphisms are evident (Figures 2.3, 2.4 and 2.5).

Figure 2.3. Restriction digest of total fly DNA probed with D.melanogaster pOR-m7. (A) = Eco RI digest. (B) = Sal I digest. Notice uniformity of pattern for Sal I and polymorphism due to the Eco RI site at (-2.2).

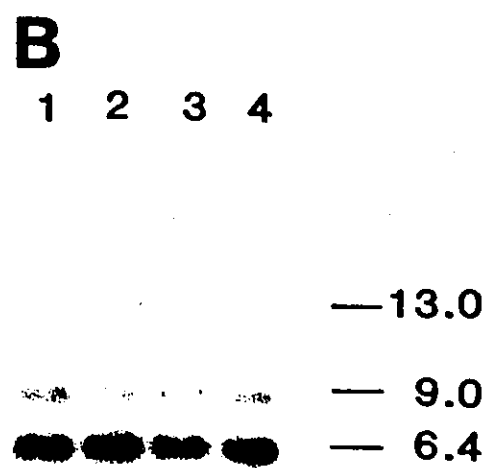
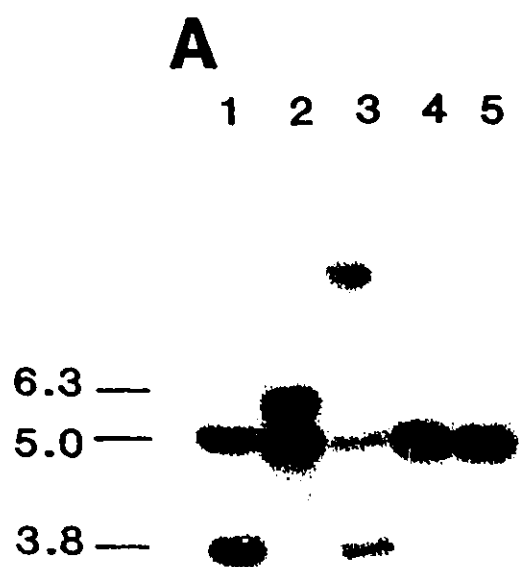


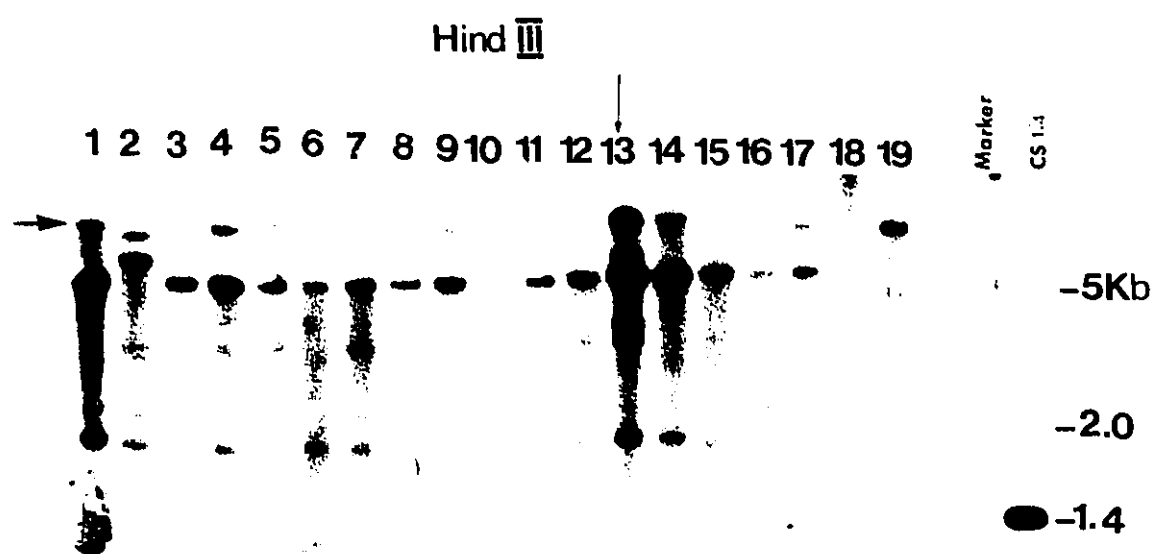
Figure 2.4. Restriction digest of total fly DNA from 14 different strains of D.melanogaster cleaved with the tetranucleotide Msp I and probed with pCS-1.4 (a flanking region probe). Arrow points to polymorphic 562 bp fragment at position -1.7 of the map in Figure 2.6. Only polymorphic fragments are shown here, the rest of the fragments recognized by the probe are not shown.

1 2 3 4 5 6 7 8 9 10 11 12 13 14



Msp I

Figure 2.5. Hind III digest of total fly DNA from 19 different D.melanogaster strains probed with pCS-1.4. Note monomorphic intergenic fragment of 2 Kb. Horizontal arrow indicates insertion/deletion variants at Hind III (position + 10.7 of Figure 2.6). Lane 2 is the amylase null strain (see Hickey et al. 1988, for details of the inversion). Lane 13 is a Benin 35b (#46 in Table 2.1) strain with multiple insertions. Marker indicates a lambda DNA digested with Hind III and Eco RI and labelled with a ³²Plambda fragment. pCS-1.4 is the gel purified fragment used to probe the genomic DNA.



Using the published DNA sequence of the laboratory strain Oregon-R as a standard (Boer and Hickey 1986), two types of restriction fragment length polymorphisms (RFLPs) in the amylase region were detected (see Figures 2.6 and 2.7). Nine of the RFLPs were due to the presence or absence of a particular restriction site (Figures 2.4 and 2.6) while fourteen others were due to the insertion or deletion of a DNA sequence within a restriction fragment (Figures 2.3, 2.5 and 2.7). Restriction site polymorphism was observed at positions Eco RI (-6.5), Hae III (-2.4), Eco RI (-2.4), Hinf I (-2.9), Sau 3A (-1.8), MspI (-1.4), Alu I (-1.8), Eco RI (+4.6), and Hind III (+10.7) of Figure 2.6. Polymorphisms due to the presence or absence of a given restriction site are not limited to particular populations; however, RFLPs due to insertions or deletions are limited to geographically distinct populations, for example the 70 bp insert is fixed in populations from Benin, West Africa (see Table 2.2). There is more DNA variation within African populations (Benin strains) than within the European and North American strains.

Estimates for genetic variability for the amylase region were obtained using the estimates of Ewens, Spielman and Harris (1981) in equation 13:

$$\theta = (K_4 + K_6) / (8m_4 + 12m_6)$$

where K_6 and K_4 are the number of cleavage sites for enzymes with four- and six-site recognition sequence respectively, and where m_4 and m_6 are the cleavage sites screened for enzymes that have four and six-site recognition sequence.

The proportion of nucleotide sites which are polymorphic is 0.014 and the estimate of heterozygosity per nucleotide is 0.003. When considered as a whole, the heterozygosity for restriction sites, in the entire probed region (including both restriction site and insertion/deletion variation) is 0.86, estimated according to Nei and Tajima (1981).

Phylogenetic Comparison of Amylase Alleles.

Several additional features of the evolution of amylase in Drosophila melanogaster can be inferred from a different type of analysis - by comparison of shared polymorphic sites among pairs or groups of alleles. To determine the relationship between the 50 amylase morphs, I have used the unweighted-pair group cluster method (UPGMA) of Sneath and Sokal (1973, pp 230-234). This method utilizes average distances between sample pairs (or groups) to determine branching orders in the phenogram. The resulting phenogram is shown in Appendix 2.3.

In general, UPGMA phenograms do not reveal the true phylogenetic relationships or evolutionary distances.

However, in this clustering, the underlying rate of evolution for the measured character, nucleotide differences, should be reasonably uniform since it is governed primarily by mutation and genetic drift. The major features of the alignment of alleles in Table 2.1 (electrophoretic pattern) are preserved in the phenogram, with the exception of line #38 (Ottawa Amy^{1,6}) which is aligned with the Amy¹ strains.

The phenogram underscores the breadth of the geographic distribution among closely related alleles. This is clear for line #19 (Benin 7CA). The fact that closely-related alleles are broadly distributed geographically provides additional evidence that there are a limited number of haplotypes segregating with the species.

Figure 2.6. General organization of the autosomal amylase locus in 50 D.melanogaster. Amylase genes and their transcriptional orientations are indicated by arrows. The coordinate on the top is in Kilobases, centered between the two transcriptional units as in Langley et al. (1988).

The Figure represents RFLPs due to presence or absence of a restriction site. Restriction sites above the bar are monomorphic, sites below the bar are polymorphic. Not all monomorphic sites are shown in the map. Restriction enzymes are Alu= Alu I; B=Bam HI; Bgl=Bgl II; E=Eco RI; Hae= Hae III; H=Hind II; Hinf=Hinf I; M=Msp I; S=Sal I; Sau=Sau 3A; Xho=Xho I.

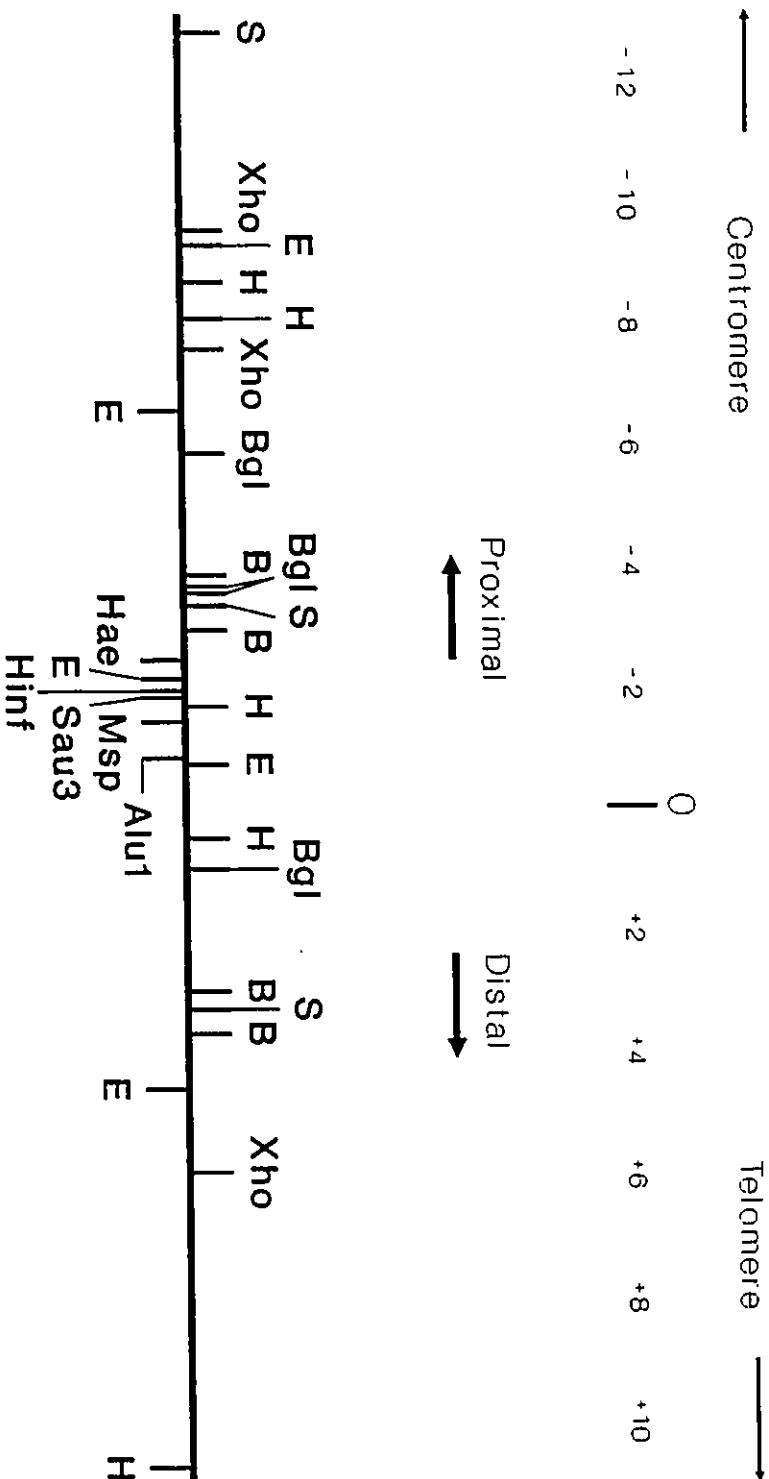
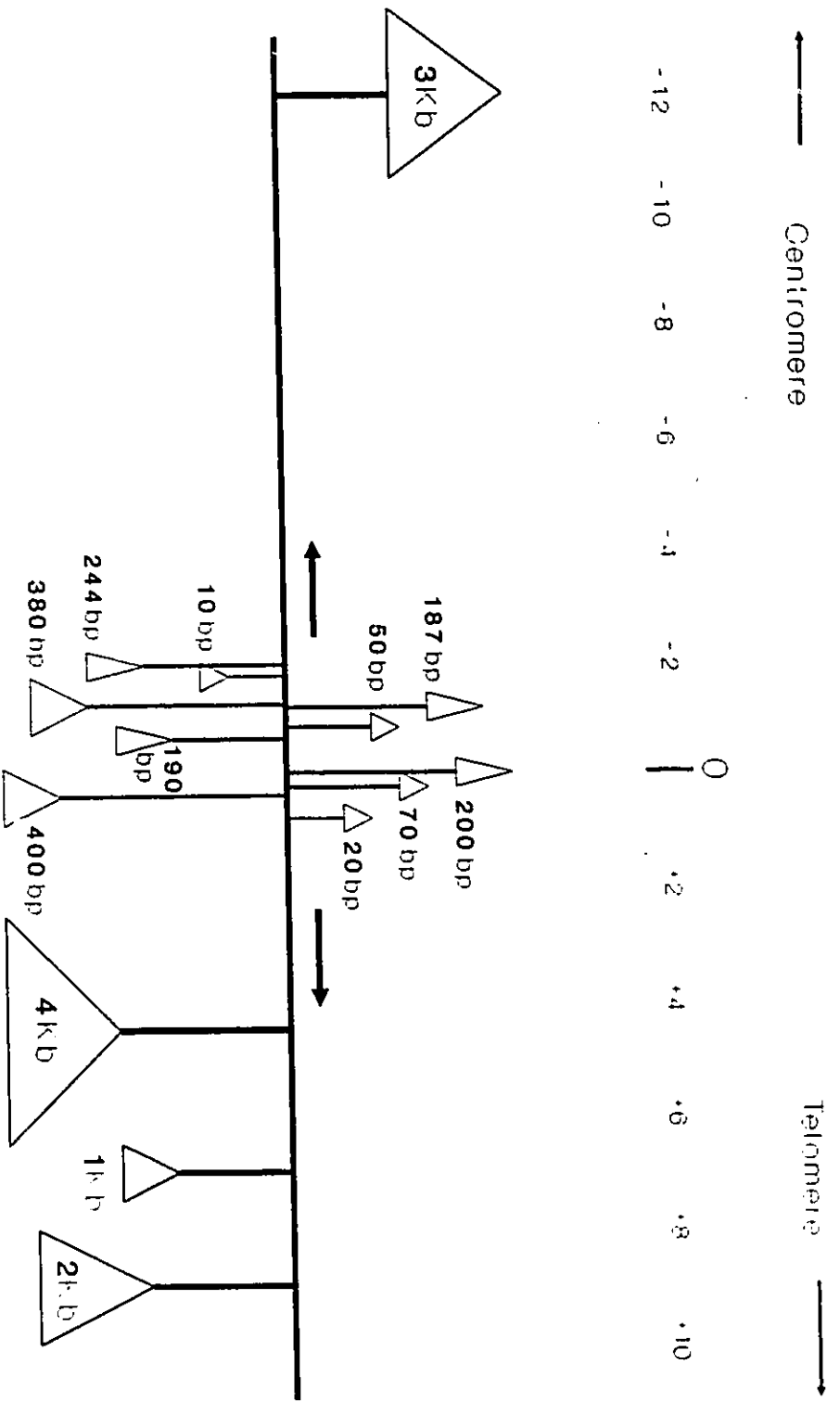


Figure 2.7. Represents RFLPs due to insertions and deletions in the amylase region. Triangles above the bar are deletions, below the bar are insertions. Enzyme designations and scale are the same as in Figure 2.6.



[illegible]

Table 2.2 Continued

ant	Haplotype																				Variant frequency
	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	
1 (-6.5)	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	0.06
11 (-2.2)	-	-	+	-	-	-	-	-	-	+	-	-	+	-	-	-	-	+	-	-	0.42
1 (-2.0)	+	-	+	+	+	+	+	+	-	-	-	-	-	+	-	-	-	-	+	-	0.20
1 (-1.9)	-	-	+	-	-	-	+	-	-	+	-	-	-	-	+	-	-	-	-	-	0.34
3A (-1.9)	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	+	-	0.10
1 (-1.7)	-	-	-	+	-	-	+	-	+	+	-	+	+	-	+	-	-	-	-	+	0.30
1 (-1.5)	-	+	+	-	+	+	+	-	-	-	-	+	-	-	-	-	+	-	-	-	0.48
11 (+4.2)	+	+	+	-	-	-	-	+	+	+	-	+	+	+	+	+	+	+	+	+	0.10
1111(10.7)	+	+	+	+	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+	+	0.14
244 bp	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	0.10
10 bp	-	-	-	-	-	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	0.04
380 bp	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	-	0.04
190 bp	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.06
400 bp	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.08
4kb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	0.02
1kb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.02
2kb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.02
3kb	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.04
187bp	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.04
50 bp	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.02
200bp	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.02
70bp	+	-	-	-	-	-	-	-	-	-	-	-	-	+	-	-	-	-	-	+	0.12
20bp	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0.02

50 lines of *Drosophila melanogaster* are listed by distinct electrophoretic patterns as in Table 2.1. Restriction map co-ordinates are from figures 2.6 and 2.7. The presence of a restriction map variant is indicated by a "+" and the absence by a "-". Variant frequency refers to the frequency of the least common state for a restriction map variant, per presence or absence.

Table 2.3 Non-random association (D) among amylase restriction sites and enzyme polymorphisms

[illegible]

Chi square associated with this value of D significant at $P < 0.05$, with one df. D is the ratio of the observed linkage disequilibrium (D) to its theoretical maximum D .
value of D indicates completely random association, values of -1 or 1 max represent complete association (Lewontin, 1964). Single- and double-banded refers to electrophoresis
in different enzyme mobility patterns. Restriction sites are as in Figure 2.6

Non-random associations

Tests for linkage disequilibrium were carried out on the 2X2 combinations of polymorphic restriction sites, and restriction sites and enzyme electromorphs, with Fisher's exact test (Table 2.2). D' is a normalized measure of non-random association proposed by Lewontin (1964). D' is the ratio of the observed linkage disequilibrium term, D , to its theoretical maximum, $D_{\max}$. Observed values of D are related to the chi-square statistic with one degree of freedom (Table 2.3).

There are strong non-random associations between amylase enzyme electrophoretic variants and those RFLPs which are due to the presence or absence of restriction sites (see Tables 2.1 and 2.2). The strongest associations in the amylase region are between double-banded enzyme variants and the presence of Eco RI and HinfI sites which map to the upstream non-coding region of the proximal amylase gene. The presence of a HaeIII site in the coding region is associated with single-banded allozymes. There are also significant associations among the restriction sites themselves (see Table 2.3). Of the twenty three RFLPs detected, only one maps to the amylase coding region (Hae III at 2.2), and it is not an insertion/deletion variant.

Unlike the results obtained by Langley et al. (1988), in the 50 strains studied here, insertions/deletions tend to cancel out in the upstream intergenic region, keeping the spacing of the two genes relatively constant. This is measured by the fairly constant size of the 6 Kb Sal I fragment that includes part of the coding and all the upstream flanking the two genes. No large insertions were apparent here, although in the study of Langley et al. (1988) large insertions/deletions were mapped in this region.

DISCUSSION

Restriction fragment Length Polymorphism

With all enzymes used (108 sites) only 9 sites were found to be polymorphic at and around the probed region of the amylase genes (22Kb.). The restriction map variation observed for the amylase region among the fifty strains of D.melanogaster shows many similarities with earlier studies of other gene regions of this species. The structure of the amylase locus is similar to that of the Hsp70 region in that all chromosomes examined have the same basic structure of two divergent transcription units (sharing great sequence similarity) separated by several kilobases of unique sequence (Blackman and Meselson, 1986; Gemmill et al. 1985, Gemmill

et al. 1986, Boer and Hickey 1986). This inverted and divergently transcribed gene duplication might represent an evolved economy of regulatory mechanism, e.g., shared 5' sequences. Or it may simply be a genetically stable gene duplication, e.g., less likely than tandemly duplicated genes to asymmetrically pair leading to unequal exchange (Langley et al. 1988).

The same restriction fragment length polymorphisms are common to a number of strains from different localities. On the other hand, the variation for DNA insertions and deletions appear to be population specific and this may reflect different evolutionary mechanisms or dynamics associated with this latter class of variation.

The positions of insertions/deletions appear to be localized into distinct regions on each side of the two transcriptional units of amylase and is not associated with restriction site polymorphism. This class of variation has also been observed in other tracts of DNA, for example, human haemoglobin (Higgs et al. 1981) and is well known in mutant laboratory stocks of Drosophila. Sometimes the insertional variation can be attributed to transposable elements (Green, 1980; Strobel et al. 1979) and, in turn, these have been identified in association with the laboratory mutants (Bingham 1981). At present the nature and origin of the insertions and deletions in the amylase region is unknown.

The D.melanogaster samples from oceanic island collections don't have any particular pattern of amylase polymorphism. Unlike the strains from West Africa, samples from Fiji, Seychelles and Mauritius all have a single Amy¹ band and have no unique insertion or deletion. This supports the theory that oceanic islands don't have endemic Drosophila populations, and that all Drosophila found there are of recent origin.

The novel feature of this study was the relationship between DNA haplotype and the amylase enzyme electrophoretic pattern. The linkage disequilibrium between restriction sites was strongest for the Eco/Hinf combination. Moreover, this combination of restriction sites show a strong association with electrophoretic enzyme patterns. It is possible, of course, that this association may be a chance event, reflecting the ancestry of the chromosome. For such reasons, levels of linkage disequilibrium might be expected to be particularly high in such a relatively short DNA segment. Nonetheless, the DNA concerned brackets the transcriptional units and whilst any evolutionary, causal relationship between the DNA, marked by the polymorphic restriction sites and the type of amylase electrophoretic pattern is not proven, the association by virtue of its relationship with the enzyme phenotype is a potential target for any form of natural selection capable of acting on the amylase enzyme.

Although no clear relationship emerged between DNA insertions/deletions and amylase enzyme phenotypes, as discussed earlier, whether this class of naturally occurring DNA variation affects the amylase electrophoretic mobility is unknown. However, their effect could be very mildly deleterious if at all. Obviously more detailed survey and in vitro experiments are required. However, this study has shown that a study of naturally occurring variation in DNA may be useful in the study of gene function.

Overall, three main conclusions can be drawn from the results of the restriction analysis. First, much of the variation can be explained by the presence of a limited number of haplotypes which are segregating in natural populations. Each of these haplotypes contains its own particular combination of restriction sites. As a result of this pattern, there is strong linkage disequilibrium between various restriction sites. Secondly, enzyme electrophoretic variants are associated with the DNA-defined haplotypes. Thirdly, the insertion/deletion variants in the amylase region are not associated with either the restriction site variation or with the enzyme polymorphisms. Instead, such variants are characteristic of particular geographic areas.

These results could be interpreted as meaning that the polymorphisms in the amylase region are very old. This would explain their distribution throughout many of the populations

sampled. Presumably, the missing theoretically possible combinations of restriction sites have been lost by genetic drift. The more restricted distribution of the insertion/deletion variants suggests a more recent origin for these variants. It is reasonable to suppose that insertions and deletions also occurred in the more distant past but that they have been removed by selection. This interpretation is consistent with the observation that the existing insertions and deletions all fall outside the coding regions, suggesting even more stringent selection against insertions and deletions within the coding regions.

However, restriction polymorphism studies will not replace the need for direct DNA sequencing. In D.melanogaster, because of the great similarity between the amylase duplicated genes, restriction analysis was found to be not the ideal method of detecting the polymorphism in each gene separately. Hence a choice of certain strains for fine analysis was made, the results of which will be discussed in the next chapter.

CHAPTER 3

INTRASPECIFIC DNA SEQUENCE COMPARISONS

INTRODUCTION

It is obvious that DNA sequence differences between organisms contain a great deal of information about how much evolutionary divergence has occurred since these organisms had a common ancestor. As the mass of sequence information in different species has accumulated, a great deal of attention has been devoted to methods of inferring evolutionary events from observed sequence comparisons.

In this section, I look at intraspecific DNA sequence comparisons and the types of evolutionary inferences that can be drawn from this set of data. Both intraspecific and interspecific sequence comparisons have become important tools in molecular approaches to the study of evolution of gene structure and gene function. Therefore, it is important to understand how the rates of sequence divergence between species relate to genetic polymorphism within species.

All species of the *Melanogaster* group appear to have an amylase gene duplication. These duplicate amylase genes are of particular interest because it is evident that the genes are not coordinately regulated throughout development (Doane 1969b). Also some tropical West African strains are less affected by dietary changes (e.g. less repressibility by dietary glucose). The choice of the West African strain Makokou, Amy^{4,6}, which has a high overall amylase activity, coupled with differential expression of the two amylase isozymes, was influenced by both of these considerations.

It can be reasoned that if the amylase genes are closely linked but not coordinately controlled, then the genetic information for the tissue- and stage-specific regulation of these two genes may be contained within a fairly small identifiable region of DNA. My plan was to identify segments of DNA that might be important in gene regulation by comparing the structure of the locus in the Makokou strain with the published sequence of the Oregon R strain (Boer and Hickey 1986). Both of these strains have the amylase duplication but they differ from one another in a number of significant regulatory features. For instance, the expression of both gene copies in the Oregon R strain is very sensitive to glucose repression. In contrast, not only are both amylase genes in the Makokou strain less sensitive to glucose, but there is also a large difference between the two gene copies

in their degree of sensitivity (Benkel and Hickey, 1986b). It is likely that these regulatory differences have arisen subsequent to the origin of the duplication and, therefore, would represent a fairly recent evolutionary event that might be studied readily at the molecular level.

Finally, the Makokou strain is interesting in that it has an amylase electromorph pattern of the type which is common in equatorial Africa, but is rare elsewhere in the world. Thus, while the previous sequencing reports on D. melanogaster amylases were confined to strains typical of European and North American populations (Boer and Hickey 1986; Okuyama and Yamazaki 1988), this strain provides detailed sequence information for a representative of a distinctly different type of amylase genotype.

MATERIALS AND METHODS

Drosophila Library And Clones

Total genomic DNA was isolated following Davis and Davidson (1984) from the D. melanogaster Makokou, Amy^{4,6}, strain homozygous for the standard second chromosomal arrangement. The library was constructed by ligating genomic DNA, partially cleaved with Sau 3A then treated with calf intestinal phosphatase, into the Bam HI site of a Lambda Dash

phage vector (an EMBL3 phage derivative supplied by Stratagene). The library was screened using two D.melanogaster amylase-specific probes. One of these, a 650 bp PvuII fragment of the pOR-m7 cDNA clone (Benkel et al 1987) contains coding sequences from the 3' end of the gene; and the second probe, pCS-1.4, is a 1.4kb fragment specific for the amylase 5' non-coding region. Three successive screenings of the positive plaques were performed by modification of the procedures of Benton and Davis (1977). Four 140mm X 15mm plates containing approximately 3×10^4 plaque-forming units were screened using amylase probes. Phage DNA was transferred to Biodyne membrane filters, treated with denaturing buffer (0.5 N NaOH/1.5 M NaCl for 5 minutes and then with neutralization buffer [3 M Na-acetate (pH 5.5)] for 5 minutes. The phage DNA was fixed by UV to the membrane. After prehybridization with 50 µg/ml of heat-denatured salmon sperm DNA in prehybridization mix [5X Denhardt's, 5X SSC, 50 mM NaPhosphate (PH 6.5), 0.1% SDS and 50% (v/v) formamide for one hour at 45° C. Phage DNA was hybridized with the homologous probes in the same solution as that for prehybridization for 24 to 36 hours at 45° C. Membranes were washed three times with 0.1% SSC and 0.1% SDS at 65° C. Between two and five positive plaques were identified in each plate. Positive plaques were picked, replated at low density and rescreened twice. Single positive

clones were picked and used to prepare recombinant phage DNA.

Isolation and characterization of Amylase genes

Recombinant phage was grown and DNA was extracted according to the method of Silhavey (1984) by adding a single plaque to 0.5 ml of an overnight culture of bacteria (m.o.i., about 0.01). After incubation for 20 minutes at room temperature, 2 ml of LB broth containing 10mM $MgCl_2$ were added to the bacterial culture followed by shaking for about 6 hours at 37° C. after lysis is complete the culture was treated with 0.1 ml chloroform and was centrifuged at 6000 rpm for 10 minutes. One ml of the supernatant was added to an equal amount of DEAE-cellulose slurry, centrifuged, and 0.8 ml of the supernatant transferred to a clean eppendorf tube. 100 μ l of 5M NaCl and 450 μ l of isopropanol were added to precipitate the bacteriophage particles. The pellet was then resuspended in TE, treated with phenol and precipitated with ethanol. DNA from twelve positive clones was digested with restriction enzymes according to the manufacturers specifications. DNA fractionation by gel electrophoresis, processing for Southern analysis, filter blotting and hybridization is done following standard procedures described by Maniatis, Fritsch and Sambrook (1982).

Genomic DNA from the Makokou, Amy^{4,6}, flies was digested and probed along with the genomic lambda clones. Lambda genomic clone Mk2 which contains both proximal and distal amylase genes was chosen for further analysis. The restriction map of the duplicated amylase genes of the Makokou strain was constructed using both genomic clones and total fly DNA digests.

Subcloning

From the Makokou, Amy^{4,6}, lambda genomic clone MK2, a 6 kb Sal I fragment containing upstream regions of both proximal and distal amylase genes as well as 900 bp coding sequence from each of the duplicated genes was isolated and subcloned.

The Makokou 6 kb SalI fragment was cleaved with EcoRI and fragments of 2.5 kb Sal/Eco (proximal) and 3.5 kb Eco/Sal (distal) were religated in pIBI 24 and pIBI 25 plasmid vectors. Insert DNA and vector DNA in 20 μ l of ligation solution [0.6 mM ATP/ 66mM Tris-HCl (pH 7.6)/ 6.6 mM MgCl₂/ 10 mM dithiothreitol DTT] were incubated overnight with 0.01 unit of T4 DNA ligase at 4° C. Transformation was conducted by the calcium chloride method (Mandel and Higa 1970). An overnight culture of JM101 was added to 100 times the volume of YT medium and allowed to grow for about 2 hours at 37° C.

After the culture had been kept on ice for 10 minutes, the bacterial pellet was washed with 10 ml of ice-cold buffer [50 mM CaCl_2 / 10 mM Tris-HCl, PH 8.0)], suspended in 1/15 the volume of the original culture and kept on ice for 3 to 6 hours. An aliquot of 0.2 ml of the competent cell solution was mixed with 1 μl of ligation mixture and kept on ice for an hour. After incubation for 2 minutes at 42° C, 1 ml of the YT broth was added to the mixture followed by incubation for one hour at 37° C. About 100 μl of reaction solution were spread on the YT agar plates containing 50 $\mu\text{g/ml}$ Ampicillin, 30 $\mu\text{g/ml}$ X-gal and 2.5 mg/ml IPTG (Maniatis, Fritsch and Sambrook 1982).

Plasmid DNA preparation

pIBI vectors can be propagated as a double-stranded DNA plasmid or, upon superinfection with helper phage, as a single-stranded DNA phage (IBI INC.).

Double-Stranded DNA Preparation

An aliquot of 1.5 ml of an overnight culture was centrifuged and the collected bacteria suspended in 350 μl of [8% sucrose, 0.5% Triton X-100, 50 mM EDTA and 10 mM Tris-HCl (PH 8.0)]. a volume of 250 μl of freshly prepared

lysozyme solution (10mg/ml in 10 mM Tris-HCl) was mixed with the bacterial solution and incubated for 40 seconds in a boiling-water bath. The mixture was centrifuged for 15 minutes and the pellet was removed with a toothpick. The plasmid was precipitated with 40 μ l of 2.5 sodium acetate and 420 μ l of isopropanol. The pellet was resuspended in TE, treated with RNAase, Proteinase K and phenol. Plasmid DNA was then precipitated with ethanol.

Single-Stranded DNA preparation

One colony of insert-containing plasmid was grown in YT broth for 30 minutes. An aliquot of 50 μ l helper phage, M13 K07 (1×10^{11} pfu/ml) was added and left to grow for another 30 minutes. Kanamycin was added (50 mg/ml) and the culture was left to grow for 14-18 hours. Cells were harvested by centrifugation and single-stranded plasmid precipitated with 25 μ l of 3.5M NaCl₂ and 20% PEG solution. The pellet was dissolved in 200 μ l TE buffer and treated with phenol. The single-stranded DNA was precipitated with ethanol and dissolved in 20 μ l deionized water. One μ l of this DNA template was then mixed with 1 ng of the appropriate primer and split four ways for each separate sequencing reaction.

DNA Sequencing

Subclones were grown single-stranded and sequenced according to the chain termination reaction method of Sanger et al. (1977) using Sequenase (United States Biochemical Co.) and [α - ^{35}S dATP]. M13 universal primer and a number of 20 bp custom-made oligonucleotide primers were used to sequence the two strands. Generally, each sequencing ladder was determined to 450-500 bp from the labelled end. This required double loadings of 50 cm, 8% gradient gels performed in an LKB Macrophor electrophoresis unit. Two-hour and four-hour runs at 2200V were found to produce a continuous ladder. Following electrophoresis, gels were soaked in a 10% acetic acid solution for 30 minutes, air dried with a hair drier, wrapped in saran wrap and autoradiographed using Kodak or Trilite X-ray film without intensifying screens at -70°C .

Analysis of DNA sequences

Confirmation of sequences was obtained by sequencing both strands whenever possible, and by repeated sequencing. Sequence data were entered directly into an IBM computer using a sonic digitizer. Alignment of the sequence and subsequent analysis is done using the Microgenie (Beckman) computer programs (Queen and Korn 1984). The nature of the

sequenced genes was established unambiguously by comparison with the complete sequence of the D.melanogaster OR-R strain proximal locus (Boer and Hickey 1986) and the sequence of the TN-329 strain distal locus of Okuyama and Yamazaki (1989).

RESULTS

Isolation of Amylase clones

The library screened yielded twelve overlapping phage clones homologous to the D.melanogaster amylase probes. Of the twelve lambda clones isolated three had the proximal gene only, four had the distal gene only, three clones had the proximal and part of the distal, while two clones contained the complete two genes and their flanking regions (Figure 3.1 and Figure 3.2). Clone MK2 is a representative with the complete two coding sequences and their flanking regions.

Characterization of amylase clones

Restriction map of the lambda MK2 is shown (Figure 3.3). The complete restriction map of the duplicated genes was shown in Chapter 2. Inspection of the restriction map shows that the two genes have the same arrangement as

Figure 3.1. Nine positive clones obtained from the Lambda-Dash genomic library of the Makokou strain. Clones were digested with Bam HI and with Eco RI and probed with pCS-1.4, a fragment specific for the proximal amylase gene. Probing was at 45° C with 50% formamide and washing at 65° C.

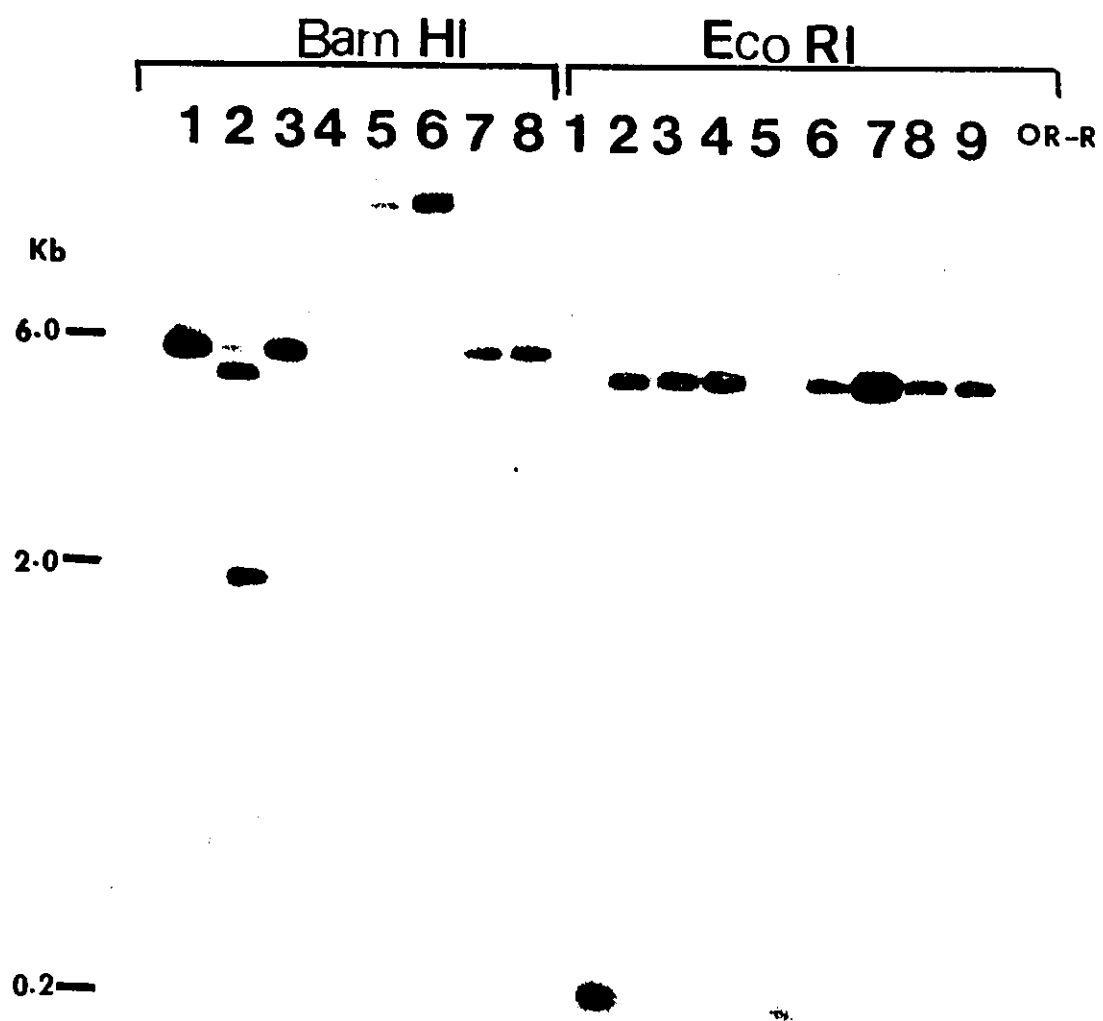


Figure 3.2. Positive clones as in Figure 3.1 probed with pOR-m7 Pvu II fragment that is homologous to the coding sequences of both proximal and distal amylase genes. Arrow indicates genomic clone MK2 used for construction of the map and subsequent subcloning. Note that this clone had both genes and their flanking regions. OR-R is lambda clone described by Benkel et al.(1987); it is used here as an internal control and size marker.

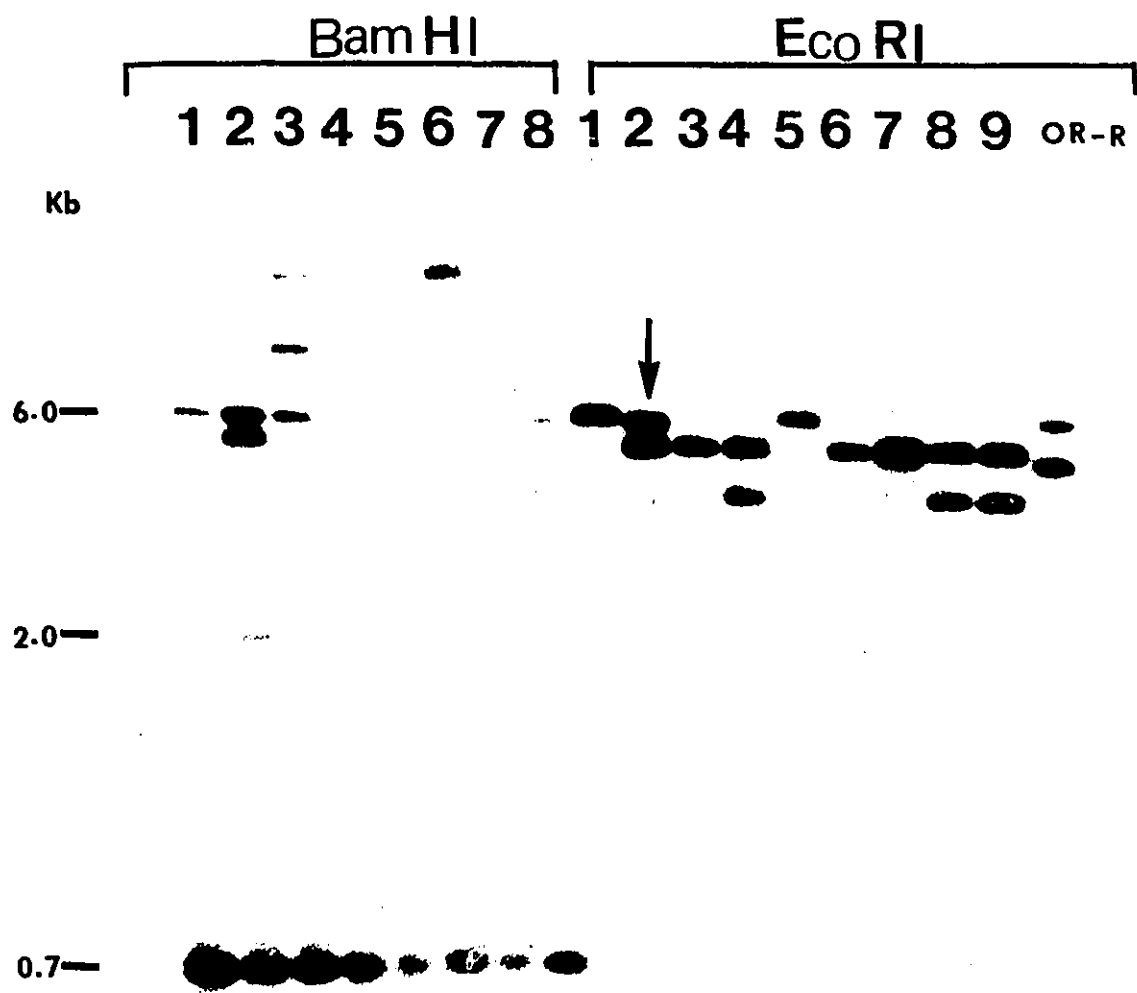
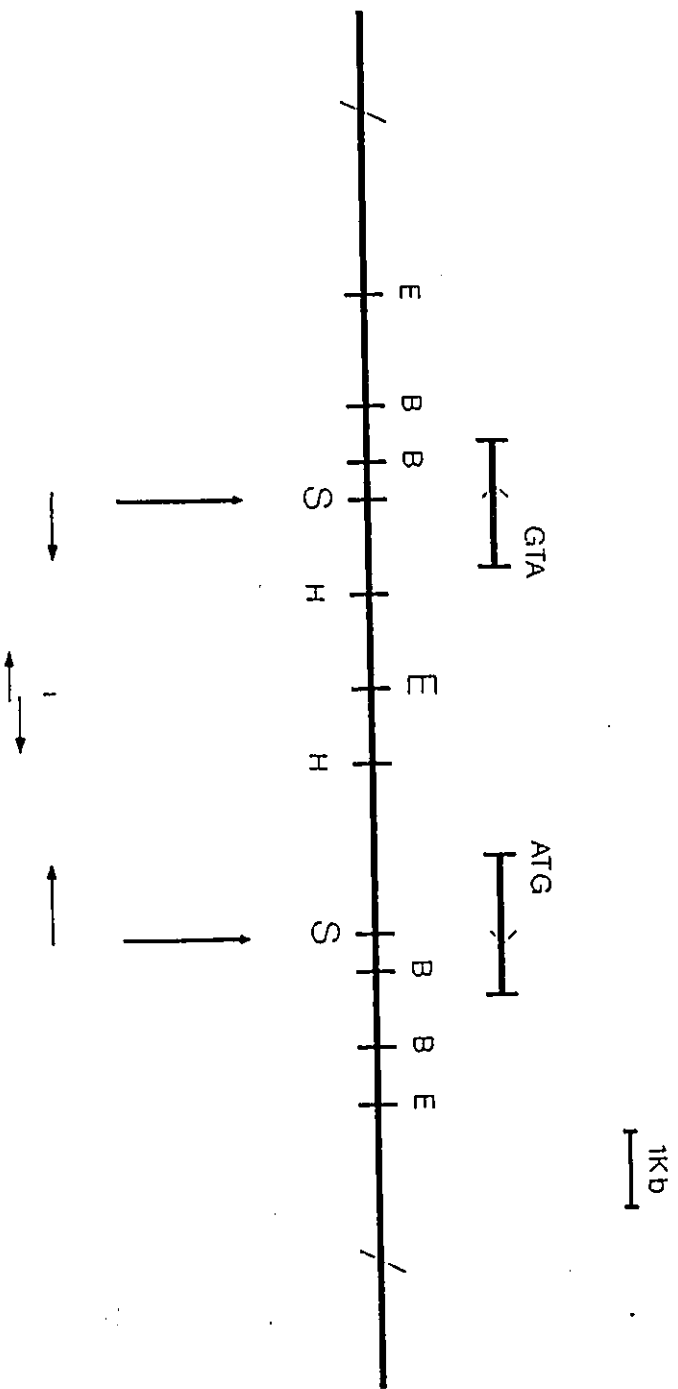


Figure 3.3. Restriction map of clone MK2.

This clone was isolated from a genomic library prepared from the wildtype D. melanogaster strain, Makokou Amy^{4,6}. This clone includes the duplicated amylase genes. The coordinate system used in the Figure is that of Benkel et al 1987. Restriction sites are designated as follows: B, Bam H1; E, Eco R1; H, Hind III; S, Sal I. Horizontal arrows above indicate the direction of transcription for each coding sequence. Vertical arrows point to the Sal-Sal fragment used for sequencing. Horizontal arrows below show direction of sequencing from ends of Sal and Eco fragments. Sequencing strategy is shown at the bottom.

Makokou^{4,6}



that of the Oregon R strain. The orientation and arrangement was studied by probing blots of cloned DNA from Makokou, Amy^{4,6}, with 3' and 5' subclones of the Oregon R amylase cDNA (see Figures 3.1 and 3.2). The duplicated genes are divergently transcribed, a feature typical of the melanogaster species subgroup (Boer and Hickey 1986, Gemmill et al. 1986, Payant et al. 1988).

Structure of the Duplicated Genes of the Makokou, Amy^{4,6}, Strain.

Site Variation and length polymorphism

The Coding Region

Nine hundred bp of sequence from the coding region were sequenced, as well as 300 bp from the 5' non-transcribed and non-translated region; these regions were sequenced from both the proximal and distal genes. Figure 3.4 shows regions of both proximal and distal genes aligned with the Oregon R sequence (Boer and Hickey 1986).

Comparing the sequence of the proximal genes from the standard Oregon R strain of D.melanogaster which has an Amy¹ isozyme and that of the Makokou strain with an Amy⁴ isozyme, in the coding region of the two strains there are

6 nucleotide differences out of 822 (0.7%). Five out of the six nucleotide differences in the coding region are transitions while one is a transversion.

The distal gene of the Makokou strain is compared to the sequence available from the Oregon R distal gene and the published sequence of the distal gene of the TN-329 (Okuyama and Yamazaki 1988). The Makokou gene has different electrophoretic mobility i.e. different isozymes from that of OR-R and the TN-329 (both have Amy¹, while the Makokou has an Amy⁶). There are 11 nucleotide differences out of 791 (1.6%) between the two coding regions of the Makokou and the TN-329.

When the two coding sequences from the Makokou strain were compared with one another, 14 nucleotide differences out of a 900 bp region were noticed (1.5%). In addition to the nucleotide site variation the duplicated genes have a region of three nucleotide insertion/deletion (see Figure 3.4). This means that the protein product from the distal gene is one amino acid residue shorter than that produced by other amylase genes in D.melanogaster.

The 5' flanking region

In both proximal and distal Makokou genes the classical regulatory motifs are conserved and they lie at the exact

distance from the transcription initiation site as that of the Oregon R strain. However, the two genes have a stretch of 19 mismatches just before the TATA box including one insertion/deletion. When the proximal genes of the Makokou and OR-R are compared upstream of the initiation site, 13 nucleotide differences out of 290 (4%) were detected. Of these three nucleotide differences are due to transition, three are due to transversions and seven are a result of a single insertion/deletion. Seven of these differences disrupt the conserved sequences at position -200 (see Figure 3.4); a conservation of this region between the Drosophila amylase and the glucose-repressible Adh III gene of yeast was noted by Boer and Hickey (1986). However, this region is not disrupted in the distal gene of the Makokou strain.

Figure 3.4. Nucleotide sequence of Makokou 5' region from the proximal locus aligned with that from the Makokou distal gene, and also with the sequence of the Oregon R proximal locus (Boer and Hickey 1986). +1 is the first nucleotide of the start codon. The conservative glucose regulatory box between Drosophila and yeast is highlighted; the codon which is missing in the Makokou distal locus is highlighted. The sequence from the Makokou proximal gene is shown in full on top; the sequences of the Makokou distal gene and the Oregon R gene are indicated by dashes where the nucleotides are identical to the Makokou proximal sequence; nucleotide differences only are shown in full. Missing nucleotides are represented by dots. Arrow indicates initiation of transcription.

-291

TTTCCCGATGAGTTATTGATACAAATATAACGAAAATAAGCCGACTCACTAATCATCAGC
 A-AAT- ---A--CACC-AT-TG-GGA-T---TTG- C-TA--A--TTAGT--GGG-A-

-231

GAAAAATTGCCATCTCCAGTCTGC.....TCGGGAAATTCTGATTAGACAACTAATCG
 T--TT--AATT-C-A-----AATAGGTCTG-CCTGAGC-AA---A-C--GCCT-----
 -----AATACGTCTG-TC-G---G-----T-----

-171

CCAGTCAGACCCCATGCGCGAAAAAACCCCTTAGGGAGCGATAAGATCCCATGCAGTCAC
 -G-T-G-C-A-T--ATACAT-C--G-GAA-G---CA-TTT-GGG---AAG--CA-T--G-
 -----T-----

-111

AAATCACTCCCCGCGAAGACCTCAGATAAAGTAGCAGTGGGGTCCACTATATAAGGAGCG
 -----G--A---C...-G-...-GTT---TCAT-----C---T
 -----C-----

-51

GCTCTGAGTAGTTCCGACCAGAGTGAAACTGAACTTCCATCTGGAATCATCATGTTTCTG
 --A----CA-C--TTA--T-----T-T-G-----C--AA-----

GCCAAGAGCATAGTGTGCCTCGCCCTCCTGGCGGTGGCCAACGCCCAATTCGACACCAAC

TACGCATCCGGTCGTAGTGGAATGGTCCACCTCTTCGAGTGGAAGTGGGACGACATCGCT

GCCGAGTGCGAAAACCTTCCTTGGACCCAATGGCTACGCCGGTGTTCAAGGTCTCCCCTGTG

AACGAGAACGCCGTCAAGGACAGCCGCCCTGGTGGGAACGTTACCAGCCCATCTCCTAC

AAGCTGGAGACCCGCTCCGGAAACGAAGAGCAGTTCGCCAGCATGGTCAAGCGCTGCAAC

GCCGTCCGGAGTGCGCACCTACGTGGACGTGGTCTTCAACCACATGGCCGCCGGCGGAGGC

A-----

ACCTACGGCACTGGCGGCAGCACCGCCAGCCCCAGCAGCAAGAGCTATCCCGGAGTGCCC
 -----C-----

TACTCCTCGCTGGACTTCAACCCGACCTGCGCCATCAGCAACTACAACGACGCCAACGAG
 -----G-----

GTGCGCAACTGCGAGCTGGTCTGCGCGACCTTAACCAGGGCAACTCCTACGTGCAG
 -----C-----

GACAAGGTGGTCGAGTTCCTGGACCATCTGATTGATCTCGGCGTGGCCGGATTCCACGTG
 -----C---...-----
 -----G-----

GACGCCGCCAAGCACATGTGGCCCGCCGACCTGGCCGTCATCTATGGCCGCCTCAAGAAC
 -----T-----

CTAAACACCGACCACGGCTTCGCCTCGGGATCCAAGGCGTACATCGTCCAGGAGGTCATC

GACATGGGCGGCGAGGCCATCAGCAAGTCAGAGTACACCGGACTGGGCGCCATCACCGAG
 -----C-----
 -----C-----

TTCCGCCACTCCGACTCCATCGGCAAGGTCTTCCGCGGCAAGAACCAGCTGCAGTATCTG

 -----G-----C-----

ACCAACTGGGGCACCGCCTGGGGCTTCGCCGCCTCCGACCGCTCCCTGGTATTCGTCGAC

 -----T-----

--- Makokou^{4,6} Proximal
 --- Makokou^{4,6} Distal
 --- Oregon-R Proximal

Sal I

When the upstream non-coding region of the Makokou distal gene is compared with that of the Oregon R (Amy¹) and the Carlton-S (Amy³), there are 32 nucleotide differences in a region of 638 bp (5%) between the Makokou, Amy^{4,6} and either of the other two strains.

DISCUSSION

It is clear from this study and previous studies that in the several strains and species that have been analyzed, there is a constancy of structure in the amylase region. In the Makokou^{4,6} as in the OR-R (Boer and Hickey 1987) the two amylase coding regions show a very high degree of sequence similarity while the sequence of the 5' upstream regions are much more diverged.

However, unlike the Oregon R strain, the Makokou has more sequence variation between its proximal and distal coding sequences (1.5%). This suggests that more mutations have accumulated between the two genes of this African strain than between the Oregon R coding sequences. Still, the 1.5% difference between the two genes is less than that between either gene copy and their counterparts from other strains.

The small size of the region that separates the two amylase genes in D. melanogaster (Boer and Hickey 1987, Payant et al. 1988, Hickey et al. 1989, Langley et al. 1988) is too short to expect recombinations to be readily recovered. Since the duplicated genes are so closely linked, it is likely that the amount of information that was originally duplicated is fairly small. Nevertheless, an inversion between the two amylase genes of a null-strain (Hickey et al. 1989) appears to be a result of homologous recombination.

A surprising finding was that, despite the overall similarity in the sequences of the two coding regions, there is a deletion of an entire codon in the distal gene. Such insertions and deletions of codons are relatively frequent in interspecific comparisons, including those of amylase sequences (see Hickey et al., 1989), but they have not been reported previously in comparisons of genetic variants within a single species. There is no information available as to how these insertions /deletions are generated. The fact that they are multiples of three nucleotides is probably due to purifying selection removing those length variants which alter the reading frame. This implies that such regions where codons are added/deleted are susceptible to certain mutational effects.

It is possible that the deletion of the 7 bp at the conserved glucose box is related to the change in glucose repression of the Makokou, Amy^{4,6}, proximal gene that has occurred during the evolution of this strain. It is also possible that the two genes share one regulatory glucose box, since both genes are less repressed by glucose compared to Oregon R for example. A final feature of the nucleotide differences in this region that cannot be explained by the accumulation of new mutations is that these nucleotide differences are clustered in one region of 19 bp. This clustered change is not associated with continuous runs of the same nucleotide, or short tandem repeats, which are known to give rise to high mutation rates by slippage mechanisms (Efstratiadis et al. 1980; Jones and Kafatos 1982; Brown and Piechacsyk 1983).

CHAPTER 4

CONCERTED EVOLUTION OF THE DUPLICATED AMYLASE CODING SEQUENCES

Introduction

The evolution of multigene families undergoing gene conversion has been extensively studied, see Ohta (1985, 1986), Shimizu (1985, 1987), Weir (1986), Nagylaki and Barton (1986), Walsh (1986, 1987, 1988), Kaplan and Hudson (1987) and Tachida (1987). Recently, results emerging from work with multigene families demonstrated intrachromosomal recombination between duplicated DNA sequences. It is shown that intrachromosomal recombination can lead to gene conversion by heteroduplex repair. Other experiments have shown that there is a bias in heteroduplex repair which favours the formation of GC base pairs (Brown and Jiricny, 1988). The results of these experiments suggest that, during the course of evolution, duplicated genes would undergo repeated rounds of gene conversion and that, in so doing, they would become increasingly GC-rich. This latter effect would bias the codon usage within such genes. The work to be presented in this chapter supports these predictions.

The amylase genes of the fruit flies, genus Drosophila, offer an excellent system for studying the manner in which

gene conversions affect the evolution of a small multigene family. When the two copies of the alpha-amylase gene from strains of D. melanogaster were compared (Boer and Hickey 1986; Okuyama and Yamazaki 1988), the two coding regions were found to be very similar. This, in addition to the close linkage of the two copies, raised the possibility of sequence recombinations between the two copies of the gene.

Materials and Methods

Subcloning and Sequencing

The 6.4 Kb Sal I fragment of Payant et al. (1988) from Drosophila erecta, which contained the 5' region of both copies of the amylase gene, was cleaved out of the lambda clone with Sal I and removed out of a low-melting agarose gel. The fragment was further cut with Hind III and subcloned into pIBI 24 and pIBI 25 plasmid vectors. The resulting subclones (4 Kb from the proximal locus and 2.4 Kb from the distal locus) were sequenced by Laure Bally-Cuif (Bally-Cuif et al. 1989).

Two D. melanogaster Oregon R clones isolated by Benkel et al. (1987) from a phage library and subcloned in pUC plasmids were chosen for analysis of downstream, non-coding regions. Clones Lambda OR-5.2 (proximal), and Lambda OR-5.5 (distal)

were cleaved with EcoR I and Bam HI. The Bam/Eco fragments were subcloned into pIBI vectors according to the procedures described in Chapter 3 above.

Results

The 3' region of the proximal and distal genes of the OR-R strain were sequenced for a total of 300 bp and aligned with each other (Figure 4.1). The coding sequences had the highest similarities between the two genes (Figure 4.1 and Boer and Hickey 1986). The similarity drops immediately after the stop codon. The 3' downstream sequence from the two genes loses similarity immediately after the stop codon, except for poly-A addition motifs which remain conserved in sequence and position. Thus, the sequence divergence which was noted by Boer and Hickey (1986) for the upstream flanking regions is also true for the downstream, non-coding regions.

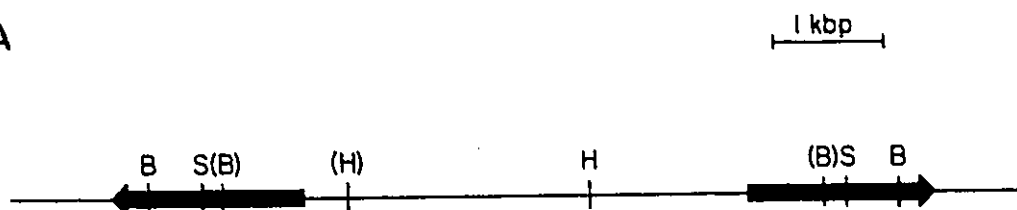
Figure 4.1. Comparison of the duplicated amylase-coding genes in Drosophila melanogaster.

A. Arrangement of the duplicated coding sequences. The two genes (shown by arrows) are approximately 4kb apart and are divergently transcribed. The proximal and distal gene copies are shown on the left and right of the Figure, respectively. Restriction enzyme sites for Bam HI (B), Hind III (H), and Sal I (S) are shown. Those sites which are not shared with the D. erecta sequences are shown in brackets.

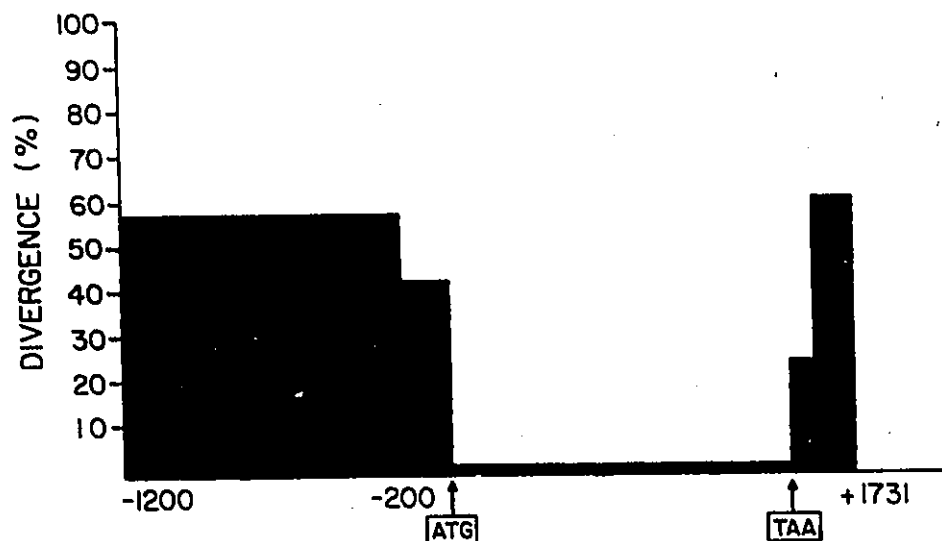
B. The percentage sequence divergence between the duplicated genes. Three regions were compared: (i) upstream, non-coding; (ii) coding; and (iii) downstream non-coding. The positions of the start and stop codons are indicated.

C. Alignment of sequences flanking the coding regions of the duplicated amylase genes. This alignment illustrates the relatively sharp boundary between the conserved transcribed regions and the divergent flanking regions. Regulatory motifs (CAAT and TATA) are boxed, along with the start and stop codons. Initiation of transcription is indicated by an arrow. The polyadenylation site is indicated by an inverted triangle. Sequence identities are indicated by dashes; gaps are shown as asterisks. All analyses were done using the Beckman Microgenie sequence analysis programs.

A



B



C

```

-200 CTGGAAATTGTGATTGACAACTAATCGCCAGTCA**GACCCCATGCGTGAAAAAACCC*****CTTAGGGAGCGATAA
    -G-C--GC-AA--AAC--GCCT-----G-T-T-ATC-A-T--ATACAT-C--G-GAACGTAGCGA-----*-----
GATAAGATCCCATGTCAGTCATCAATCACTCCCCCGGAAGCCCTCAGATAAAGTAGCAGTGGGGTCCACATATATATAGGAGCGGC*
    -A-----G-----G-----A-----*-----GTT--TCATC*-----C--T--A
          ↓
TCTGAGTAGTTCGGACCAGAGTGAAACTGAACTTCCATCTGGAATCATCATGTTTCTGGCCAAGAGCATAGTGTGCCTCGCCCT
    -GACAC-C--TTA-T-----T-T-G-----C-AA-----T-----
          +1
CCTGGCGGTGGCCAACGCCCAATTGACACCAACTACGCATCCGGTCGTAGTGGAAATGGTCCACCTC ( .... 1.3kb ....
    -T-----T----- ( .... 1.3kb ....

... ) GGATCCGACGGACGGGCTTCCATCAACATGGCAGCTCCGAGGACGACGGAGTGCTGGCCATTACCGTCAACGCCAAGTT
... ) -----T-----
stop
GTAACAGCTGGG*****GAGCATGG*CGAACAGCCAGGCAATTAATTGAGATTATTAATTCTACGAAATATATATGATGAGATT
    -A-----GCAAA--C-A--C--A--GA-----C-----C--*TTA-----T--C--
ATAAACACACAACACTTTTATTTCGCAAGGGATGATAAGATCTAATATATATATTAT*CTGGGCTAAGCTGA - PROXIMAL
    -G--A-AGGAAGA-G-GC--TA--C--*--TA-AT--T--GG-AA--G--TAAT-TTAT--T - DISTAL
  
```

Discussion

The work described in this chapter was motivated by certain observations of the near identity of the duplicated amylase genes (Boer and Hickey 1986) and by a possibility of gene conversion (Payant et al. 1988; Bally-Cuif et al. 1989; Hickey et al. 1989). In using sequences to estimate the times since the duplication event, a paradox was encountered: the duplication appeared to be recent if one compared coding regions; but when upstream flanking sequences were compared the divergence of the two genes seemed to be older. However, the presence of the duplication in seven related Drosophila species pointed towards an older estimate of the duplication (Payant et al. 1988), which predated the speciation events in the melanogaster species subgroup. This suggested the possibility of concerted evolution of the duplicated genes.

The only effective test of this model for the evolution of the two amylase loci would be to examine the accumulation of nucleotide changes and recombination events in the duplicated genes with time. If most changes which have accumulated can be explained by conversions, i.e., are not mutations new to the loci, then sequence transfer must be frequent relative to the accumulation of new mutations. Furthermore, if these changes are not associated with cross

over events in flanking regions, then an intrachromosomal conversion model is supported.

The restriction map for the alpha-amylase locus in each of the eight species of the melanogaster subgroup (Payant et al. 1988) showed that the general pattern of restriction sites was conserved between the species, but that a Bam HI site located in the coding region was absent from the two copies of the gene in both D.erecta and D.teissieri. The mutation of this Bam HI site proved to be the same in the two copies of D.erecta (Bally-Cuif et al 1989) and it is likely that it is the same in D. teissieri. Not only that, but the two gene copies of D.erecta were found to be 100% similar even for the silent substitutions that occurred since the divergence of D.erecta and D.melanogaster (see Figure 4.2).

Figure 4.2. Comparison of amylase sequences from D. melanogaster and D. erecta. Sequences from both D. erecta genes are aligned with the proximal gene of D. melanogaster. The alignment extends from nucleotide position -200 to position +909. Dashes represent sequence identity with the D. melanogaster sequence; asterisks denote sequence gaps. The start codon, ATG is boxed. Those substitutions (with respect to the D. melanogaster gene) which are shared by the two D. erecta genes are also boxed. The initiation of transcription is shown by an arrow. The single nucleotide difference between the two D. erecta coding sequences is indicated by an inverted triangle. (from Bally-Cuif et al., 1989)

CTCGGAATTGTGATTGACAACTAATCGCCAGTCAGACCCCATGCGTGAAAAA...CCCCTTAGGGAGCGAT
 -----C-----TAAA-----AA-----
 -C-C--C-CA--AAC-TGTCT-----G-T-T-AT-AA--CAA-AC-T-T***AG-----

AAGATCCCATGCAGTCACAAATCACTCCCCGGAAGCCCTCAGATAAAGTAGCAGTGGGGTCC*****ACTATATAAGGAGCGG
 -----GC-----A-----GGGTCC-----A-----
 -----A*****TC-----T-A--GAATG--TT-----CAT*****CA-----GT-

↓
 C•TCTGAGTAGTTCGACCAAGAGTGAACTGAACTTCCATCTGGAATCATCATGTTTCTGGCCAAGAGCATAGTGTGCCTCGCCC
 --*A--C-A-G-----G-----A-----
 -A--AAC-CGC--TT--T-----G-----

TCCTGGCGGTGGCCAACGCCAATTGACACCAACTACGCATCCGGTCTAGTGGAAATGGTCCACCTCTTCGAGTGGAAAGTGGGA
 -----G-----G-----G-----G-----

CGACATCGCTGCCGAGTGGCAAACTTCTTGGACCCAATGGCTACGCCGGTGTTCAGGTCTCCCCCTGTGAACGAGAACGCCGTC
 -----G-----G-----G-----G-----G-----

AAGGACAGCCGCGCCTGGTGGGAACGTTACCAGCCCATCTCTTACAAGCTGGAGACCCGCTCCGGAACGAAGAGCAGTTCGCCA
 -----G-----G-----G-----G-----G-----

GCATGGTCAAGCGCTGCAACGCCGTCGGAGTGCGCACCTACGTGGACGTGGTCTTCAACCACATGGCCGCCGACGGAGGCACCTA
 -----G-----G-----G-----G-----G-----

CGGCACCTGGCGGCAGCACCGCCAGCCCGAGCAGCAAGAGCTATCCCGGAGTGCCCTACTCTCGCTGGACTTCAACCCGACCTGC
 -----A-----G-----G-----G-----G-----

GCCATCAGCAACTACAACGACGCCAAGGAGTGGCAACTGCCGAGTGGTCTGGCGGACCTTAACGAGGCAACTCCTACG
 -----G-----G-----G-----G-----G-----

TGCAGGACAAGGTGGTTCGAGTCTCTGGACCATCTGATTGATCTCGGCGTGGCCGGATTCCGCGTGGACGCCCAAGCACATGTG
 -----G-----G-----G-----G-----G-----

GCCCGCCGACCTGGCCGTCATCTATGGCCGCTCAAGAACCTAAACACCGACCGGCTTCGCCTCGGGATCCAAGGCGTACATC
 -----G-----G-----G-----G-----G-----

GTCCAGGAGGTCACTGACATGGGCGCGAGGCCATCAGCAAGTCCGAGTACACCGACTGGGCGCCATCACCGAGTTCGGCCACT
 -----G-----G-----G-----G-----G-----

CCGACTCCATCGGCAAGGTCTTCCGCGGCAAGGACCAGCTGCAGTACCTGACCAACTGGGGCACCGCCTGGGGCTTCGCTGCCTC
 -----G-----G-----G-----G-----G-----

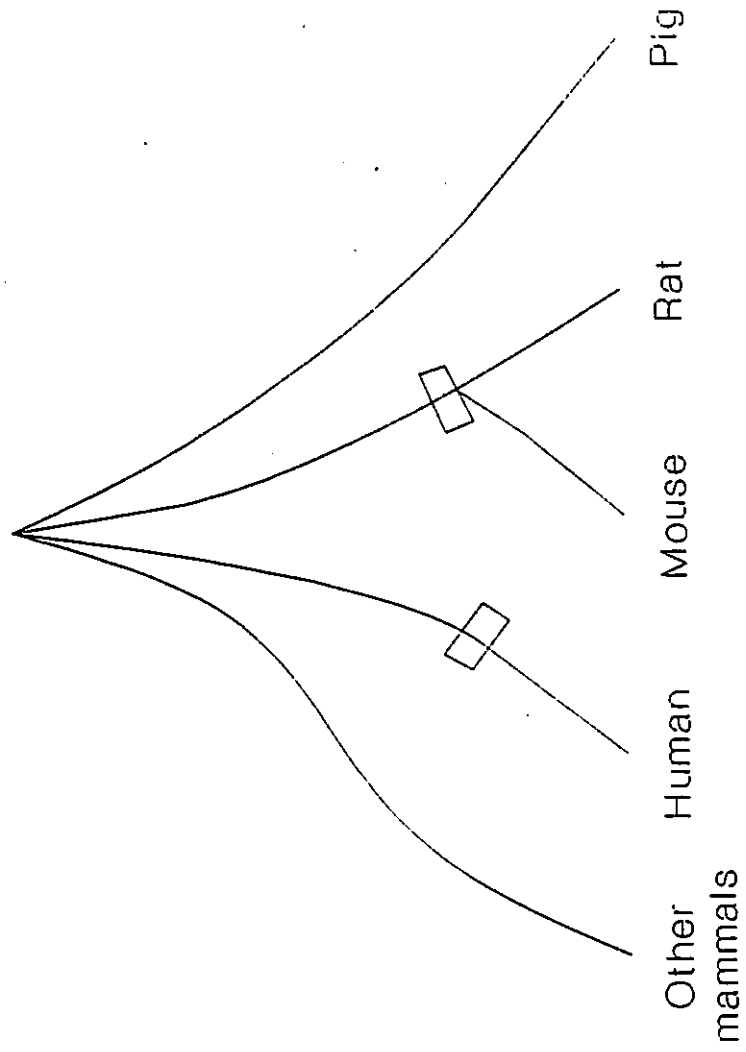
CGACCGCTCCCTGGTATTCGTCGAC
 -----G-----G-----G-----G-----G-----

D. MELANOGASTER, PROXIMAL
 D. ERECTA, PROXIMAL
 D. ERECTA, DISTAL

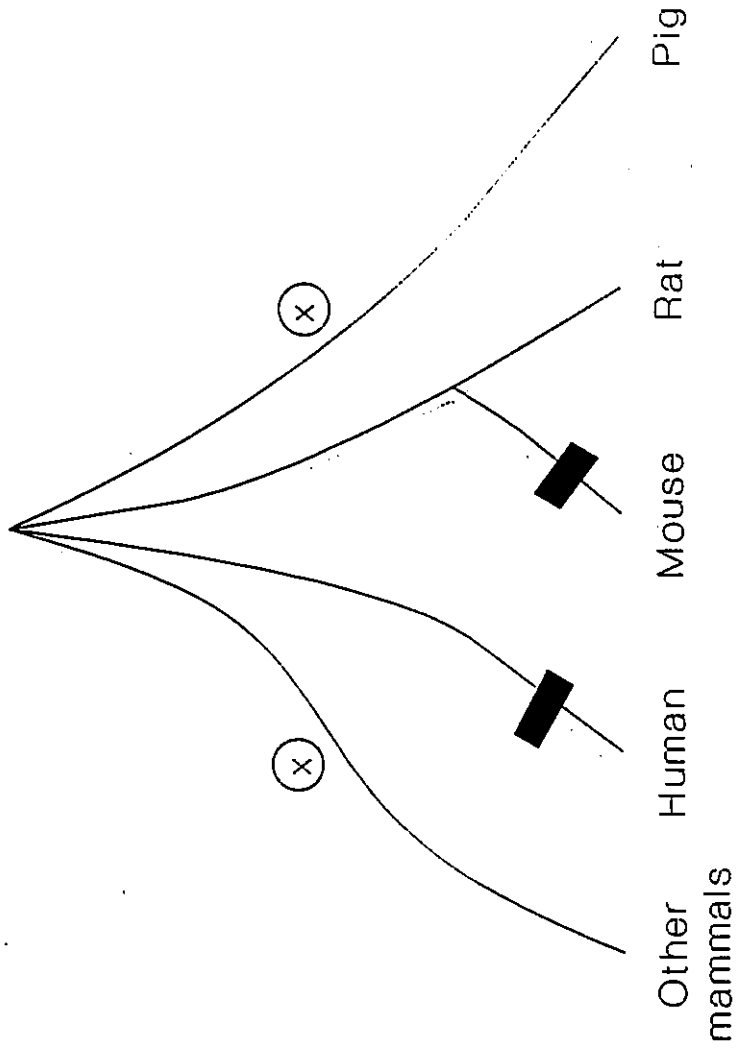
Nakamura et al. (1984) observed that the nucleotide sequences of the human salivary and pancreatic amylases are 94% homologous. In contrast, the homology between the human and mouse pancreatic mRNAs is only 83%, and the homology of the human and mouse salivary mRNAs is 82%. The remarkable degree of similarity between the human salivary and pancreatic mRNAs suggests that recent molecular events, such as gene conversion or gene duplication, must be responsible for "correcting" the sequences. Just as we have shown here in the case of the melanogaster subgroup, the mouse amylase genes also appear to have been "corrected", since the mouse pancreatic amylase gene sequence is more homologous to the mouse salivary gene (90%) than to the human pancreatic gene (83%). Meisler and Gumucio (1986) presented two alternative hypotheses for the evolution of amylase genes from mammals (Figure 4.3). The hypothetical mammalian ancestor in Figure 4.3 (A) possesses only pancreatic amylase gene(s). In this model, salivary

Figure 4.3. Two models for the recent evolution of the mammalian salivary amylase genes. Solid lines trace the descent of the pancreatic amylase genes; dotted lines represent salivary amylase genes. Open box, gene duplication; closed box, gene conversion; X, gene inactivation. (A): gene duplication with aquisition of salivary-specific expression; (B): gene conversions and multiple gene inactivations. Reproduced from Meisler and Gumucio (1986).

Hypothetical ancestral
mammal with active
pancreatic amylase gene(s)



Hypothetical ancestral
mammal with active
salivary and pancreatic
amylase genes



amylase genes arose from recent independent gene duplication. The acquisition of salivary-specific expression by one of the duplicated gene copies is required by this model. Information regarding conservation of structure between the mouse and human Amy-1 gene argues against an independent origin of the two genes. In (B) the hypothetical ancestor possesses genes for salivary and pancreatic amylase, and multiple independent gene inactivation events account for the absence of salivary amylase in most mammalian lineages (Meisler and Gumucio 1986). The intra-species homogeneity of salivary and pancreatic amylase of mouse and man would then be explained by recent gene conversion events. However, the available evidence is insufficient to distinguish between these models.

Unlike the case of the conversion hypothesis of mammalian amylases, the DNA sequences from non-coding regions (upstream and downstream) of the *Drosophila* genes are now available and hence the interconversion of *Drosophila* amylase genes is better substantiated.

In summary, in *D.melanogaster* the coding sequences of the two genes are quite similar, but the similarity drops off quickly both upstream and downstream of the coding region. The most probable explanation is that the duplicated coding sequences can loop over to facilitate mitotic or meiotic recombination and gene conversion through heteroduplex repair (Hickey et al. 1989; Leigh-Brown and Ish-Horowicz 1981).

Genes that undergo sequence conversion are likely to have a GC biased codon usage (Brown and Jirncy, 1988). The *Drosophila* species group is 90% GC rich in the third position (see Chapter 5). One would like to evaluate this GC bias in the light of the GC content of non-duplicated amylase genes. This will be discussed in the next chapter.

CHAPTER 5

LONG-TERM INTERSPECIFIC SEQUENCE COMPARISONS

INTRODUCTION

The rate of synonymous nucleotide change in the coding regions of genes has become a crucial measurement in the study of molecular evolution. The reasons are threefold: precise alignment of coding regions is possible, these silent positions evolve at relatively high rates, and their rates of evolution, it has been argued, approach the selectively neutral rate. Species-level comparisons of DNA sequences will soon represent a large sample of genetic loci in Drosophila. In addition to addressing questions of synonymous position evolution, these data are providing details on the rates of nucleotide substitution experienced at different nucleotide positions within a locus, and on the different forms of selection operating to produce the observed levels of substitution, e.g., selection at the protein sequence level, at the codon level (e.g., codon bias), and at the nucleotide

level (e.g., transition/transversion bias).

There has been increasing acceptance of the hypothesis that molecular divergence is linear with time (Miyata et al. 1980; Hayashida and Miyata 1983; Nei 1987), although not without some resistance (Gillespie 1986). This concept is of particular importance to the application of certain models of molecular evolution. It is necessary both to establish the existence of clock-like behaviour in nucleotide substitutions and to provide the appropriate experimental model (e.g. pseudogenes, silent sites, or intron positions) with which to measure the underlying rate of substitution for a chromosomal region. The assumption is, then, that this rate represents the best approximation of the neutral mutation rate. Positions that deviate from this rate can be examined to establish which forces, e.g., natural selection or drift, are responsible for the observed deviation.

The alcohol dehydrogenase locus, Adh, has been the subject of numerous species-level sequence comparisons (Bodmer and Ashburner 1984; Cohn 1985; Coyne and Kreitman 1986; Schaeffer and Aquadro 1987). These data have substantiated previous reports (Kafatos et al. 1977; Jukes and King 1979) that rates of evolution for functionally distinct nucleotide positions (amino acid replacement sites, synonymous sites and intron positions) are very different. Patterns of substitution across the Adh exons reveal that all

replacement sites and all synonymous sites do not experience the same levels of mutation and/or substitution i.e., the observed changes are clustered in certain exons. Finally, comparison of Adh and a closely linked, uncharacterized gene 3' to Adh shows that the rates of substitution at replacement, synonymous and intron positions vary at different loci within the same chromosomal region (Schaffer and Aquadro 1987). It is necessary to determine the generality of the results obtained at the Adh locus, where the majority of Drosophila species comparisons have been made.

I thought it of interest to compare sequences of the amylase genes in different Drosophila species. This should allow recognition of functionally-constrained sequences specific for the particular gene (proximal or distal) as well as those common to both genes. Such comparisons also serve to characterize the nature and rate of evolutionary change in various domains of the gene.

Ideally, one would like to compare species that have evolved independently for periods sufficiently long to allow nearly complete divergence of unconstrained sequences, but not for very much longer. Constrained sequences could then be recognized against a divergent background. The resolution and certainty with which such sequences can be recognized will be enhanced by including more than two divergent species

in the comparison. The earliest Drosophila radiation is thought to have occurred about 0.5×10^8 years ago (Throckmorton 1975). Therefore if the rate of synonymous substitution estimated for mammals is 0.55×10^{-8} nucleotide per site per year (Hayashida and Miyata 1983), is like that of Drosophila, the desired degree of divergence should be attainable in suitably chosen interspecific comparisons within this genus.

The specific choice of species was influenced by the desirability of choosing Drosophila taxa that are very distantly and moderately distantly related to D.melanogaster, as well as representing the major, successive drosophilid radiations. Drosophila melanogaster and D.virilis are in different subgenera, Sophophora and Drosophila respectively (Throckmorton 1975; Wheeler 1981); D.virilis is thought to represent the ancestral species (Clayton and Guest 1986); and the chromosomal elements of the two have been extensively compared by linkage analysis of putatively homologous loci (Sturtevant and Novitski 1941; Patterson and Stone 1952; Alexander 1976; Gubenko and Evgen'Ev 1984). The expectations for degrees of divergence were based on presumed separation of the species, as well as an empirical data from previous studies on these or related species (Zwiebel et al. 1982, Schaeffer and Aquadro 1987, Blackman and Meselson 1986).

In this chapter a nucleotide sequence of the amylase locus in D.virilis is presented. Features of amylase gene organization in D.virilis are also described, and a detailed comparison with the amylase nucleotide sequence from D.melanogaster (Boer and Hickey 1986) and D.pseudoobscura (Brown et al. 1989) is made.

MATERIALS AND METHODS

D.virilis library and clones

A Drosophila virilis genomic library in phage lambda EMBL3 was obtained from R. Blackman at Harvard University. The library was screened using both the 650bp PvuII fragment and the 1.4kb EcoR I fragments of the pOR-m7 amylase cDNA clone. Screening of the library paralleled what was done for the Makokou library (see Chapter 3), except that the prehybridization and hybridization were performed at a lower temperature (42° C). Washing was at 50° C. Seven positive clones were grown up and DNA was extracted from them (for detailed methods, see Chapter 3). Lambda clone Vir #1 which gave a restriction map similar to that of the total genomic DNA was chosen for further analysis. The restriction map was constructed using clone Vir#1 and the total genomic DNA restriction digests.

Subcloning and Sequencing

From the D.virilis lambda phage clone Vir #1, four sub-fragments, 3kb Eco RI, 8 kb Sal I, 2.2 kb Sal I, and 200 bp E/S were subcloned into the plasmid vectors pIBI 24 and pIBI 25, in both orientations, for sequencing. Subclones were identified either by colony hybridization (Grunstein and Hogness 1975), using the Pvu II fragment probe, or by restriction analysis of plasmid mini-preps (Holmes and Quigley 1981).

DNA sequencing of single-stranded templates was performed as described above. In some cases double-stranded templates were sequenced using the alkaline denaturation method of Chen and Seeburg (1985).

RESULTS

Physical Organisation of the D.virilis Amylase gene

Low resolution mapping of the D.virilis amylase gene

The plaque lifts from the D.virilis genomic library hybridized under conditions of relatively high stringency to probes from the Oregon-R strain of D.melanogaster. This provides the first indication of a high degree of similarity in the amylase coding sequences of these two Drosophila species (Figure 5.1).

High resolution Mapping (DNA sequencing)

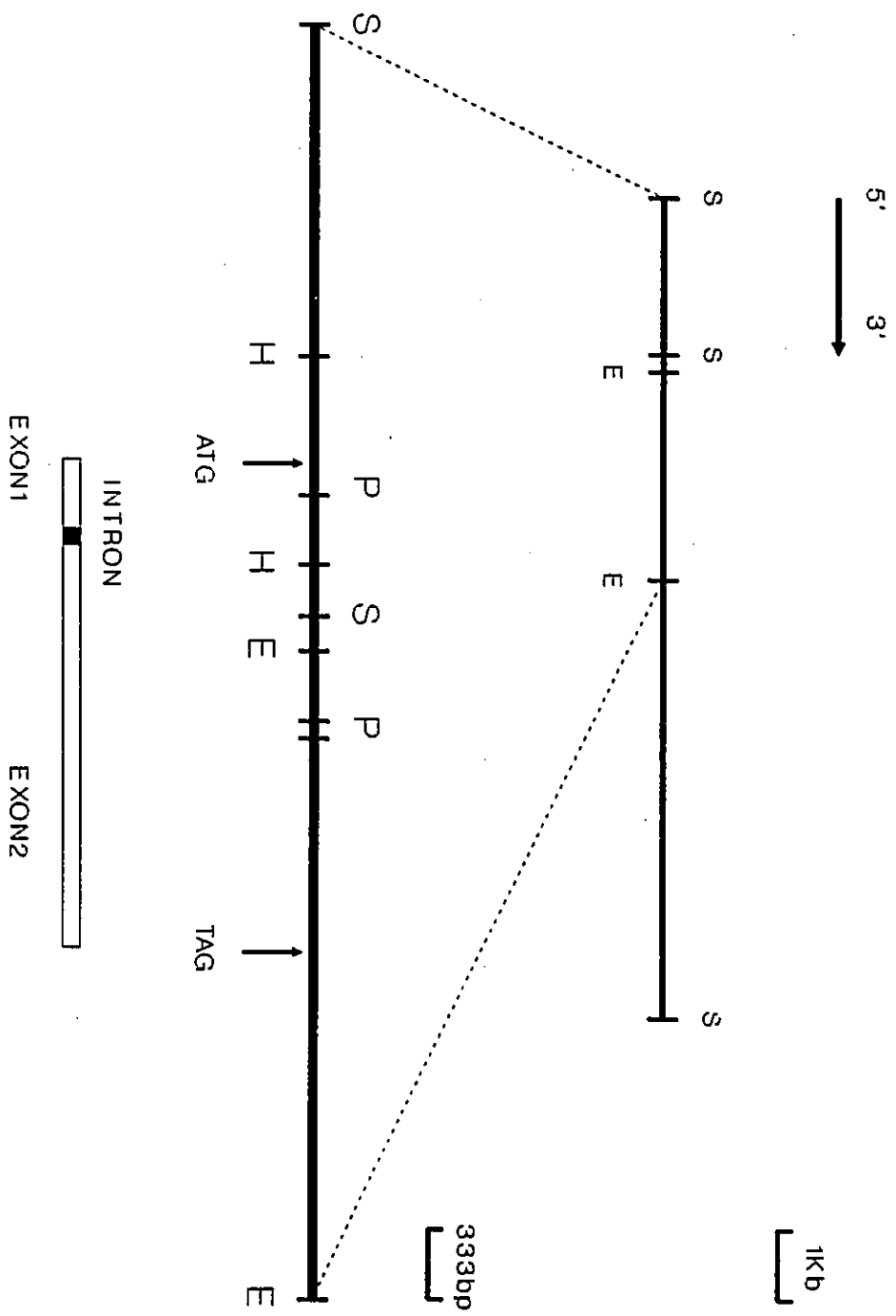
The sequencing strategy for the amylase gene is displayed in Figure 5.5. The nucleotide sequence of the gene is listed in Figure 5.6 along with the predicted amino acid sequence. The position +1 is assigned by homology to the transcribed region close to the D.melanogaster translational initiation site. For the D.virilis, 2550 bp of the amylase gene region were sequenced, including the complete transcribed region and 780 bp of the 5' and 250 bp of the 3' flanking regions. The single gene has a coding capacity of 495 amino acids

(comparable to the predicted size of the mRNA (Magoulas et al. in preparation)).

Figure 5.1. Second screening of positive plaques from a D.virilis genomic phage library which was plated at low density and probed with D.melanogaster Pvu II fragment of pOR-m7. Arrows indicate marks of radioactive ink used to align the filters. washing was at 50° C.



Figure 5.2 Restriction map constructed using clone Vir #1 and the total genomic DNA from adult flies. From the genomic blots and the clones of the D.virilis, it is obvious that there is only one amylase gene in this species. E, Ecor I; H, Hind III; P, Pst I; S, Sal I;



D. virilis

Figure 5.3. Positive clones from D.virilis genomic library in phage lambda digested with Eco RI and Sal I. OR-R refers to a genomic lambda clone from D.melanogaster which was used as an internal control. Clone # 1 is the clone used for further analysis.

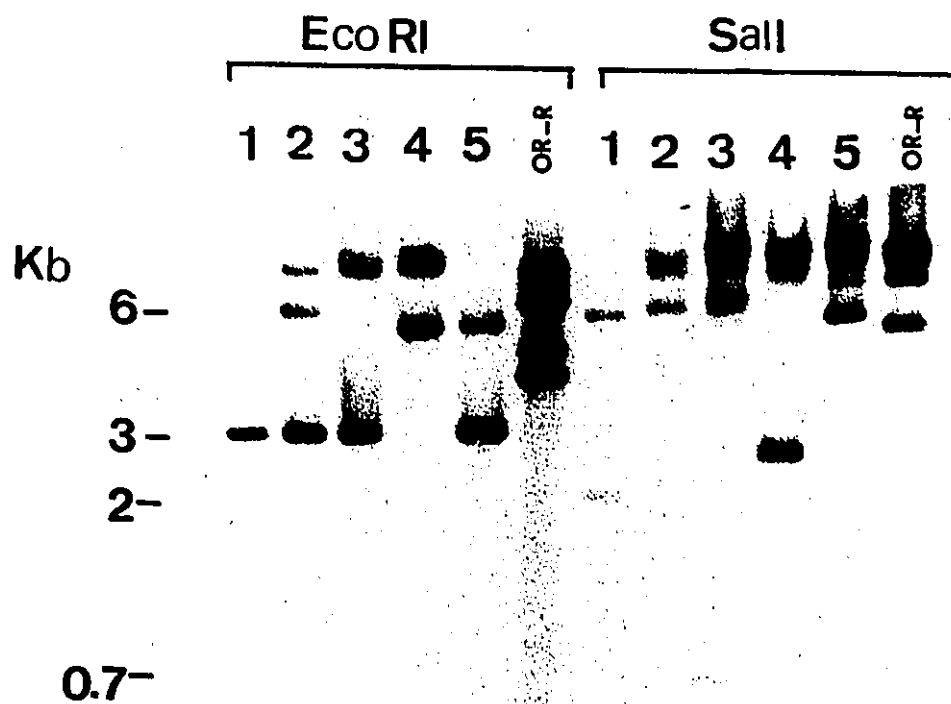


Figure 5.4. Restriction pattern obtained after the digestion of total fly DNA from D.virilis probed with the pOR-m7 Eco R1 fragment from the amylase cDNA clone of D.melanogaster.

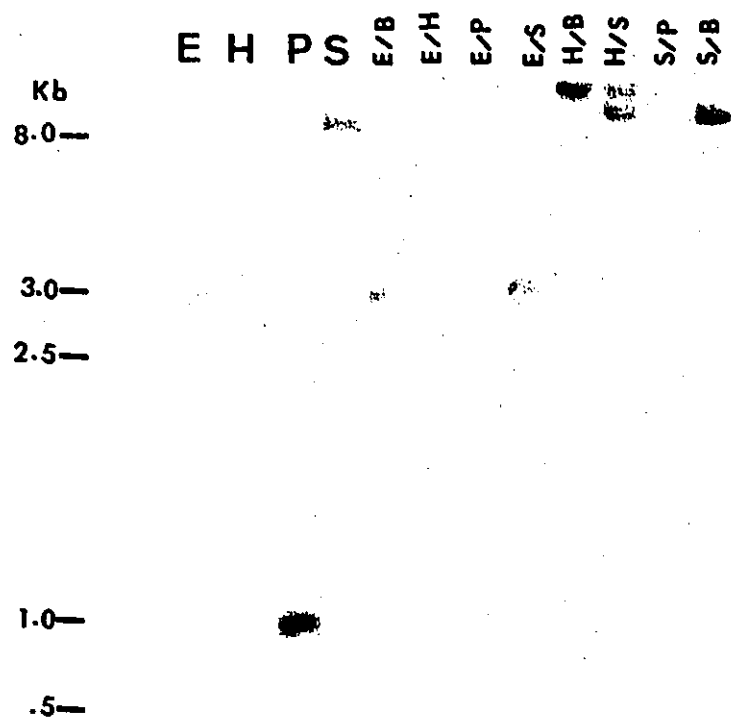
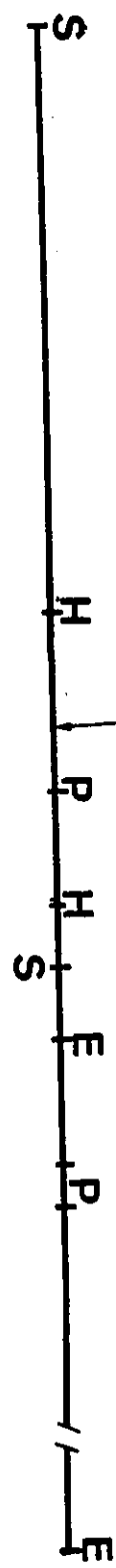


Figure 5.5. Strategy for sequencing the Drosophila virilis amylase gene. The restriction map is as in Figure 5.4. Solid boxes represent the exons of the amylase gene. Restriction sites used in the sequencing are indicated. The arrows above represent restriction fragments used for sequencing. The thick arrows below represent sequences determined by custom-synthesized primers. Direction of the arrows indicates the direction of sequencing. Enzymes as in Figure 5.2.

200 bp

5

3



ATG

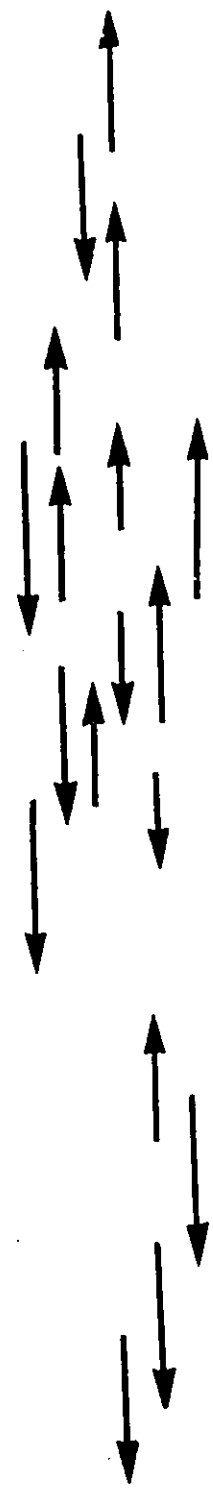


Figure 5.6. Sequence of chromosomal segment encompassing the amylase gene of D.virilis. the sense strand is shown, and the encoded polypeptides are conceptually translated. The arrow marks the putative mRNA 5' end established by homology to known D.melanogaster start sites (Wong et al. 1985; Levine and Spradling 1985). TATA and CAAT boxes as well as the polyadenylation (AATAAA) signals are highlighted. Two possible CAAT boxes are underlined. The Leader sequence and the single intron are underlined.

Sequence elements upstream of the amylase coding region which resemble cis-acting control elements of mammalian genes are highlighted. CRE is the cyclic AMP responsive element (Silver et al 1987; Roestler et al. 1988) and AP-2 is the binding site for eukaryotic transcription factor AP-2 (Imagawa et al. 1987)

TCTTCATGCTCAAGAATTAATGTTTTTTTTTATAAAAGTTGATAAAATAAATCCTTGTCTG
 CGGCATTACAAAGCTCACAAAATTACGGAATACAAAAAAGGCAACACTGTAAGAAAATAA
 da.dT
 AGATCTTCATGACCACTAAAATTAAAAAAAAAAAAAAAAAACTTAAAAAAAAATAAAGTTTTT
 TTTCTTACCACAAAGAGGCAAGGTAGAGAACTTTTCCCAAGAGGCAAGGTAGAGAACTTT

 TCTTACCACAAAGAGGCAAGGTAGAGCACTTTTCCCAAGAAGTGGCAATGCTTGATTGTT

 AAGCCTGAGGCTTAGGTACTTAAGTTCAGGAAGATATAAGTGATATACTATATTGTTTAT
 TCATTTTGTTTTAAAGATCCCCTATAGGCATTAAATTTATTATTATCTAATAATTTTGT
 TAAATAAAATGCTAGTATCATAAATTAAGCTTGAGCTATGCGATAGCTGGCAACACTGG
 AP-2
 CAACCCGCTCGTGTGCGATATCTAATCTAACACACTGCACGCCGCGCTAGCCCCACTGT
 CRE
 GCGGCGTCTCTGCTCCCCCTAGTTAAACAAAATTTGACATTGTGATAAGCGAAGAATTTG

CATCACAATCATTTAAACAATTCCACGCAAAACCTATTGACCAATTACGAGCTACCAAGT

 TCTTGTGATGAGTGTTTCCCCCACTTGCGAGGTGTGTTGCGTATAAAAGGCCCTGCGGCGT

 CTACATTCGATTCAAGTGTGGAACGGGTTTCCATTTGAGTACAAAGCAACAGCAGCAACAT

Me

GTTCTGATCAAGGGCCTGGCTTGTCTAGCCCTCCTGGCGTTGTCCCATGCCAGTTTGC
 tPheLeuIleLysGlyLeuAlaCysLeuAlaLeuLeuAlaLeuSerHisAlaGlnPheAl
 CACCAACTACCAGAGCGGACGCAACGGCATGGTCCATCTGTTTCGAGTGGAATGGGACGA
 aThrAsnTyrGlnSerGlyArgAsnGlyMetValHisLeuPheGluTrpLysTrpAspAs
 TATTGCGGCGGAGTGTGAGAACTTCCTGGGCCCCCTACGGCTATGCCGGTCTGCAGggtgc
 pIleAlaAlaGluCysGluAsnPheLeuGlyProTyrGlyTyrAlaGlyLeuGln
aggtgagccaaattatagccattgccagcacattagtttcaatgccaatcccccgctca
 gGTCTCGCCTGTCAATGAGAACGCTGTAATTGGCAATCGGCCATGGTGGGAGCGTTATCA
 ValSerProValAsnGluAsnAlaValIleGlyAsnArgProTrpTrpGluArgTyrGl
 GCCCATCTCCTACAAGCTGAACACTCGCTCCGGCAATGAGGAGCAGTTCGCCAATATGGT
 nProIleSerTyrLysLeuAsnThrArgSerGlyAsnGluGluGlnPheAlaAsnMetVa
 GAGGCGTTGCAATAATGTGGCGTACGCACCTATGTGGATGTGGTGTTCATCACATGGC
 lArgArgCysAsnAsnValGlyValArgThrTyrValAspValValPheAsnHisMetAl
 CGCCGATGGTGGCACCCACTGGCACTGGCGGCAGCAATGCGATCCCAGCTCCAAAAGCTT
 aAlaAspGlyGlyThrHisTrpHisTrpArgGlnGlnCysAspProSerSerLysSerPh

CCCGGCCGTGCCCTTCTCCTCGCTGGACTTTAATCCCACCTGTGCCATTACCAACTATGC
 eProAlaValProPheSerSerLeuAspPheAsnProThrCysAlaIleThrAsnTyrAl
 GGATCCCACCAATGTACGTAAGTGTGAGCTGGTTGGTCTGCGCGATCTCAACCAGGGCAA
 aAspProThrAsnValArgAsnCysGluLeuValGlyLeuArgAspLeuAsnGlnGlyAs
 TTCCTGGGTGCGTGACAAGATTGTCGACTTCTTGAACCATTGACGGATCTGGGCGTTGC
 nSerTrpValArgAspLysIleValAspPheLeuAsnHisLeuThrAspLeuGlyValAl
 CGGATTCCGTGTGGATGCCGCCAAGCATATGTGGCCCGGCGATTGAGGCCATCTATGG
 aGlyPheArgValAspAlaAlaLysHisMetTrpProGlyAspLeuGluAlaIleTyrGl
 ACGTCTGAACAATTTGAATACTAATCATGGCTTCAGCTCGGGCGCCAAGCCATATATCTT
 yArgLeuAsnAsnLeuAsnThrAsnHisGlyPheSerSerGlyAlaLysProTyrIlePh
 CCAGGAGGTCATCGATGTGGGCGGCGAGGCCATCTCCAAGTCCGAGTACACCGGCCTGGG
 eGlnGluValIleAspValGlyGlyGluAlaIleSerLysSerGluTyrThrGlyLeuGl
 CGCCATCACTGAATTCGTCACCTCCGATTCCATTGGCAAGGTGTTCCGCGGCAAGGATCA
 yAlaIleThrGluPheArgHisSerAspSerIleGlyLysValPheArgGlyLysAspGl
 GCTGCGCTATCTGACCAATTGGGGCACCTCCTGGGGATTGCGGCCTCCGATCGCTCCCT
 nLeuArgTyrLeuThrAsnTrpGlyThrSerTrpGlyPheAlaAlaSerAspArgSerLe
 TGTCTTTGTCGATAATCACGATAACCAGCGCGGTCATGGTGCCGGCGGTGCCCATGTGCT
 uValPheValAspAsnHisAspAsnGlnArgGlyHisGlyAlaGlyGlyAlaHisValLe
 CACCTACAAGGTGCCCAAGCAGTACAAGATGGCCTCCGCCTTCATGTTGGCCCATCCCTT
 uThrTyrLysValProLysGlnTyrLysMetAlaSerAlaPheMetLeuAlaHisProPh
 TGGCAGCGACCGCATCATGTCCTCCTTTGCCTTCGATGACACCGACCAGGGACCACCGAC
 eGlyThrHisArgIleMetSerSerPheAlaPheAspAspThrAspGlnGlyProProTh
 CACCGATGGCCACAACATTGCCTCGCCACAATTCAACAGCGACAACTCCTGCAGCGGCGG
 rThrAspGlyHisAsnIleAlaSerProGlnPheAsnSerAspAsnSerCysSerGlyGl
 CTGGGTGTGCGAGCATCGTTGGCGCCAGATCTACAATATGGTTGCCTTCAGGAATGCCGT
 yTrpValCysGluHisArgTrpArgGlnIleTyrAsnMetValAlaPheArgAsnAlaVa
 CGGCGATGCCCAGCTGCAGAACTGGTGGGATAATGGCAGCAATCAGATCTCGTTCAGCCG
 lGlyAspAlaGlnLeuGlnAsnTrpTrpAspAsnGlySerAsnGlnIleSerPheSerAr
 CGGCAGCCAGGGCTTTGTTGCCTTCAATAATGACAACTATGATCTGAATAGCTCGCTGCA
 gGlySerGlnGlyPheValAlaPheAsnAsnAspAsnTyrAspLeuAsnSerSerLeuGl
 GACTGGCCTGTCCGCCGGCACCTATTGCGATGTCATTTCCGGCACAAAGAGCGGTTCGGG
 nThrGlyLeuSerAlaGlyThrTyrCysAspValIleSerGlyThrLysSerGlySerGl
 CTGCACCGGCAAAACCATCACCGTTGGCTCTGATGGACGCGCCAATATTTATATTGGCAG
 yCysThrGlyLysThrIleThrValGlySerAspGlyArgAlaAsnIleTyrIleGlySe

CTCCGAGGAGGATGGCGTCGTTGCCATCCATGTCAATGCCAAATTGTAGGCCAATATCTC
 rSerGluGluAspGlyValValAlaIleHisValAsnAlaLysLeu

CTTTACCACTTTTCAACACAATGAGATTATATACGTATACATATATTTTATTTTCGGCGAA
 --
 --

 AATAAAAGCAACCGACTTTCAATTTTACCGCTCACAAATCTCTCAATTTTCATCGAGTGG

 AAAACCCTTGGCCAAGTGTCTCGCCAAAGTTTCTTATCACAAAAGCATAATACCGCATC
 TATGTGACTACAAGTGTGCCAAATGTGGATTCTAAATAAACAAACAAAAAAAAAACAAGT
 GCATTTAAACAACGCTCTCGATTGACAGATATCCTGTGCCTCGGTAAGTTTGGCCTTTAA
 ATACATTTTATAAAAATTACTGCATTTTATTTATTGTGCGTAAAAATGTTATTGAAGTCA
 TTAGATAAATGCCTTCTCAAAGATCCTCAATTTTGTGATAATATTATCTTGGAATTTTTTG
 AATATGTTATTATTATACTCTTCTTAAAGCCCCTTATTTCTCATACTATTATATCTATCG
 ACAGTCGAAGCTGAATGTTTCCGTTACAGTTCGTTTTCAACACTGCTACACGCAGTATTT
 CAACTTCTCAATATCAAATTAATATATTTAAATGTCCTTATGCGTTCACCTCAGCCCTACC

Interspecific Sequence comparisons

Amylase Nucleotide Substitution Analysis

Distribution of Nucleotide Differences within the Coding Region

The protein coding sequences of D.melanogaster (Boer and Hickey 1986), D.pseudoobscura (Brown et al. 1989) and D.virilis were examined for the differences in the number of nucleotide substitutions, according to the method of Lewontin (1988). The structure of the sequence in the coding region is homologous to that of the published D.melanogaster sequence (Boer and Hickey 1986).

Since divergence of the subgenera, amylase has accumulated 286 nucleotide differences in the 1482 bp of the homologous coding region (see Figure 5.7). That leaves the overall similarity between the D.melanogaster and the D.virilis coding regions as 80.6 % at the DNA level. Of the 286 nucleotide differences, 200 (70%) are at the third position within the codon; the expected frequency of third position changes, based on chance alone, is 33%, or 95 differences in this case. The excess of third position (i.e., silent) changes is indicative of natural selection acting to conserve the nucleotides at the non-silent positions.

Figure 5.7. Nucleotide sequence of the amylase coding region of D.virilis compared to that of D.pseudoobscura (Brown et_al. 1989) and D.melanogaster (Boer and Hickey 1986). Dashes indicate sequence matches. Differences between the two species and D.virilis are shown below the nucleotide sequence of D.virilis.

+1

ATGTTTCTGATCAAGGGCCTGGCTTGTCTAGCCCTCCTGGCGTTGTCCCATGCCCAGTTT
 -----C-----CA---AC---C-GG--CA-C---T-----A-CG--AGC-----C
 -----GC-----A--A-A-TG--C--C-----G--G--A-C-----A--C--

GCCACCAACTACCAGAGCGGACGCAACGGCATGGTCCATCTGTTTCGAGTGGAAATGGGAC
 AA-----GTTG-T--C---G-----C--C-----G-----
 -A-----GCATC---T--T-GT--A-----C--C-----G-----

GATATTGCGGCGGAGTGTGAGAACTTCCTGGGCCCCTACGGCTATGCCGGTCTGCAG
 --C--C--C--C-----C-----C-G-----C-----T---
 --C--C--T--C-----C--A-----T--A--A-T-----C-----G-T---

TCGCCTGTCAATGAGAACGCTGTAATTGGCAATCGGCCATGGTGGGAGCGTTATCAGCCC
 -----C-----C-----C--G-AGA-TGGG--T--G-----C--C-----
 --C-----G--C-----C--C-AG-A--GC--C--C-----A-----C-----

ATCTCCTACAAGCTGAACACTCGCTCCGGCAATGAGGAGCAGTTCGCCAATATGGTCAGG
 -----C--C--T-----C-----GC-----C--T
 -----G-G--C-----A--C--A-----GC-----C-A--

CGTTGCAATAATGTCGGCGTACGCACCTATGTGGATGTGGTGTTCATCACATGGCCGCC
 --C-----C--C-C--A--G-----C-----C--CA-C-----C-----A---
 --C-----CGCC-----A--G-----C-----C-----C-----C-----

GATGGTGGCACCCTGGCACTGGCGGCAGCAATGCGACTCCCAGCTCCAAAAGCTTGCCG
 -----C-----TAC-----A-C-----CC--C-GC-----G-----C--C
 --C--A-----TAC-----CC--C-GC-----AG--G-----AT--C

GCCGTGCCCTTCTCCTCGCTGGACTTTAATCCACCTGTGCCATTACCAACTATGCGGAT
 --A-----A-----C-----C--C-----C-----C-G-----CAAC--C
 -GA-----A-----C--C--G-----C-----C-G-----CAAC--C

CCCACCAATGTACGTAACCTGTGAGCTGGTTGGTCTGCGCGATCTCAACCAGGGCAATTCC
 G---A-C-G--G--C-----C-----C--C-----T--C-----C--G
 G---A-G-G--G--C-----C-----C-----C--T-----C---

TGGGTGCGTGACAAGATTGTGACTTCTTGAACCATTTGACGGATCTGGGCGTTGCCGGA
 -AC--C-AG-----G--TC---T--C-----C---TC-----T--G-----
 -AC---AG-----G-G-----G--C--G-----C---TT-----C-----G-----

TTCCGTGTGGATGCCGCCAAGCATATGTGGCCCGCGATTGAGGCCATCTATGGACGT
 -----C-----C--T-----C-----C-----GC-T-----C--C
 -----C-----C-----C-----C-----CC--CC-T-----C--C

CTGAACAATTTGAATACTAATCATGGCTTCAGCTCGGGCGCCAAGCCATATATCTTCCAG
 --C--G--CC----C--CG----C-----GCG--C--AT---G---C--C---G-----
 --C--G--CC-A--C--CG-C--C-----GC-----AT-----G-G--C---G-----

GAGGTCATCGATTTGGGCGGCGAGGCCATCTCCAAGTCCGAGTACACCGGCCTGGGCGCC
 -----CA---T-----AG---AA-----T-----
 -----CA-----AG-----A-----

ATCACTGAATTCGTCACCTCCGATTCCATTGGCAAGGTGTTCCGCGGCAAGGATCAGCTG
 -----C--G-----C-----C-----C-----CA-----
 -----C--G-----C-----C-----C-----C-----

CGCTATCTGACCAATTGGGGCACCTCCTGGGGATTGCGCGCCTCCGATCGCTCCCTTGTC
 AC---CA--T---C-----GG-----C--C--T-----G--C---
 -AG--C-----C-----G-----C--C--T-----C-----G--A

TTTGTGCGATAATCACGATAACCAGCGCGGTCATGGTGCCGGCGGTGCCCATGTGCTCACC
 --C-----C--C-----C-----A--C---A--T---TG-----
 --C-----C--C-----C--T-----A---A--A--A--C---G--C--T-G---

TACAAGGTGCCCCAAGCAGTACAAGATGGCCTCCGCCTTCATGTTGGCCCATCCCTTTGGC
 -----C-----C---
 -----C---G--C-----C---

ACGCACCGCATCATGTCCTCCTTTGCCTTCGATGACACCGACCAGGGACCACCGACCACC
 --T-C---TG-G-----CT-----ACC-----C-----
 --T-C---G-G-----CT-----ACG-----T-----C--G--C-----

GATGGCCACAACATTGCCTCGCCACAATTCAACAGCGACAACCTCCTGCAGCGGCGGCTGG
 --C--A--G-----T--CACC-----G-----G-----A---
 --C-----C-----CATC-----T-----

GTGTGCGAGCATCGTTGGCGCCAGATCTACAATATGGTTGCCTTCAGGAATGCCGTCGGC
 -----C--C-----C-----G-----C---C---
 -----T-----C--C-----C-----G-----C-A--CA---G---

GATGCCCAGCTGCAGAACTGGTGGGATAATGGCAGCAATCAGATCTCGTTCAGCCGCGGC
 -----G-T--C-----TCC--C--A-----C-----C-----T--A
 TCG-A-G--A-C-----C--C-----C-----C-----A--

AGCCAGGGCTTTGTTGCCTTCAATAATGACAACTATGATCTGAATAGCTCGCTGCAGACT
 -A-AG--C--C--G-----C--C-----C--C--C--C-----C-----C
 ---GC-----C--G-----C--C-----C--C-----C-----C-----G

GGCCTGTCCGCCGGCACCTATTGCGATGTCATTTCCGGCACAAAGAGCGGTTCCGGCTGC
 -----C-----C-----C-----C-----GAG-----C---TC----
 -----C-----C-----C-----C-----T-C-----TC----

ACCGGCAAAACCATCACCGTTGGCTCTGATGGACGCGCCAATATTTATATTGGCAGCTCC
 -----G--G-----C-----GC--C-CC---A-----
 --G-----G--G-----C--A--C--C-----G--TTCC--CA-C-----

GAGGAGGATGGCGTCGTTGCCATCCATGTCAATGCCAAATTGTAA
 -CT--C-----T---C-G-----T--C-----C-----TG-----G
 -----C--C--A--GC-G-----T--C-----C-----G-----G

D.virilis
D.pseudoobscura
D.melanogaster

5' Flanking regions

The putative start codon is flanked by sequences (at -18 to -1) which are similar to the *Drosophila* consensus sequence (C/AAAA/CAUG) for translation initiation site (Cavener 1987).

Sequence comparisons show that at the upstream non-coding region there is little overall similarity between the *D.virilis* and *D.melanogaster* sequences; however, the classical regulatory sequences are conserved. The TATA box which is at -78 (at -64 in *D.melanogaster*) and the CAAT box, at -130 (at -110 in *D.melanogaster*); there are two other putative CAAT boxes at -160 and -170, and the one at -170 seems more similar to that of *D.melanogaster* at -110. The TATA and CAAT boxes in *D.virilis* are conserved at appropriate places relative to one another, though their exact location relative to the translational start site is different from that of the *D.melanogaster*. This suggests that there is a longer untranslated mRNA leader sequence in *D.virilis*. In addition to the TATA and CAAT sequences, other regions in the 5' flanking may be important in the regulation of transcription. Some typical mammalian regulatory sequences can also be found in the 5' region of this gene (Figure 5.6).

A striking sequence containing long stretches of alternating A and T (-330 to -460 and -600 to -780) is present in the D.virilis 5' upstream region. This sequence is absent from the other amylase genes. A computer search of the DNA Genbank Data base using the D.virilis 5' non-transcribed region shows some similarity between this sequence (-94 to -462) and a 5' sequence of the Bovine parathyroid hormone gene (63%). The region also has a 36 bp direct repeat (Figure 5.6).

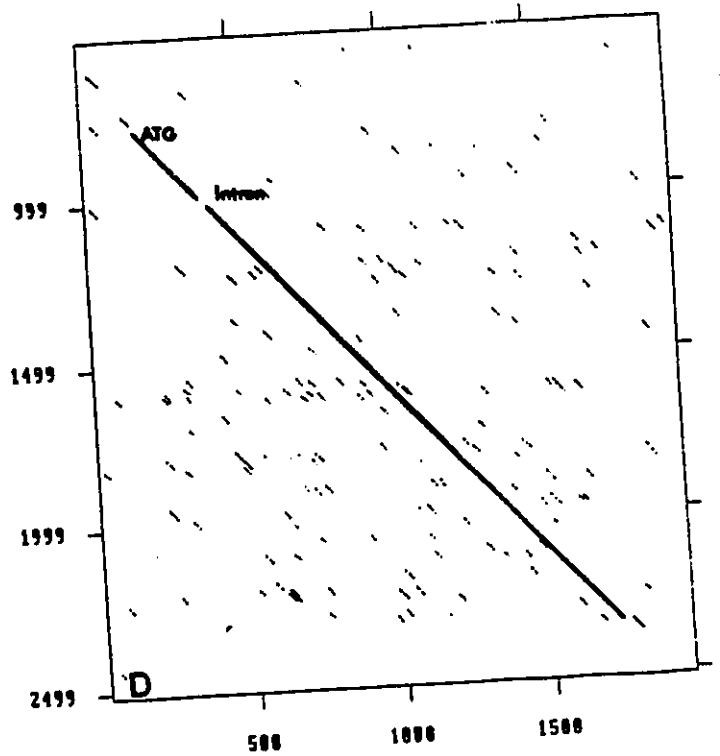
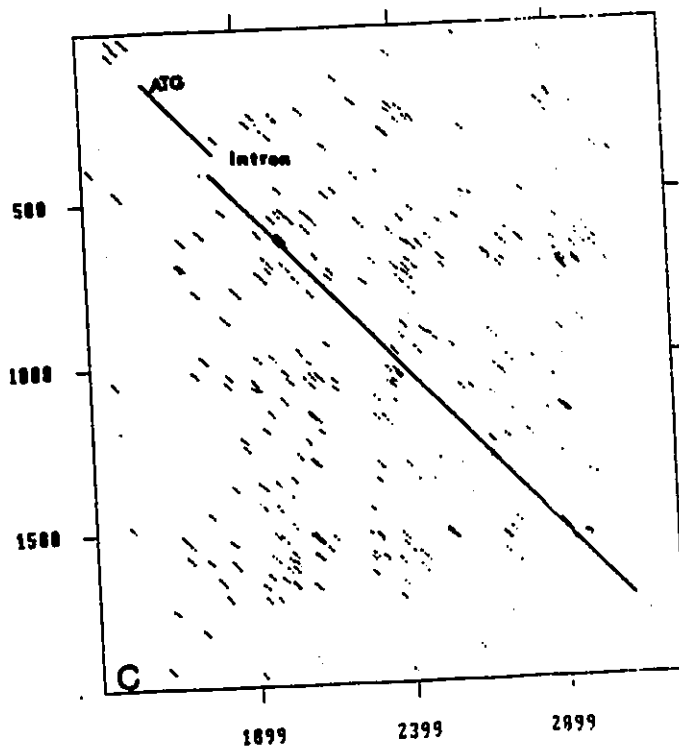
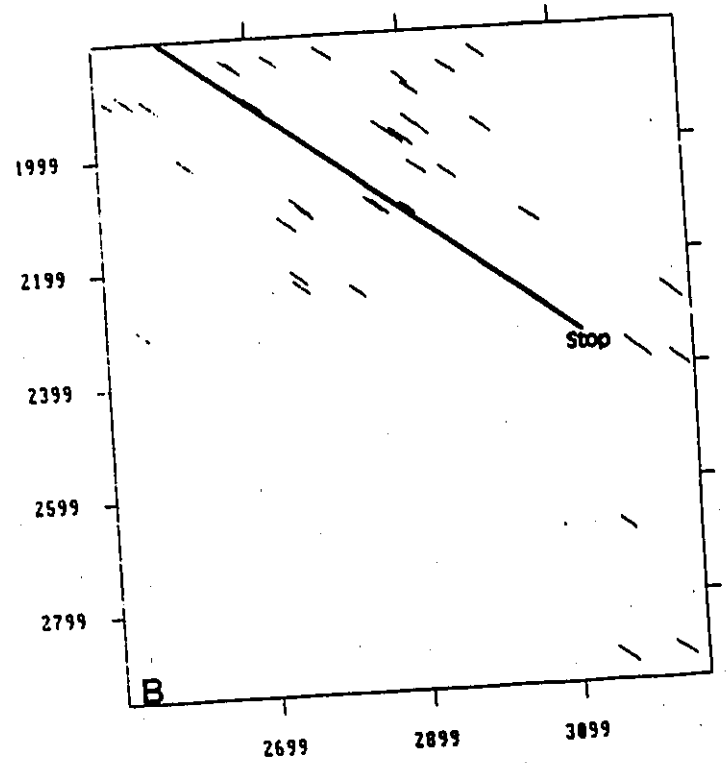
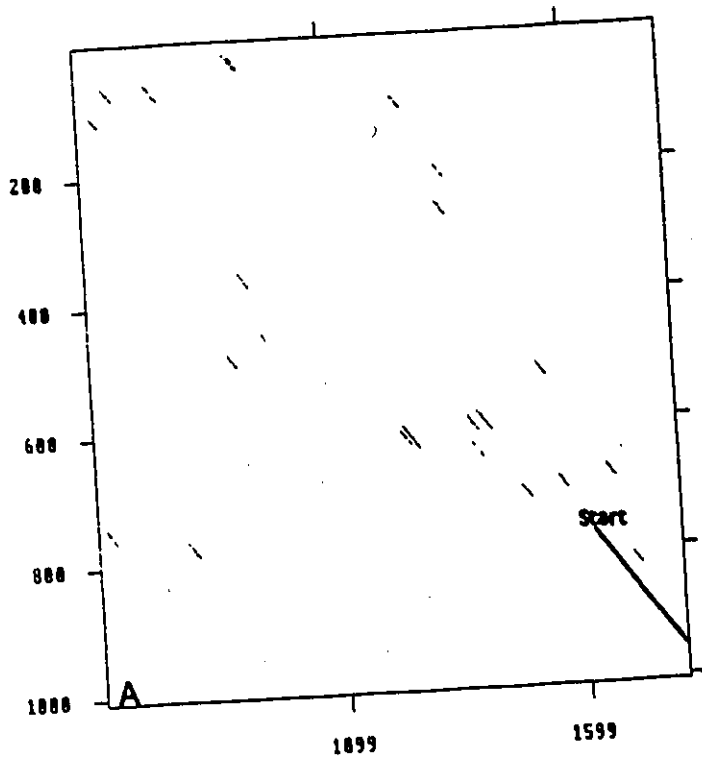
A comparison of the 5' region and the intron resulted in a similarity of a conserved sequence in both regions. A 9 bp stretch (TTGCAGGTG) from the 5' region at position -80 had a perfect match at the 5' intronic region that includes the splice site at position +183 (Figure 5.6). Another 5' sequence at -396 (TATAGGCATT) matches 9 out of 10 nucleotides of the intron.

The 3' non-translated region

In addition to the protein coding regions in the amylase gene that are conserved between D.virilis and D.melanogaster, the poly (A)-addition sequences AATATA are also conserved. The similarity between the two species in the 3' region starts to drop rapidly after the translational stop site (see Figure 5.8). Although the stop codon used by the GC-rich

Figure 5.8. Matrix analysis of interspecies sequence conservation at the amylase locus. Lateral shifts in the diagonal result from insertions/deletions, and off diagonal matches indicate sequence repeats. The direction of transcription is from upper left to lower right of each panel. The single intron of D.virilis is indicated by a slight offset of the two exonic diagonals. Numbers are as in Figure 5.6. , except for D.melanogaster where they refer to the sequence of Boer and Hickey (1986) and in D.pseudoobscura where they refer to the sequence of Brown et al. (1989). Note that discontinuous elements in the proximal 5' flanking region are substantially more conserved than the intron and most of the 3' untranslated sequence. Conservation downstream of the 3' end of the gene is less than upstream of the 5' end.

A, B, and C are comparisons of D.virilis and D.melanogaster, where D is between D.virilis and D.pseudoobscura.



melanogaster subgroup is the amber UAA, D.virilis opts for the ochre UAG stop codon.

Introns and Exons

Unlike the amylase genes of D.melanogaster, the single amylase coding sequence of D.virilis is interrupted by an intron. There are perfect splice junction signals at the 5' and 3' ends of the inserted sequences (Breathnach and Chambon 1981), hence the sequence is regarded as an intron. The intron in D.virilis and D.pseudoobscura (Brown *et al* 1989) provides an intragenic region presumably under minimal functional constraint, the divergence of which may be compared with other domains of the gene. There are 65 bps in the gene of the D.virilis which have no homology to the D.melanogaster amylase sequence (this can be seen as a shift in the two diagonals representing the exons in Figure 5.8). The intron of the D.virilis is 6 bp shorter than the D.pseudoobscura Amy¹ intron (Brown *et al*. 1989); however, the location of the two introns is conserved in the two species. Not only that, but this conservation of intron position seems to be the norm for all the introns in the amylase genes. Although the sequence of the introns had diverged between

Figure 5.9. Comparison between the introns of D.virilis and D.pseudoobscura (Brown et al. 1989). Different nucleotides are shown and similar nucleotides are represented by dashes.

GTTGCAGGTGAG CCAAATTATAGCCATTGCCAGCACATTAGTTTCAATGCCAATCCCCCG
 C--T-----ATACGTCCGG--AG---C--TC----G-CCACC-----T---C--T----

.CTCAGGT.....

C----CC-CCTTCCCAGGT

mouse (Schibler et al. 1982), flour beetle (Hickey et al. 1989), D.pseudoobscura (Brown et al. 1989) and the D.virilis amylase genes, the introns are located at the same position in relation to exons I and II from each. The D.virilis intron is 54% homologous to that of the D.pseudoobscura and is 40% homologous to the flour beetle's intron I. The intron in D.virilis is less GC rich (49%) than is the structural gene (57%).

Although the D.melanogaster and other members of the melanogaster species subgroup don't have any introns, D.melanogaster has the consensus splice site intact.

Relative Rates of Nucleotide Substitutions

The degree of nucleotide substitution that has occurred since the latest common ancestor between the two species representing the two subgenera, can be calculated by extrapolation from measurements on more closely related species (Britten 1986) and estimates of the time since divergence (Beverley and Wilson 1984). However such a calculation makes untested assumptions and is uncertain in the absence of a detailed fossil record for *Drosophila*. For the purposes of this study, it is important to know if homologous regions indicate functional conservation, i.e., whether these two species are sufficiently distant in time

since divergence that mutation would have obliterated homologies at positions that are selectively neutral. However it is possible to obtain a direct estimate of divergence at selectively neutral positions by examining third base codon differences.

Sequence divergence between the D. melanogaster and D. virilis can be examined with regard to placement of individual nucleotide changes, as shown in Figure 5.7. Nucleotide changes in the third position of codons are more frequent than changes in position 1. Changes in position 1 are two times as common as those in position 2 between the two sequences.

In order to evaluate the evolutionary distance separating each of the *Melanogaster* and *Sophophoran* subgenera, the procedure of Miyata and Yasunaga (1980) to estimate the number of silent nucleotide substitutions that have occurred since the amylase genes of the two subgenera had diverged from a common ancestor was used. In each species, 70% of the possible substitutions in the coding region are silent. The ratio K'_s of the observed number of silent substitutions obtained from pairwise comparisons of the two to the total possible number of all silent substitutions in the coding region is 0.476. This value, however, considers those nucleotide differences present today. It is corrected for possible multiple substitutions at the same site (Jukes and

Cantor 1969; Kimura and Ohta 1972), giving the silent substitution divergence value, K_S :

$$K_S = \{-(3/4)\ln[1-(3/4)(K'_S)]\} = 0.768$$

and bringing the total evolutionary separation of the subgenera to 76 million years ago.

Table 5.1. Frequency of nucleotides in the first, second and third position of codons used by amylase genes of D.melanogaster and D.virilis.

Position								

	First		Second		Third		Total	
	-----		-----		-----		-----	
	M	V	M	V	M	V	M	V
T	0.18	0.20	0.15	0.24	0.06	0.24	0.13	0.23
C	0.18	0.18	0.22	0.22	0.63	0.46	0.34	0.29
A	0.25	0.26	0.31	0.31	0.05	0.04	0.20	0.20
G	0.36	0.34	0.22	0.20	0.24	0.24	0.27	0.26

M= D.melanogaster

D= D.virilis

Transition and Transversion Synonymous Events

The frequency of transitions and transversions between D.melanogaster and D.virilis are shown below:

	A	40	G	
	17		48	
40				28
	C	115	T	

Nucleotide substitutions at the 288 polymorphic sites between the Drosophilan (D.virilis) and Sophophoran (D.melanogaster and D.pseudoobscura) subgenera are categorized in the box above into the six possible transitions and transversions. Transitions of C--T were three times more frequent than A-G transitions. Within the transversion class G-C transversions are over-represented while A-T and G-T transversions are less frequent in

comparison to A-C transversions between the two sequences (a test for equal frequencies of the four kinds of transversion gave $\chi^2 = 16.96$; d.f. = 3; $p < 0.01$). However, on an average, the overall observed ratio of transitions:transversions is 1.17 (155/133), which deviates significantly from the expected ratio of 2:1 (0.5) transversion to transitions ($\chi^2 = 18.0$; d.f. = 1; $p < 0.001$) assuming equal mutation rates, equal base utilization and no differential selection.

Amylase amino acid sequence comparison

Based on the unambiguous nucleotide sequence alignments in the coding regions, the amino acid sequence of the amylase locus of D.virilis can be predicted. When the predicted protein of the D.virilis is compared to that of D.melanogaster, the two are homologous at 86.3 % of the amino acid sequences. However, in the hydrophobic region that is presumed to be the signal peptide of the secreted protein 7 out of the 18 amino acids differ between the two proteins. However, the signal sequence of the alpha-amylase shares several of the characteristic features of signal peptides.

There are 65 differences observed in the 494 amino acids compared between D.virilis and D.melanogaster and 68 amino acid differences between D.virilis and D.pseudoobscura (Figure 5.10). Sixty four of the amino acid replacements

Figure 5.10. Decoded amino acid sequence of D.virilis aligned against sequence of D.pseudoobscura (Brown et al. 1989) and D.melanogaster (Boer and Hickey 1986).

1 M F L I K G L A C L A L L A L S H A Q F
 -----T-----T-----G-----I-----F-----I--A--S-----
 -----A-----S--I--V-----V--A--N-----

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

 41 D I A A E C E N F L G P Y G Y A G L Q V
 -----Q-----
 -----N-----V-----

 61 S P V N E N A V I G N R P W W E R Y Q P
 -----L--S--G-----
 -----K--D--S-----

 81 I S Y K L N T R S G N E E Q F A N M V R
 -----T-----S-----S
 -----E-----S-----K

 101 R C N N V G V R T Y V D V V F N H M A A
 -----A-----I-----
 -----A-----

 121 D G G T T G T G G S N A T P S S K S L P
 -----Y-----A-----T-----S-----F--
 -----Y-----T-----S-----Y--

 141 A V P F S S L D F N P T C A I T N Y A D
 -----Y-----S-----N--
 G-----Y-----S-----N--

 161 P T N V R N C E L V G L R D L N Q G N S
 A--N--E-----
 A--N--E-----

181 W V R D K I V D F L N H L T D L G V A G

Y---Q-----V-S-D-----I-----
Y---Q-----V---E-----D-----I-----

201 F R V D A A K H M W P G D L E A I Y G R

-----A-----G-V-----
-----A-----A-V-----

221 L N N L N T N H G F S S G A K P Y I F Q

---K-----D-----A-----S---A-----V---
---K-----D-----A-----S---A-----F---

241 E V I D L G G E A I S K S E Y T G L G A

-----M-----T-----
-----M-----

261 I T E F R H S D S I G K V F R G K D Q L

-----K---

281 R Y L T N W G T S W G F A A S D R S L V

T---M-S-----G-----
Q-----A-----

301 F V D N H D N Q R G H G A G G A H V L T

-----D-----
-----D-----

321 Y K V P K Q Y K M A S A F M L A H P F G

341 T H R I M S S F A F D D T D Q G P P T T

---P---V-----S---T-----(-)
 ---P---V-----S---T-----
 361 D G H N I A S P Q F N S D N S C S G G W
 ---Q-----T-----L-----G-----
 -----I-----
 381 V C E H R W R Q I Y N M V A F R N A V G
 -----A-----
 -----T-----
 401 D A Q L Q N W W D N G S N Q I S F S R G
 S---D-----S-----
 S--D--E--I-----
 421 S Q G F V A F N N D N Y D L N S S L Q T
 N--R--A-----
 --R-----
 441 G L S A G T Y C D V I S G T K S G S G C
 ---P-----E-----S-----
 ---P-----S-----S-----
 461 T G K T I T V G S D G R A N I Y I G S S
 ---V-----S-----S-----
 -----S-----N-----
 481 E E D G V V A I H V N A K L U D.virilis
 A--D-----L-----M---D.pseudoobscura
 ---D-----L-----D.melanogaster

occur at a conserved position in the three species. Thirty three of these amino acid differences are shared by D.melanogaster and D.pseudoobscura compared to D.virilis. Six amino acid replacements are unique to D.melanogaster (vs D.virilis) and do not occur in the D.pseudoobscura amino acid sequence. Fifteen of the amino acid differences between D.virilis and D.pseudoobscura are not shared by D.melanogaster. These amino acid differences are not distributed uniformly across the amino acid sequence, but are clustered into 'blocks'. Drosophila virilis protein has 6 fewer acidic residues and one less basic residue relative to amylase in D.melanogaster. The pI of D.virilis polypeptide, 6.03, is similar to that of D.pseudoobscura and close to that of D.melanogaster, 5.99.

Comparison of Codon Bias at the Amylase Locus

The D.virilis and D.melanogaster amylase codon biases were compared on a goodness-of-fit test, by using the null hypothesis that both species have the same codon bias at the amylase locus (see Lewontin 1988, Table 2). This test compares between two species, the number of occurrences of a specific codon with the sum of occurrences of all codons within an amino acid codon group. Table 5.2. shows a comparison of codon usage of three Drosophila species.

Differences in codon usage do not appear to be very significant between the three species. Overall codon usage conforms approximately to what has been seen for a large number of D.melanogaster genes (Maruyama et al. 1986).

Table 5.2. The distribution of codons in the amylase genes of *D.melanogaster* (A), *D.pseudoobscura* (B) and *D.virilis* (C).

A B C				A B C				A B C			
TTT Phe	1	0	8	TCT Ser	0	1	1	TAT Tyr	2	1	10
TTC Phe	22	25	17	TCC Ser	26	26	21	TAC Tyr	17	17	7
TTA Leu	0	0	0	TCA Ser	0	0	0	TAA End	1	1	0
TTG Leu	1	2	9	TCG Ser	4	3	6	TAG End	0	0	1
TGT Cys	1	0	4	CGT Arg	4	5	8	AGT Ser	1	2	0
TGA End	0	0	0	CGC Arg	16	14	12	AGC Ser	1	2	13
TGC Cys	8	9	5	CGA Arg	2	0	0	AGA Arg	0	0	0
TGG Trp	11	11	12	CGG Arg	1	0	1	AGG Arg	0	3	2
CTT Leu	2	1	1	CCT Pro	1	0	1	CAT His	2	1	10
CTC Leu	5	8	3	CCC Pro	13	15	9	CAC His	10	10	5
CTA Leu	1	0	1	CCA Pro	0	1	4	CAA Gln	1	0	1
CTG Leu	19	17	19	CCG Pro	2	1	2	CAG Gln	15	17	17
ATT Ile	3	3	9	ACT Thr	2	2	7	AAT Asn	3	0	25
ATC Ile	16	17	12	ACC Thr	19	23	17	AAC Asn	32	35	16
ATA Ile	1	0	0	ACA Thr	0	2	1	AAA Lys	0	1	4
ATG Met	10	12	9	ACG Thr	3	0	2	AAG Lys	19	17	12
GTT Val	2	3	6	GCT Ala	3	4	2	GAT Asp	2	11	22
GTC Val	17	16	12	GCC Ala	33	34	33	GAC Asp	32	21	8
GTA Val	1	0	3	GCA Ala	2	3	0	GAA Glu	3	0	1
GTG Val	19	14	11	GCG Ala	3	2	5	GAG Glu	15	14	15
GGT Gly	4	8	7	GGC Gly	31	33	40	GGA Gly	15	9	6
GGG Gly	0	2	0								

DISCUSSION

The D.virilis single amylase gene

I have determined in this section that the amylase region in D.virilis contains one coding gene, an intron and putative regulatory sequences. I am confident that regions isolated and the cloned fragments from Vir #1 contain the amylase gene. The restriction map of the cloned fragment is identical to a restriction map prepared following analysis of genomic DNA. In these genomic digests there is no evidence of additional fragments which might indicate other active genes or pseudogenes. The aligned sequences reveal a similar organization of the amylase locus in D.melanogaster and D.virilis. However an additional intron was identified in D.virilis.

Sequence comparisons between Drosophila species are increasingly being used to identify sequences important in gene regulation. The high level of sequence divergence in non-coding regions between D.melanogaster and D.virilis allows clear identification of conserved sequences. Comparisons of this sort have resulted in the localization of previously uncharacterized genes (Cohn 1985; Schaeffer and

Aquadro 1987) and in the description of sequences that may be involved in gene expression (Blackman and Meselson 1986).

It has been suggested that in coding regions, if mutations occur at random, then one should expect twice as many transversion as transition events (e.g. see Nei 1987). Some observations of synonymous substitution levels in Drosophila are consistent with this predicted transition to transversion ratio of 0.5 (Kreitman 1983); others are not (Bodmer and Ashburner 1984; Schaeffer and Aquadro 1987).

The amylase protein has accumulated 65 amino acid substitutions since the divergence of the two subgenera Drosophila and Sophophora. There is conservation of overall charge of the two proteins and of the pattern of distribution of hydrophobic and hydrophilic domains (see Appendix 3). The third position codons in D.virilis are less GC rich (70%) than in the melanogaster species group (90%). This can be explained by the presence of the amylase duplication in the melanogaster species subgroup which undergo gene sequence conversion during evolution (Hickey et al. 1989). Duplicated genes which inter-convert become GC rich by time (Brown and Jiricny 1988); this biases the codon usage in the melanogaster species subgroup.

Using the fact that D.virilis has a single amylase gene compared to all other Drosophila species studied to date, the evolutionary history of the amylase gene duplication in

Drosophila can be traced. It is apparent that the ancestral *Drosophila* amylase gene was a single copy gene that has been duplicated after the split of the two subgenera *Sophophora* (including *D.pseudoobscura* and members of the *melanogaster* species subgroup) and *Drosophila* (*virilis* Species subgroup).

5' flanking region of *D.virilis*

Both *D.melanogaster* and *D.virilis* are subject to glucose repression (Hickey and Benkel, 1989; Magoulas and Abukashawa, Unpubl. observ.), however, the upstream sequence shows a high level of divergence. Although specific regulatory patterns are conserved over long periods of evolutionary time, it seems that overall the DNA sequences upstream of the affected genes are subject to little evolutionary constraint.

There is no AT rich region in the 5' amylase gene in the *melanogaster* subgroup comparable to that of the *D.virilis* one. However, an AT rich region about 500 bp upstream of the RNA initiation site in the silk fibroin gene (Tsuda and Suzuki 1981) and the AT rich spacers between histone genes (Grosschedl and Birnstiel 1980) are required for optimal rates of transcription. Because of the alternating A's and T's, potential single-stranded loops could form in this region and, since the sequences are repeated, several alternative loops could be formed which might be related to

transcriptional activity of the amylase gene. Alternatively, these sequences may be residual sequences from the initial formation of the amylase gene. They might have arisen by slippage reactions (Streisinger et al. 1966) in regions of the genome that are not strongly selected. They probably form hot spots for recombination; first because they can easily become single-stranded (Hentschel 1982), and second because base pairing in this region of simple sequence repeats can occur easily. It is likely that recombination in this region would also lead to elongation or deletion, because the probability for an unequal crossover event within the AT stretches is higher (Tauntz and Renz 1984).

EVOLUTION OF THE AMYLASE GENES IN DROSOPHILA

Evolutionary Stability of the Organization of the Amylase locus

The genus Drosophila has been, in many ways, an ideal group for evolutionary studies. One major drawback to its use, however, is that the timing of divergence events within the genus is largely unknown. This deficiency is unfortunate, as knowledge of the timing of evolutionary events in other groups, especially vertebrates, has allowed the formation of

hypotheses concerning evolutionary rates, the importance of gene regulation and arrangement in evolution, and the association of divergent events with geographic settings. Beverley and Wilson (1984), using a molecular clock of immunological distance, indicate that the subgenus Sophophora diverged from the subgenus Drosophila around 62 Myr ago. The analysis of Throckmorton (1975) suggests that the lines leading to the melanogaster and virilis groups separated in the Eocene, corresponding to 80 million years. Dividing this time into the corresponding silent site substitution frequencies (K_s) for the species pair, a value of about 1.0×10^{-8} /nucleotidesite/year is obtained. This is not very much different from the value of 0.55×10^{-8} estimated for the mammalian silent substitution rate by Hayashida and Miyata (1983). The value of 76 MY estimated from the amylase data is close to the 80 MY value calculated by Blackman and Meselson (1986) from the incidence of silent nucleotide change in the protein coding regions of heat shock gene hsp82 of D.melanogaster and D.virilis. Thus, these limited comparisons suggest that the rate of silent nucleotide substitution may be fairly uniform along different Drosophila lineages and also in different Drosophila genes.

With a Drosophila gene that codes for a protein as an example, one can show that the limited fossil and biogeographic data available for flies permit a time scale

to be established for Drosophila evolution. The amylase gene considered here is ideal for such a study.

The results presented reveal strong conservation of the overall organization of the amylase locus. Although duplicated in members of the subgenus Sophophora, members of the subgenus Drosophila possess only one amylase gene. Thus the duplication of the genome happened after the divergence of the two subgenera. The common ancestry of the single D.virilis and the duplicated Sophophoran genes antedates the separation of the Drosophila species considered here. So the sequence divergence between the two subgenera could be interpreted as a result of speciation (species divergence) rather than occurrence of substitutions after gene duplications. This seems to be true since the duplication in the melanogaster species group is probably not the same duplication as that of the obscura species group (Boer and Hickey 1986; Doane et al. 1987; Payant et al. 1988; Brown et al. 1989). The duplicated genes of members of the melanogaster species subgroup are divergently transcribed, with their 5' ends closest to each other, while in the obscura subgroup the two genes are arranged in tandem (Brown et al. 1989), a situation similar to that of the amylase genes of the flour beetle Tribolium castaneum (our unpublished observations).

The phylogenetic comparison of amylase coding sequences also gives some insight into the evolution of the exon/intron structure of genes. For example, amylase genes from mammals are split by several introns, whereas representatives of the melanogaster subgroup completely lack introns (Boer and Hickey, 1986). More recent data show that other insect amylases have introns within their coding sequences. For instance, in the cricket, Achaeta domesticus, the number of introns is similar to that in mammals, but the size of the introns is much smaller (Hickey and Benkel 1989). The flour beetle, Tribolium castaneum, has split amylase genes, but the introns of this insect are fewer and also shorter than either cricket or mammalian introns (Abukashawa et al., unpubl. obs). The absence of introns from the Melanogaster subgroup points towards a loss of intron(s) from this lineage since the ancestral (D.virilis) lineage retains an intron similar in sequence and position to that found in a descendant lineage (D.pseudoobscura). The conservation of intron positions between insects and mammals is in line with conservation of intron positions from other genes (Gilbert, 1986) and indicates that introns are an evolutionarily ancient feature of genes. In contrast, the discrepancy in the number and size of introns that split amylase genes from different organisms supports the alternative view that introns can be inserted into genes during the course of

evolution (Hickey et al. 1986; Lambowitz 1989).

In agreement with expectations, the introns, the 3' untranslated sequences and much of the 5' upstream DNA are poorly conserved between the two species, much more poorly than the protein coding regions of the amylase genes. These rapidly diverging regions might approximate neutral evolution, and thus be useful for establishing phylogenetic relationships between more closely related *Drosophila* species.

The entire 5' untranslated region of *D.virilis* and the same region of *D.melanogaster* are highly diverged. This suggests that much of the sequence in this region is not essential for the developmental specificity of the amylase gene, although it may affect enhancement of the transcriptional initiation because of the conservation of certain short upstream elements.

In contrast to the lack of high overall sequence conservation in the upstream region, there is a high degree of conservation of short elements in the 5' flanking regions. These results vindicate the strategy of comparing amylase sequences in multiple species; furthermore the species chosen are apparently far enough separated in evolution to permit extensive drift of putatively unimportant sequences, yet not so far that elements subject to partial constraint will have become unrecognizable.

It is against the background of extensive diversification that the short, perfectly preserved elements in the 5' non-transcribed regions of the two species acquire special significance. Blocks of conserved sequences in the 5' flanking regions or intragenic regulatory sequences of homologous genes have been observed in evolutionary distant species (e.g. Efstradiadis et al. 1980, Emorine et al. 1983, Kelly and Pitha 1985, Blackman and Meselson 1986, Bray and Hirsh 1986, Kassis et al. 1986). It is still necessary to obtain experimental confirmation of the functional importance of these conserved elements; the sequence comparisons, however, point to a limited number of potential candidate sequences for such experimental tests.

CHAPTER 6

SUMMARY AND CONCLUSION

The principal goal of this work was to examine, at the molecular level, patterns of genetic variation within a species, and to correlate these intraspecific patterns with the more long-term, interspecific molecular evolutionary trends. Amylase-coding sequences were chosen as a model system to investigate the patterns of molecular evolution in an enzyme-coding gene.

The structure of the amylase locus in 50 different strains of Drosophila melanogaster, representing a wide variety of amylase genotypes and geographic distribution, was compared by constructing restriction maps of genomic DNA. Adult enzyme activity of alpha-amylase was also determined for each strain. The DNA-level variation at the amylase locus was compared with the electrophoretically-detectable protein variation. These intra-specific studies demonstrated extensive DNA and protein polymorphisms. Moreover, statistically significant linkage disequilibrium was observed between certain restriction sites and electrophoretic enzyme patterns; this indicates that there is a non-random association between the protein variants and DNA structure.

These results mean that there are a restricted number of amylase "haplotypes" segregating in natural populations.

A West African wildtype strain, Makokou^{4,6}, was chosen for further analysis of the intraspecific variation. This strain, is distinguished from the common laboratory *Drosophila* stocks in that it produces distinct subsets of alpha-amylase isozymes, Amy-4 and Amy-6, and in that it has a high level of amylase activity and low repressibility by glucose. A genomic library produced from this strain was screened with amylase probes, and regions from the duplicated coding sequences and their 5' flanking DNA were sequenced in order to identify conserved features of possible regulatory significance and to determine the nature of evolutionary change in different parts of the gene. The results show that the duplicated genes are highly conserved at their coding region and less conserved at the 5' non-transcribed region.

The results of the DNA sequencing analysis of the Makokou strain corroborate the results based on restriction site and allozyme polymorphisms. For instance, both the conservation and the high GC-content of the coding sequences explains the presence of invariant, GC-rich sites, such as Sal I (GTCGAC) within those regions; whereas the variability and the AT-richness of the non-coding regions correlates nicely with the variable Eco RI sites (GAATTC) in the flanking sequences.

In higher eukaryotes, a single enzymatic function is frequently encoded at more than one locus. The duplicated loci need not be closely linked but are likely to derive from a single ancestral locus. Here, I presented the study of evolution in Drosophila of a gene system showing a type of duplication for which detailed molecular analysis is possible. In contrast to the situation in mammals, where amylase genes occur in more complex gene families, Drosophila amylase genes occur in single copies or in very small gene families. Previous studies on the genetic relationship of the two genes using electrophoretic variants showed them to be closely linked. I have shown that all strains of D.melanogaster, regardless of their single-banded or double-banded allozyme phenotype, contain the duplicate amylase gene structure. This statement can be extended to all member species of the melanogaster species sub-group, which have also been shown to contain a duplicate gene encoding the enzyme alpha-amylase (Amy). Several features have emerged which suggest that classical views on mutation and evolution in such a system are inadequate. For instance, intraspecific and interspecific comparisons of the duplicated coding and flanking sequences show that the duplicated amylase gene copies undergo rapid gene conversion and concerted evolution. The gene conversion process can explain, in turn, why many wildtype strains which contain two gene copies produce only

a single isozymic form of the amylase enzyme.

In order to look at the long-term evolution of amylases within the genus Drosophila, a D.virilis amylase gene was cloned and sequenced. Comparisons of the amylase transcription unit from the D.melanogaster and the distantly related D.virilis show no conservation in the non-translated regions except for certain regulatory motifs. However, the coding regions of these species are 79% homologous at the DNA level and 86% identical at the amino acid level. In addition, there is variation in the intron/exon structure of amylase genes, both within the genus Drosophila and between insects and mammals. In contrast to D.melanogaster, the virilis has a small intron. The intron is located at a very similar, if not identical, position within comparable regions of amylase genes of other species including mammals. Finally, the duplicated amylase gene structure, which is characteristic of species in the melanogaster species sub-group, is not found in D.virilis. These contrasting findings of high conservation in coding sequences, coupled with lack of conservation of gene number or exon-intron gene structure, argue against many accepted notions that gene structure and organization are evolutionarily more conserved than the primary sequence.

The two most important findings from the sequence comparisons are, (i) that the amylase locus diverges at very

inhomogeneous rates in different regions and, (ii) that small highly-conserved DNA elements can be identified in all groups. Such non-homogeneous distribution of sequence divergence has already been described in other systems. For example, the intron and the far upstream region (-150 to -600) of the hsp82 gene show little or no conservation in four Drosophila species (Blackman and Meselson 1986), while greater similarity is found in the proximal 5' flanking region (-1 to -150) or the first exon, and even stronger throughout the second exon.

Previous, very long-term, phylogenetic comparisons of amylase-coding sequences indicated high levels of sequence conservation over long evolutionary periods (see Hickey et al., 1987). Such studies showed that alpha-amylase coding sequences were recognizably similar between animals, plants and microbes. This is unexpected in light of the large amounts of genetic variability which can be observed between strains within a single species (e.g., Singh et al., 1982). Such high levels of variability suggest the potential for very rapid rates of evolutionary divergence. The results presented in this thesis help to solve this paradox of rapid microevolutionary change, coupled with very slow macroevolutionary change, at the molecular level. It appears that polymorphisms at silent sites, within introns, in flanking non-coding sequences, and presumably within that

minority of substitution sites that are selectively neutral, can accumulate very rapidly. The fact that we observe such variants grouped into a few haplotypes in finite natural populations is further evidence of a random sampling process. Despite this rapid accumulation of variants at some sites, there is strong selection maintaining the integrity of the gene at other sites. Essentially, the conclusion is that patterns of molecular evolution are dominated by relatively high mutation rates, coupled with a combination of purifying selection and random genetic drift; the adaptively-significant mutations, which are of prime biological importance, are, nevertheless, statistically insignificant in general comparisons of sequence divergence.

REFERENCES

- Abe, K. (1958). Genetical and Biochemical studies on amylase in Drosophila melanogaster. Jpn. J. Genet. 33: 138-145.
- Abraham, I. and W. W. Doane (1978). Genetic regulation of tissue-specific expression of amylase structural genes in Drosophila melanogaster Proc. Natl. Acad. Sci. U.S.A. 75: 4446-4450.
- Aquade, M. (1988). Nucleotide sequence comparison of the rp49 gene region between Drosophila subobscura and drosophila melanogaster. Mol. Biol. Evol. 5: 433-441.
- Alexander, M. L. (1976). The genetics of Drosophila virilis, pp. 1365-1419. In: The Genetics and Biology of Drosophila Vol. 1C, M. Ashburner and E. Novitsky eds. Academic Press, New York.
- Arnheim, N. (1983). Concerted evolution of multigene families. In: Evolution of Genes and proteins. M. Nei and R. Koehn ed. Sinauer, Sunderland, MA. pp. 38-61.
- Ashburner, M., Bodmer, M. and F. Lemeunier (1984). On the evolutionary relationships of Drosophila melanogaster. Dev. Genet. 4: 295-312.
- Awise, J. C., Lansman, R. A. and R. O. Shade (1979). The use of restriction endonucleases to measure

mitochondrial DNA sequence relatedness in natural populations. I. Population structure and evolution in the genus Peromyscus. *Genetics* 92: 295-297.

Bahn, E. (1967). Crossing over in the chromosomal region determining the amylase in Drosophila melanogaster by means of translocations. *Hereditas* 67: 75-78.

Bally-Cuif, L., Payant, V., Abukashawa, S., Benkel, B. F. and D. A. Hickey (1989). Molecular cloning and partial sequence characterization of the duplicated amylase genes from Drosophila erecta. *Genet. Sel. Evol.* In Press.

Baltimore, D. (1981). Gene conversion: some implications for immunoglobulin genes. *Cell* 24: 592-594.

Barrington, E. J. W. (1957) in *THE PHYSIOLOGY OF FISHES*, Brown, M. E. Ed., Vol.1

Benkel, B.F. and D.A. Hickey (1986a). Glucose repression of amylase gene expression in Drosophila melanogaster. *Genetics* 114: 137-144.

Benkel, B. F. and D. A. Hickey (1986b). The interaction of genetic and environmental factors in the control of amylase gene expression in Drosophila melanogaster. *Genetics* 114:943-954.

Benkel, B. F. and D. A. Hickey (1987). A *Drosophila* gene is subject to glucose repression. *Proc. Natl. Acad. Sci. USA* 84:1337-1339.

- Benkel, B. F., Abukashawa, S., Boer, P. H. and D. A. Hickey (1987). Molecular cloning of DNA complementary to Drosophila melanogaster alpha-amylase mRNA. Genome 29:510-515.
- Benton, W. D., and R. W. Davis (1977). Screening lambda gt recombinant clones by hybridization to single plaques in situ. Science 196: 180-182.
- Berk, J. E., Kawaguchi, M., Zeineh, R., Ujihira, I., and R. Searcy (1963). Electrophoretic behaviour of rabbit serum amylase. Nature 200:572-573.
- Beverley, S. M. and A. C. Wilson. (1984). Molecular evolution in Drosophila and the higher Diptera. II. A time scale for fly evolution. J. Mol. Evol. 21:1-13.
- Bingham, P. M., Levis, R. and G. M. Rubin (1981). Cloning of DNA sequences from the white locus of Drosophila melanogaster by a novel and general method. Cell 25: 693.
- Birley, A. J. (1984). Restriction endonuclease map variation and gene activity in the Adh region in a population of Drosophila melanogaster. Heredity 52: 103.
- Blackman, R. K. and M. Meselson. (1986). Interspecific nucleotide sequence comparisons used to identify regulatory and structural features of the Drosophila hsp82 gene J. Mol. Biol. 188:499-515.

- Bloom, G. D., Carlson, B., and A. Danielson (1975) Br. J. Pharm. 54:523-528.
- Bodmer, M. and M. Ashburner (1984). Conservation and change in the DNA sequences coding for alcohol dehydrogenase in sibling species of Drosophila. Nature 309: 425-430.
- Boer, P. H. and D. A. Hickey (1986). The alpha-amylase gene in Drosophila melanogaster: nucleotide sequence, gene structure and expression motifs. Nucl. Acids Res. 14:8399-8411.
- Bray, S. J. and J. Hirsh (1986). The Drosophila virilis Dopa decarboxylase gene is developmentally regulated when integrated into Drosophila melanogaster. EMBO J. 5: 2305-2311.
- Breathnach, R. and P. Chambon (1981). Organization and expression of eukaryotic split genes coding for proteins. Ann. Rev. Biochem. 50: 349-383.
- Brenner, D. A., Smagocki, A. C. and R. D. Camerini-Oter (1985). Effect of insertions, deletions and double-strand breaks on homologous recombination in mouse L cells. Mol. Cell. Biol 5: 684-691.
- Britten, R. C. (1986). Rates of DNA sequence evolution differ between taxonomic groups. Science 231: 1393-1398.
- Brown, C. J., Aquadro, C. F. and W. W. Anderson (1989). The evolution of the amylase multigene family of Drosophila

pseudoobscura. Genetics In Press.

- Brown, T. C. and J. Jiricny (1988). Different base/base mispairs are corrected with different efficiencies and specificities in mouse kidney cells. Cell 54: 705-711.
- Brown, W. M. (1980). Polymorphism in mitochondrial DNA of humans as revealed by restriction endonuclease analysis. Proc. Natl. Acad. Sci. U.S.A. 77: 3605-3609.
- Brown, S. D. and M. Piechaczyk (1983). Insertion sequences and tandem repetitions as sources of variation in a dispersed repeat family. J. Mol. Biol. 165: 249-256.
- Buchberg, A. M. (1983). The developmental regulation of amylase expression by the map locus in the midgut of Drosophila melanogaster. Ph.D. Thesis, State University of New York, Buffalo, New York.
- Cavener, D. R. (1987). Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. Nucl. Acids Res. 15:1353-1361.
- Chen, E., and P. H. Seeburg (1985). Supercoiled sequencing. DNA 4:165-170.
- Church, G. M. and W. Gilbert (1984). Genomic sequencing. Proc. Natl. Acad. Sci. U.S.A. 81: 1991-1995.
- Clayton, F.E. and W.C. Guest (1986) Overview of chromosome evolution in the Family Drosophilidae. in: The Genetics and Biology of Drosophila, vol. 3, pp. 1-38. (edits M. Ashburner, H. Carson and J. Thompson), Acad.

Press, New York.

Cohen, E., Rae, P. and L.-J. Tsai (1980). Relatedness of rDNA intervening sequences among species of *Drosophila*. *Chromosoma* 55: 359-367.

Cohen, J. B. Effron, K., Rechavi, G. Ben-Neriah, Y, Zakut, R. and D. Givol (1982). Simple DNA sequences in homologous flanking regions near immunoglobulin V_h genes: A role in gene interaction? *Nuc. Acids Res.* 10: 3353-3370.

Cohn, V. H. (1985). Organization and evolution of the alcohol dehydrogenase gene in *Drosophila*. Ph.D. thesis, University microfilms, Ann Arbor, Mich.

Coyne, J. A. (1985). Genetic studies of three sibling species of *Drosophila* with relationship to theories of speciation. *Genet. Res.* 46: 169-195.

Coyne, J. A. and M. Kreitman (1986). Evolutionary genetics of two sibling species, *Drosophila simulans* and *Drosophila sechellia*. *Evolution* 40: 673-691.

Dainou, O., Cariou, M. L., David, J. R. and D. A. Hickey (1987). Amylase gene duplication: an ancestral trait in the *Drosophila melanogaster* species subgroup. *Heredity* 59: 245-251.

Davis, R. and N. Davidson (1984). Isolation of the *Drosophila melanogaster* Duncce chromosomal region and recombinational mapping of Duncce sequences with

restriction site polymorphism as genetic markers.

Molec. Cell. Biol. 4: 358-367.

Dawson, W. D. and L. L. Huang (1981). Comparative genetics of hamster amylases. Biochem. Genet. 19:623-633.

De Jong, G. and W. Scharloo (1976). Environmental determination of selective significance or neutrality of amylase variants in Drosophila melanogaster. Genetics 84: 77-94.

Doane, W. W. (1967). Cytogenetic and biochemical studies of amylase in Drosophila melanogaster. Am. Zool. 7:780.

Doane, W. W. (1969a). Amylase variants in Drosophila melanogaster: linkage studies and characterization of enzyme extracts. J. Exp. Zool. A1: 321-341.

Doane, W. W. (1969b). Drosophila amylases and problems in cellular differentiation. In: Problems in Biology: RNA in Development E. W. Hanley ed. pp. 73-109. Univ. of Utah Press, Salt Lake City.

Doane, W. W. (1980). Midgut amylase activity patterns in Drosophila: Nomenclature. Drosophila information Service. 55: 36-39.

Doane, W. W., Abraham, I., Kolar, M., Martenson R. F. and G. E. Deibler (1975). Purified Drosophila alpha-amylase isozymes: genetical, biochemical and molecular characterization. In: Isozymes Vol. 4: Genetics and Evolution. C. L. Markert ed. pp.585-607.

- Doane, W. W., Treat-Clemons, L. G., Gemmill, R. M., Levy, J. N., Hawley, S. A., Buchberg, A. M. and K. Paigen (1983). Genetic mechanisms for tissue-specific control of alpha-amylase expression in Drosophila melanogaster. In: Isozymes: Gene Expression and Development. Vol.9. Rattazzi, Zm. Zl. et al. eds., Alan Liss. Inc. N. Y. pp.63-90.
- Doane, W. W., Gemmill, R. M., Schwartz, P. E., Hawley, S. and R. A. Norman (1987). Structural organization of the alpha-amylase gene locus in Drosophila melanogaster Drosophila miranda. Isozymes 14: 229-266.
- Dover, G. (1982). Molecular drive: A cohesive mode of species evolution. Nature 299: 111-117.
- Duane, W. C., Frerichs, R., and M. D. Levitt (1971). Distribution, turnover, and mechanism of renal excretion of amylase in the baboon. J. Clin. Invest. 50:156.
- Duane, W. C., Frerichs, R., and M. D. Levitt (1972). Simultaneous study of the metabolic turnover and renal excretion of salivary amylase- ^{125}I and pancreatic amylase in the baboon. J. Clin. Invest. 50:1504.
- Dube, S. K. and P., Nordin (1961). Isolation and properties of sorgum alpha-amylase. Arch. Biochem. Biophys. 94:121.

- Efstratiadis, A., Posakony, J. W., Maniatis, T., Lawn, C., O'Connel, C., Spritz, R. A., Deriel, J. K., Forget, B. G., Weissman, S. M., Slightom, J. L., Blechl, A. E., Smithies, O., Baralle, F. E., Shoulders, C. C., and N. J. Proudfoot. (1980). The structure and evolution of the human β -globin gene family. *Cell* 21: 653-668.
- Emorine, L., Kuehl, M., Weir, L., Leder, P. and E. E. Max. (1983). A conserved sequence in the immunoglobulin J κ -C κ intron: possible enhancer element. *Nature* 304: 447-449.
- Ewens, W. J., Spielman, R. S., and H. Harris. (1981). Estimation of genetic variation at the DNA level for restriction endonuclease data. *Proc. Natl. Acad. Sci. USA*, 78:37480-3750.
- Gemmill, R. M. , Levy, J. M. and W. W. Doane (1985). Molecular cloning of alpha-amylase genes from Drosophila melanogaster I. Clone isolation by use of a mouse probe. *Genetics* 110:299-312.
- Gilbert, W. (1986). The RNA world. *Nature* 319: 618.
- Gillespie, J. H. (1986). Variability of evolutionary rates of DNA. *Genetics* 113: 1077-1091.
- Gillespie, J. and K. I. Kojima (1968). The degree of polymorphism in enzymes involved in energy production compared to that in non-specific enzymes in two Drosophila ananassae populations. *Proc. Natl. Acad.*

Sci. U.S.A. 61: 582-585.

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

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

Grunstein, M. and D. Hogness. (1975). Colony hybridization: a method for the isolation of cloned DNA's that contain a specific gene. Proc. Natl. Acad. Sci. U.S.A. 72: 3961-3965.

Gubenko, S. I. and M. B. Evgen'Ev (1984). Cytological and linkage maps of Drosophila virilis chromosomes. Genetica 65: 127-139.

Hagenbuechle, O. Bovery, R. and R. A. Young (1980). Tissue-specific expression of mouse alpha-amylase genes: nucleotide sequence of isozyme mRNAs from pancreas and salivary gland. Cell 21: 179-187.

Haldane, J. B. S. (1949). Evolution and disease. Ri. Sci. Suppl. 19:68-76.

Hasselholt, M., Larsen, B., and P. B. Nielson (1966). Studies on serum amylase systems in swine, horses and cattle. In: Royal Veterinary and Agriculture College Yearbook. pp. 78-90. Copenhagen.

- Hayashida, H. and T. Miyata (1983). Unusual evolutionary conservation and frequent DNA segment exchange in class I genes of the major histocompatibility complex. *Proc. Natl. Acad. Sci. U.S.A.* 80: 2671-2675.
- Henri, V. (1902). Théorie générale de l'action de quelques diastases. *C. R. Acad. Sci. (Paris)*. 135:916.
- Hentschel, C.C. (1982). Homocopolymer sequences in the spacer of a sea urchin histone gene repeat are sensitive to S₁ nuclease. *Nature* 295 714-716.
- Hickey, D. A. (1977). Selection for amylase allozymes in Drosophila melanogaster. *Evolution* 31: 800-804.
- Hickey, D. A. (1979a). The geographical pattern of an enzyme polymorphism in Drosophila melanogaster. *Genetica* 51: 1-4.
- Hickey, D. A. (1979b). Selection on amylase allozymes in Drosophila melanogaster: selection experiments using several independently derived pairs of chromosomes. *Evolution*. 33: 1128-1137.
- Hickey, D. A. (1981). Regulation of amylase activity in Drosophila melanogaster: variation in the number of enzyme molecules produced by different amylase genotypes. *Biochem. Genet.* 19: 783-796.
- Hickey, D. A. and B. F. Benkel (1982). Regulation of amylase activity in Drosophila melanogaster: effects of various dietary treatments on amylase production.

- Biochem. Genet. 20: 1117-1129.
- Hickey, D.A. and B.F. Benkel (1985). Splicing and the evolution of introns. *Nature* 316:582.
- Hickey, D. A. and B. F. Benkel (1989). Patterns of molecular evolution in alpha-amylase-coding genes. in *Molecular Evolution* (edits. M. Clegg and S. O'Brien), Alan Liss, New York, in press.
- Hickey, D. A. and B.F. Benkel (1986). Introns as relict retrotransposons: implications for the evolutionary origin of RNA splicing mechanisms. *J. Theoret. Biol.* 121: 283-291.
- Hickey, D. A., Genest, Y. and B. F. Benkel (1987a). Nucleotide sequence upstream of a glucose-repressible *Drosophila* gene. *Nucl. Acids Res.* 15:7184.
- Hickey, D. A., Benkel, B. F., Boer, P. Genest, Y., Abukashawa, S. and G. Ben-David (1987b). Enzyme-coding genes as molecular clocks: The molecular evolution of animal alpha-amylases. *J. Mol. Evol.* 26: 252-256.
- Hickey, D. A., Benkel, B. F. and S. Abukashawa (1988). DNA rearrangement causes multiple changes in gene expression at the amylase locus in *Drosophila melanogaster*. *Biochem. Genet.* 26: 757-768.
- Hickey, D. A., Benkel, B. F. and S. Abukashawa (1989). A general model for the evolution of nuclear pre-mRNA introns. *J. Theor. Biol.* 137: 41-53.

Hickey, D. A., Benkel, B.F. and C. Magoulas (1989).

Molecular biology of enzyme adaptations in higher eukaryotes. *Genome* 31: 272-283.

Hickey, D. A., Bally-Cuif, L., Abukashawa, S. and B. F.

Benkel (1989). Concerted evolution of duplicated protein-coding genes in *Drosophila*. *In Press*.

Higgs, D. R., Goodbourn, S. E., Wainscoat, J. S., Clegg, J.

B., and J. B. Weatherall (1981). Highly variable regions of DNA flank the human globin genes. *Nuc. Acids Res.* 9: 4213-4224.

Hjorth, J. P. (1979). Genetic variation in mouse salivary amylase rate of synthesis. *Biochem.Genet.* 17:665-682.

Holmes, D. S. and M. Quigley (1981). A rapid boiling method for the preparation of bacterial plasmids. *Anal. Biochem.* 114: 193-197.

Hoorn, A. J. W. and W. Scarloo (1978). The functional significance of amylase polymorphism in *Drosophila melanogaster* V. The effect of food components on amylase and alpha-glucosidase activity. *Genetica* 49: 181-187.

Imagawa, M., Chiur, R. and M. Karin (1987). Transcription factor AP-2 mediates induction by two different signal-transduction pathways: protein kinase C and cAMP. *Cell* 51:251-260.

- Jackson, J. A. and G. R. Fink (1981). Gene conversion between duplicated genetic elements in yeast. *Nature* 292: 306-311.
- Jeffreys, A. J. (1979). DNA sequence variants in the Ggamma globin genes of man. *Cell* 18: 1-10.
- Johnson, G. B. (1974). Enzyme polymorphism and metabolism. *Science*. 184:28-37.
- Jones, C. W. and F. C. Kafatos (1982). Accepted mutations in a gene family: evolutionary diversification of duplicated DNA. *J. Molec. Evol.* 19: 87-103.
- Jukes, T. H. and C. R. Cantor (1969). Evolution of protein molecules, pp. 21-123. In: *Mammalian Protein Metabolism*, H. N. Munro ed., Academic Press, New York.
- Jukes, T. H. and J. L. King (1979). Evolutionary nucleotide replacements in DNA. *Nature* 281: 605-606.
- Kafatos, F. C., A. Efstratiadis, B. G. Forget, and S. M. Weissman (1977). Molecular evolution of human and rabbit β -globin mRNAs. *Proc. Natl. Acad. Sci.* 74: 5618-5622.
- Kaplan, R. D., Chapman, V., and F. H. Ruddle (1973). Electrophoretic variation of alpha amylase in two inbred strains of Mus musculus *Heredity* 64:155
- Kaplan, N.L. and R.R. Hudson (1987). On the divergence of genes in a multigene family. *Theoret. Pop. Biol.* 31: 178-194.

- Karn, R. C., Rosenblum, B. B., Ward, J., and A. D. Merritt (1975). Genetic and post-translational mechanisms determining human amylase isozyme heterogeneity. In: *Isozymes*, Vol. IV: Genetics and Evolution, C. L. Markert, ed., pp. 745-761, Academic Press, New York.
- Karn, R. C. and G. M. Malacinski (1978). in: *ADVANCED COMPARATIVE PHYSIOLOGY AND BIOCHEMISTRY*, Lowenstein, O. Ed. pp. 1-103, Academic Press, New York.
- Kassis, J., Poole, S., Wright, K., and P. H. O'Farrel. (1986). Sequence conservation in the protein coding and intron regions of the engrailed transcription unit. *EMBO J.* 3583-3589.
- Kelly, K. A., and P. M. Pitha (1985). Characterization of a mouse interferon gene locus. II. Differential expression of α -interferon genes. *Nuc. Acids Res* 13: 825-839.
- Kikkawa, H. (1960). Further studies on the genetic control of amylase in Drosophila melanogaster. *Jpn. J. Genet.* 35:382-387._
- Kikkawa, H. (1964). An electrophoretic study on amylase in Drosophila melanogaster. *Jpn. J. Genet.* 39:401-411.
- Kikkawa, H. and K. Abe (1960). Genetic control of Amylase in Drosophila melanogaster. *Jpn. J. Genet.* 39: 401-411.
- Kimura M. and T. Ohta (1972). Theoretical Aspects of Population Genetics. Monographs in Population Biology,

4. pp. 219, Princeton university Press, Princeton, New York.
- Kimura, M. and T. Ohta (1974). On some principles governing molecular evolution. *Proc. Natl. Acad. Sci. U.S.A.* 71:2848-2852. .
- King, J. L. and T. H. Jukes (1969). Non-Darwinian evolution: random fixation of selectively neutral mutations. *Science* (Washington, D. C.). 164: 788-798.
- Klein, H.L. (1984). Lack of association of intrachromosomal gene conversion and reciprocal exchange. *Nature* 310: 748-753.
- Korsurd, F.R. and P.Brandtzaeg (1982). Characterization of epithelial elements in human major salivary glands by functional markers. *Histochem. Cytochem.* 30:657-666.
- Krawinkel, U., Zoebelen, G., Bruggemann, M., Radbruch, A. and K. Rajewsky (1983). Recombination between heavy chain variable-region genes: evidence for gene conversion. *Proc. Natl. Acad. Sci. USA* 80: 4997-5001.
- Kreitman, M. (1983). Nucleotide polymorphism at the alcohol dehydrogenase gene of Drosophila melanogaster. *Nature* 304: 412-417.
- Kreitman, M. and M. Aguade (1986). Genetic uniformity in two populations of Drosophila melanogaster as revealed by filter hybridization of four-nucleotide-recognizing restriction enzyme digests. *Proc. Natl. Acad. Sci.*

U.S.A. 83: 3562-3566.

Kuhn, R. (1925). Der Wirkungsmechanismus der amylasen; ein Beitrag zum Konfigurationsproblem der Stärke. Ann. Chem. 443:1-10.

Lambowitz, A. (1989). Infectious introns. Cell 56: 323-326.

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

Langley, C. H., Shrimpton, A. E., Yamazaki, T., Miyashita, N., Matsuo, Y. and C. F. Aquadro (1988). Naturally occurring variation in the restriction map of the Amy region of Drosophila melanogaster. Genetics 119: 619-

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

Leuchs, E. (1831). Ueber die Verzuckerung des Stärkmehls durch Speichel. Arch. Chem. Meteorol. 3:105-116.

Levine, J. and A. Spradling (1985). DNA sequence of a 3.8 kilobase pair region controlling Drosophila chorion gene amplification. Chromosoma 92: 136-142.

Levy, J. N. , Gemmill, R. M. and W. W. Doane (1985). Molecular cloning of alpha-amylase genes from Drosophila melanogaster II. Clone verification and organization. Genetics 110:313-324.

Lewontin, R. C. (1964). The interaction of selection and linkage. I. General considerations: heterotic models. *Genetics* 49:49-67.

Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.

Lewontin, R. C. (1988). Inferring the number of evolutionary events from DNA coding sequence differences. *Mol. Biol. Evol.* 6: 15-32.

Li, W.-H, C.-I. Wu, and C.-C. Luo (1985). A new method for estimating synonymous and nonsynonymous rates of nucleotide substitution considering the relative likelihood of nucleotide and codon changes. *Mol. Biol. Evol.* 2: 150-174.

Lieberhaber, S., Goossens, M. and Y. W. Kan (1981). Homology and concerted evolution at the alpha-1 and alpha-2 loci of human gamma-globulin. *Nature* 290: 26-29.

Lindsey, D. L. and E. H. Grell (1968). *Genetic Variations of Drosophila melanogaster*. Carnegie Institution of Washington, Baltimore, MD. Publ. No. 627.

Liskay, R.M. and J.L. Stachelek (1983). Evidence for intrachromosomal gene conversion in cultured mammalian cells. *Cell* 35: 157-165.

Liskay, R.M. and J.L. Stachelek (1986). Information transfer between duplicated chromosomal sequences in

- mammalian cells involves contiguous regions of DNA.
Proc. Natl. Acad. Sci. USA 83: 1802-1806.
- Long, C. M., Virolle, M., Chang, S-Y., Chang, S. and M. J. Bibb. (1987). Alpha-amylase gene of Streptomyces limosus: Nucleotide sequence, expression motifs and amino acid sequence homology to mammalian and invertebrate alpha-amylases. *J. Bacteriology* 169: 5745-5754.
- Magoulas C. and D.A. Hickey (1989). The interaction between glucose and cyclic AMP on the expression of amylase genes in *Drosophila*. in prep.
- Malaciniski, G. M., and W. J., Rutter (1969). Multiple molecular forms of alpha amylase from the rabbit. *Biochemistry*. 8:4382
- Mandel, M. and A. Higa (1970). Calcium-dependent bacteriophage DNA infection. *J. Mol. Biol.* 53: 159-162.
- Maniatis, T., Fritsch, E. F. and J. Sambrook (1982). *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, N. Y.
- Maruyama, T., Gojobori, T., Aota, S., and T. Ikemura (1986). Codon usage tabulated from the Genbank genetic sequence data. *Nuc. Acids Res.* 14: r151-r197.
- Matsumura, S. (1934). Genetical and physiological studies on amylase functions in digestive and body fluids of Bombyx mori. *Bull. Nagano Sericult. Exp. Stn.* 28:1-124.

124.

McGeachin, R. L. and J. R. Akin (1982). *Biochem. Physiol.*

(A) 72:267-269.

Meisler, M. H. and D. L. Gumucio (1986). Pancreatic

amylase: molecular genetics and evolution. In:

Molecular and Cellular Basis of Digestion, P.

Desnuelle, H. Sjöstrom and O. Noren eds. Elsevier Sci.

Publishers.

Michaelis, L. and H., Pechstein (1914). Die

Wirkungsbedingungen der speicheldiastase. *Biochem.Z.*

59:77-83.

Miyata, T. and T. Yasunaga (1980). Molecular evolution of

mRNA: A method for estimating evolutionary rates of

synonymous and amino acid substitutions from homologous

nucleotide sequences and its application. *J. Mol. Evol.*

16: 23-26.

Miyata, T., Yasunaga, T. and T. Nishida (1980). Nucleotide

sequence divergence and functional constraint in mRNA

evolution. *Proc. Natl. Acad. Sci. USA* 77: 7328-7332.

Nagylaki, T. (1984a). Evolution of multigene families under

intrachromosomal gene conversion. *Genetics* 106: 529-

548.

Nagylaki, T. (1984b) Evolution of multigene families under

interchromosomal gene conversion. *Proc. Natl. Acad.*

Sci. USA 81: 3796-3800.

- Nagylaki, T. and T.D. Petes (1983). Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics* 100: 315-337.
- Nagylaki, T. and N. Barton (1986). Intrachromosomal gene conversion, linkage, and the evolution of multigene families. *Theoret. Popul. Biol.* 29 407-437.
- Nakajima, R., Imanaka, T. and S. Aiba (1986). Comparison of amino acid sequences of eleven different alpha-amylases. *Appl. Microbiol. Biotechnol.* 23: 355-360.
- Nakamura, Y., Ogawa, M.; Nishide, T.; Emi, M.; Kosaki, G.; Himeno, S. and K. Matsubara (1984). Sequences of cDNAs for human salivary and pancreatic alpha-amylases. *Gene* 28:263-270.
- Nei, M. (1987). Molecular Evolutionary Genetics, Columbia Univ. Press, New York.
- Nei, M. and Li, W. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc. Natl. Acad. Sci. USA.* 76:5269-5273.
- Nei, M. and F. Tajima (1981). DNA polymorphism detectable by restriction endonucleases. *Genetics* 97: 145-163.
- Neilson, J. T. (1969). Genetic studies of the amylase isozymes of the bank vole, Clethrionomys glareola. *Hereditas* 61:400-410.
- Neilson, J. T. and K. Sick (1975). Genetic polymorphism of amylase isozymes in feral populations of the house

- mouse. *Hereditas* 79:279
- Ohno, S. (1970). *Evolution by Gene Duplication*. Springer-Verlag, Berlin and New York.
- Ohta, T. (1980). Evolution and Variation in Multigene Families, Springer-Verlag, Berlin.
- Ohta, T. (1985). A model of duplicative transposition and gene conversion for repetitive DNA families. *Genetics* 110: 513-524.
- Ohta, T. (1986). Actual number of genes contained in a multigene family. *Genet. Res.* 48: 119-123.
- Okuyama, E. and T. Yamazaki (1988). Nucleotide sequence of the duplicated amylase structural genes in Drosophila melanogaster. *Proc. Japan Acad.* 64 serB:274-277.
- Patterson, J. T. and W. S. Stone (1952). *Evolution in the Genus Drosophila*. Macmillan, New York.
- Payant, V., Abukashawa, S., Sasseville, M., Benkel, B. F., Hickey, D. A. and J. David (1988). Evolutionary conservation of the chromosomal configuration and regulation of amylase genes among eight species of the Drosophila melanogaster subgroup. *Mol. Biol. Evol.* 5: 560-567.
- Pease, L. R., Schultz, D. H., Pfaffenbach, G. M. and S. G. Nathenson (1983). Spontaneous H-2 mutants provide evidence that a copy mechanism analogous to gene conversion generates polymorphism in the major

histocompatibility complex. *Proc. Natl. Acad. Sci.* 80:
242-246.

Powell, J. R. (1975). Protein variation in natural
populations of animals. *Evol. Biol.* 19: 79-119

Powell, J.R. and J.M. Lichtenfels (1979). Population
genetics of Drosophila amylase. I. Genetic control of
tissue-specific expression in D. pseudoobscura.
Genetics 92: 603-612.

Queen, C. and L. J. Korn (1984). A comprehensive sequence
analysis for the IBM personal computer. *Nuc. Acids Res.*
12: 581-599.

Robyt, J. F. and R. J. Ackerman (1973). Structure and
function of amylases. II. Multiple forms of Bacillus
subtilis alpha amylase. *Arch. Biochem. Biophys.* 155:445

Roesler, W., Vanderbank, G. R. and R. Hanson (1988). Cyclic
AMP and the induction of eukaryotic gene transcription.
J. Biol. Chem. 263: 9063-9066.

Rogers, J. C. (1985). Conserved amino acid sequence domains
in alpha-amylase in plants, mammals and bacteria.
Biochem. Biophys. Res. Commun. 128: 470-476.

Rogers, J. C. and C. Milliman (1983). Isolation and
sequence analysis of a barley alpha-amylase cDNA clone.
J. Biol. Chem. 254: 8169-8147.

Romer, A. S. (1962). *THE VERTEBRATE BODY*, 3rd edition,
Saunders, Philadelphia, pp. 232-233.

- Sanger, F., Nicklen, S. and A. R. Coulson (1977). DNA sequencing with chain-termination inhibitors. *Proc. Natl. Acad. Sci. U.S.A.* 74: 5463-5467.
- Schaeffer, S. W. and C. F. Aquadro (1987). Nucleotide sequence of the alcohol dehydrogenase region of *Drosophila pseudoobscura*: evolutionary change and evidence for an ancient duplication. *Genetics* 117: 61-73.
- Schaeffer, S.W., Aquadro, C.F. and C.H. Langley (1988). Restriction-map variation in the Notch region of *Drosophila melanogaster*. *Mol. Biol. Evol.* 5: 30-40.
- Schibler, U., Pittet, A., Young, R., Hagenbuchle, O., Tosi, M., Gelman, S., and P. K. Wellauer. (1982). The mouse alpha-amylase multigene family: sequence organization of members expressed in the pancreas, salivary glands and liver. *J. Mol. Biol.* 155:247-266.
- Schibler, U. Hagenbuchle, O., Wellauer, P. K. and A. G. Pittet (1983). Two promoters of different strengths control the transcription of the mouse alpha-amylase gene Amy-1^a in the parotid gland and the liver. *Cell* 33: 501-508.
- Shear, M., Gibson, S. and E. Van der Merwe (1973). A qualitative and quantitative histochemical investigation of the effects of pilocarpine stimulation on amylase levels in rat submaxillary and parotid

- glands. *J.Histochem. Cytochem.* 21:661-673.
- Scherer, S. and R.D. Davis (1980). Recombination of dispersed repeated DNA sequences in yeast. *Science* **209**: 1380-1384.
- Shimizu, A. (1985). Diffusion approximation of an infinite allele model incorporating gene conversion. pp 243-255. in: Population Genetics and Molecular Evolution, (edits T. Ohta and K. Aoki), Jap. Sci. Soc. Press, Tokyo.
- Shimizu, A. (1987). Stationary distribution of a diffusion process taking values in probabilitydistributions on the partitions. pp 100-114. in: Stochastic Methods in Biology, (edits. M. Kimura, G.Kallianpur and T. Hida), Springer, Berlin.
- Silhavy, T. J., Berman, M. L. and L. W. Enquist (1984). Experiments With Gene Fusions. Cold Spring Harbor Laboratory, New York, pp. 8990, 140-141.
- Silver, B.J., Bokar, J. A., Virgin, J. B., Vallen, E. A., Milsted, A. and J. H. Nilson (1987). Cyclic AMP regulation of the human glycoptotein hormone A-subunit gene is mediated by an 18-base-pair element. *Proc. Natl. acad. Sci. U.S.A.* 84: 2198-2202.
- Singh, R. S., Hickey, D. A. and J. David (1982). Genetic differences between geographically distant populations of Drosophila melanogaster. *Genetics* 101:519-531.

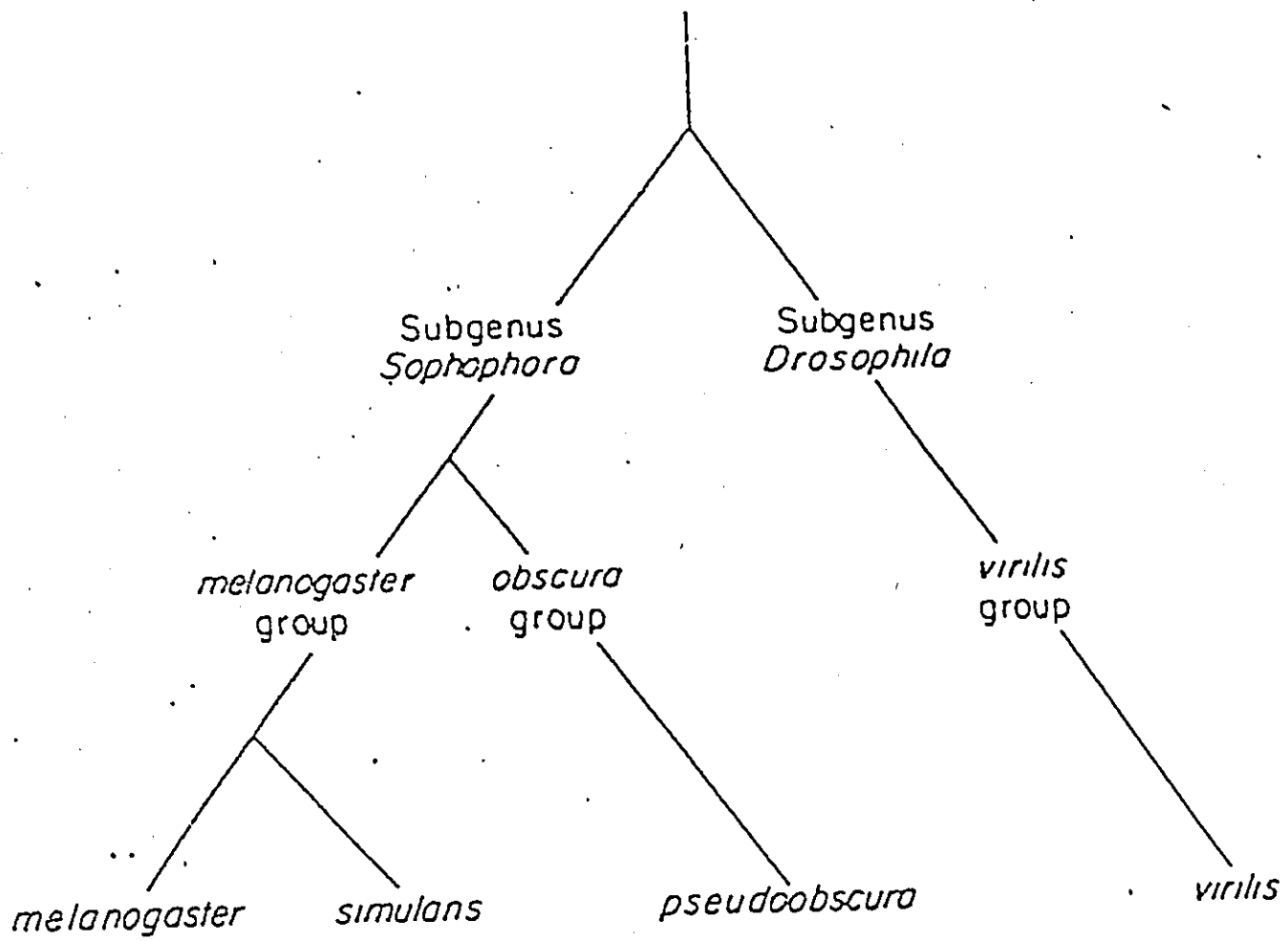
- Sligtom, J.L., Blechl, A.E. and O. Smithies (1980). Human fetal G γ - and A γ -globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* 21: 627-638.
- Sokal, R. and Sneath, P.H.A. (1973) Numerical Taxonomy. Freeman.
- Stein, E. A., Jung, J. M., and E. H. Fisher (1960). The amino acid composition of alpha amylase from Aspergillus oryzae. *J. Biol. Chem.* 235:371
- Streisinger, G. (1966). Frame shift mutations and the genetic code. *Cold Spring Harbour Symp. Quant. Biol.*31: 77-84.
- Strobel, E., Dunsmuir, P., and G. M. Rubin (1979). Polymorphisms in the chromosomal locations of elements of the 412, copia and 297 dispersed repeated gene families in Drosophila. *Cell* 17: 429-439.
- Sturtevant, A. H. and E. Novitski (1941). The homologies of the chromosome elements in the genus Drosophila. *Genetics* 26: 517-541.
- Subramani, S. and J. Rubnitz (1985). Recombination events after transient infection and stable integration of DNA into mouse cells. *Mol. Cell. Biol.* 5: 659-666.
- Tachida, H. (1987). Differentiation of a multigene family between populations. *Evolution* 41: 190-198.

- Takkinen, K., Pettersson, R.F., Kalkkinen, N., Palva, I., Soderlund, H. and L. Kaariainen (1983). Amino acid sequence of α -Amylase from Bacillus amyloliquefaciens deduced from the nucleotide sequence of the cloned gene. *J. Biol. Chem.* 258: 1007-1013.
- Tautz D. and M. Renz. (1984). Simple DNA sequences of Drosophila virilis isolated by screening with RNA J. *Mol. Biol.* 172 229-235.
- Thoma, J. A., Spradlin, J. E., and S. Dygert (1970). Plant and animal amylase. In: *The Enzymes*, Vol. V, 3rd ed. (p. D. Boyer, Ed.), pp. 115-189. Academic Press, New York.
- Toda, H., Konodo, K. and K. Narita (1983). The complete amino acid sequence of Taka-amylase A. *Proc. Jpn. Acad.* 58: 208-212.
- Throckmorton, L. H. (1975). The virilis species group, pp. 227-296 In: *The Genetics and Biology of Drosophila*. Vol. 3b., M. Ashburner, H. L. Carson and N. J. Thompson eds. Academic Press, London.
- Tsuda, M. and Y. Suzuki (1981). Faithful transcription initiation of fibroin gene in a homologous cell-free system reveals an enhancing effect of 5' flanking sequence far upstream. *Cell* 27 175-182.
- Walsh, J.B. (1986). Selection and biased gene conversion in a multigene family: consequences of interallelic bias

- and threshold selection. *Genetics* 112: 699-716.
- Walsh, J.B. (1987). Sequence-dependent gene conversion: can duplicated genes diverge fast enough to escape conversion? *Genetics* 117: 543-557.
- Walsh, J. B. (1988). Unusual behaviour of linkage disequilibrium in two-locus gene conversion models. *Genet. Res.* 51: 55-58.
- Weir, B. S., Ohta, T. and H. Tachida (1986). Gene conversion models *J. Theor. Biol.* 116: 1-8.
- Weiss, E., Golden, L., Zakut, R., Mellor, A., Fahrnerk, K., Vist, S. and R. A. Flavell (1983). The DNA sequence of the H-2K^b gene: Evidence for gene conversion as a mechanism for the generation of polymorphism in histocompatibility antigens. *EMBO J* 2: 453-462.
- Wheeler, M. R. (1981). A note on some fossil Drosophilidae from the amber of Chiapas, Mexico. *J. Paleontol.* 37: 123-124.
- Wong, Y-C. Putsell, J., Spoerel, N. and F. C. Kafatos (1985). Coding and potential regulatory sequences of a cluster of chorion genes in drosophila melanogaster. *Chromosoma* 92: 124-135.
- Wyman, A. R. and R. White (1980). A highly polymorphic locus in human DNA. *Proc. Natl. Acad. Sci. U.S.A.* 77: 6754-6758.

- Yamazaki, T. and Y. Matsuo (1984). Genetic analysis of natural populations of Drosophila melanogaster in Japan: III. Genetic variability of inducing factors of amylase and fitness. Genetics 108: 223-235.
- Yang, M. A., Galizzi, A., and D. Henner (1983). Nucleotide sequence of the amylase gene from Bacillus subtilis. Nuc. Acids Res. 11: 237-249.
- Zwiebel, L. J., V. H. Cohn, D. R. Wright and G. P. Moore (1982). Evolution of single-copy DNA and the ADH gene in seven drosophilids. J. Mol. Evol. 19: 62-71.

APPENDIX 1.1 Phylogeny of the Drosophila species employed in this study. The Drosophila and Sophophora subgenera are thought to have separated in the Eocene, about 40 million years ago: the Melanogaster and Obscura groups separated in the Oligocene, about 30 million years ago (Throckmorton, 1975). Distances are not to scale. Phylogeny is reproduced from Blackman and Meislsen (1987).



APPENDIX 2.1

SER. ORIGINAL		AMYLASE PHENOTYPE	AMYLASE PHENOTYPE OF SINGLE FLIES				
NO.	REFERENCE	5 FLIES HOMOGENATE	#1	#2	#3	#4	#5
1	Benin 1C	1,4,5	1,4,5	1,4,5	1,5	1,4,5	
2	Benin 1CA	4,5	4,5	4,5	4,5		
3	Benin 3	4,5,6	4,5,6	5,6	5,6		
4	Benin 3A	1,4,5	1,4,5	4	4		
5	Benin 3C	1,5,6	1,6	1,6	1		
6	Benin 3CA	1,5	1,5	1,5			
7	Benin 4E	5,6	5,6	5,6			
8	Benin 4EA	5,6	5,6	5,6			
9	Benin 5C	1,5,6	5,6	1,5,6	5,6	5,6	
10	Benin 5CA	1,3,4,5	3,4,5	3,5	3,5	5	
11	Benin 6CA	1,3,4,5,6	1,4,5,6	1,4,5,6			
12	Benin 7CA	1	1	1			
13	Benin 9C	3,4,5	3,5	3,5	3,5	3,5	
14	Benin 9CA	1	1	1			
15	Benin 9A	1,6	1,6	1,6	1,6		
16	Benin 10C	1,2,5,6	1,5,6	1,5,6	5,6	1	
17	Benin 10CA	1,2	1,2	1,2			
18	Benin 11	4,5	4,5	4,5	4,5		
19	Benin 11A	1,2,4,5	1,4,5	1,4,5	4,5	1	1
20	Benin 16C	1 (see #70)	1	1			
21	Benin 17	2,3,5	2,3,5	2,3,5	2,3	3,5	
22	Benin 17C	2,3	2,3	2,3	2,3	2,3	
23	Benin 18C	2,4,5	2,5	2,5	2,5	2,5	
24	Benin 18CA	2,4,5	2,5	2,5	2,5	2,5	
25	Benin 19C	1,3,4,6	1,3,4,6	1,3,4,6	1,3,4,6	1,3,4,6	
26	Benin 19CA	3,4,5,6	3,4,5,6	3,4	3,4	4,5,6	
27	Benin 20C	1,4,5	1,5	1,5	1,5		

174

28	Benin 20CA	1,4,5	1,5	1,5	5	5
29	Benin 22C	4,5,6	4,5	4,5	5	5

SER. NO.	ORIGINAL REFERENCE	AMYLASE PHENOTYPE 5 FLIES HOMOGENATE	AMYLASE PHENOTYPE OF SINGLE FLIES				
			#1	#2	#3	#4	#5
30	Benin 23	1,6	1,6	1,6	1		
31	Benin 23C	1	1	1			
32	Benin 23CA	1	1	1			
33	Benin 24C	4,6	4,6	4,6	4,6	4,6	
34	Benin 25	4,5,6	4,6	4,6	4,6	4,6	
35	Benin 25A	4,5,6	4,5,6	4,5,6	4,5,6	4,6	
36	Benin 25C	1,4,5,6	1,4,5,6	1,4,5,6	1,6	1,6	
37	Benin 25CA	3,5,6	3,5,6	3,5,6	6	6	6
38	Benin 27	5,6	5,6	5,6	5,6	5,6	
39	Benin 27A	4,5,6	5,6	5,6	5,6	5,6	5
40	Benin 30	1,4,5	1,4,5	1,5	1	1,4,5	
41	Benin 30A	1,4,5	1,5	5	1,5	5	
42	Benin 31A	1,4,5,6	1,4,6	4,6	1,4,5		
43	Benin 37	1,6	1,6	1,6	1,6		
44	Benin 37A	4,5	4,5	4,5	4,5	4,5	
45	SM5	1					
46	Lagos 2	5,6	5,6	5,6	5,6		
47	Lagos 3	1,3,6	3,6	3,6			
48	Lagos 7A	1,4,5,6	5,6	5	4,5,6	5	
49	Lagos IK	3,5,6	5,6	3,5,6	3,6	3,6	
50	Logos BG-1	1,3,6	1,3,5,6	1	1		
51	Tunisia 215	1	1	1			
52	Tunisia 216	1	1	1			
53	Tunisia 224	1	1	1			
54	Tunisia 250	1	1	1			
55	Tunisia 274	1	1	1			
56	Tunisia 277	1	1	1			
57	Italy 8I	1,3	1,3	1,3	1		

SER. NO.	ORIGINAL REFERENCE	AMYLASE PHENOTYPE		AMYLASE PHENOTYPE OF SINGLE FLIES				
		5 FLIES	HOMOGENATE	#1	#2	#3	#4	#5
58	Italy 61I	1		1	1			
59	Italy 62I	1		1	1			
60	Italy 108I	1		1	1			
61	Oregon R	1		1	1			
62	Ottawa 24	1		1	1			
63	Bronswville	1,3	*(1,2)	1,3	1	1	1,3	1,3
64	Makokou	1,2		1,2	1,2	1,2		
65	Ottawa 37	1,3		1,3	1,3	1,3	1,3	
66	Ippassa	3		3	3	3		
67	Libreville	3		3	3	3		
68	Kumba	4		4	4	4		
69	Ottawa 31	4		4	4	4		
70	Benin 16C	5,6	*(4,5)	5,6	5,6	5,6	5,6	
71	K-11	1	*(4,6)	1	1	1		
72	Makokou	4,6		4,6	4,6	4,6	4,6	
73	Cotonou	5,6		5,6	5,6	5,6	5,6	
74	David's	5,6,7		5,6,7	5,6,7	6,7	5,6,7	
		5,6,7						
75	Recomb.	1	*(1,6)	1	1	1	1	
101	D.H. Ottawa	1,3,6		1,3	1	1	1	
102	D.H. Ottawa	1		1	1			
103	D.H. Ottawa	1		1	1			
104	D.H. Ottawa	1		1	1			
105	D.H. Ottawa	1		1	1			
106	D.H. Ottawa	1		1	1			
107	D.H. Ottawa	1,2,6		1,6	1	1	1,6	
108	D.H. Ottawa	1		1	1			
109	D.H. Ottawa	1,2,5		1,2,5	1,2,5	1	1	
110	D.H. Ottawa	1		1	1			

SER. NO.	ORIGINAL REFERENCE	AMYLASE PHENOTYPE 5 FLIES HOMOGENATE	AMYLASE PHENOTYPE OF SINGLE FLIES				
			#1	#2	#3	#4	#5
111	D.H. Ottawa	1,6	1,6	1	1		
112	D.H. Ottawa	1	1	1			
113	D.H. Ottawa	1	1	1			
114	D.H. Ottawa	4	4	4	4	1,4	
115	D.H. Ottawa	1,3	1,3	1,3	1	1	
116	D.H. Ottawa	1	1	1			
117	D.H. Ottawa	1	1	1			
118	D.H. Ottawa	1	1	1			
119	D.H. Ottawa	1	1	1			
120	D.H. Ottawa	1,3	1,3	1,3	1	1	
121	D.H. Ottawa	1,6	1,6	1	1	1	
122	D.H. Ottawa	1	1	1			
123	D.H. Ottawa	1	1	1			
124	D.H. Ottawa	1	1	1			
125	D.H. Ottawa	1,2,3	1,2	1	1	1	
126	D.H. Ottawa	1	1	1			
127	D.H. Ottawa	1	1	1			
128	D.H. Ottawa	1	1	1			
129	D.H. Ottawa	1,3	1,3	1	1	1	
130	D.H. Ottawa	1	1	1			
131	D.H. Ottawa	1	1	1			
132	D.H. Ottawa	1,3,6	1,3,6	1,3,6	1,6		
133	D.H. Ottawa	1	1	1	1		
134	D.H. Ottawa	1	1	1	1		
135	D.H. Ottawa	1	1	1			
136	D.H. Ottawa	1,3	1,3	1,3	1		
137	D.H. Ottawa	1	1	1			
138	D.H. Ottawa	1	1	1			

SER. NO.	ORIGINAL REFERENCE	AMYLASE PHENOTYPE 5 FLIES HOMOGENATE	AMYLASE PHENOTYPE OF SINGLE FLIES				
			#1	#2	#3	#4	#5
139	D.H. Ottawa	1	1	1			
201	G.C. Ottawa	1	1	1	1		
202	G.C. Ottawa	2	1	1	1		
203	G.C. Ottawa	3	1,6				
204	G.C. Ottawa	4	1	1	1		
205	G.C. Ottawa	5	1	1	1		
206	G.C. Ottawa	6	1	1	1		
207	G.C. Ottawa	7	1	1	1		
208	G.C. Ottawa	8	1	1	1		
209	G.C. Ottawa	9	1	1	1		
210	G.C. Ottawa	10	1,2,6				
211	G.C. Ottawa	11	1	1	1		
212	G.C. Ottawa	12	1,2,3	2,3	1,2,3	1	2,3
213	G.C. Ottawa	13	1	1	1		
214	G.C. Ottawa	13	1	1	1		
215	G.C. Ottawa	15	1	1	1		
216	G.C. Ottawa	16	1,2	1,2,3	1,2,6	1,2,6	
217	G.C. Ottawa	17	1	1	1		
218	G.C. Ottawa	18	1	1	1		
219	G.C. Ottawa	19	1	1	1		
220	G.C. Ottawa	20	1	1	1		
221	G.C. Ottawa	21	1,3	1	1	1,3	
222	G.C. Ottawa	22	1	1	1		
223	G.C. Ottawa	23	1,2,3	1,2,3			
224	G.C. Ottawa	24	1,2,6				
225	G.C. Ottawa	25	1,2	1,2	1	1	
226	G.C. Ottawa	26	1	1	1		
227	G.C. Ottawa	27	1	1	1		

SER. NO.	ORIGINAL REFERENCE	AMYLASE PHENOTYPE 5 FLIES HOMOGENATE	AMYLASE PHENOTYPE OF SINGLE FLIES				
			#1	#2	#3	#4	#5
228	G.C. Ottawa	28	1	1	1		
229	G.C. Ottawa	29	1	1	1		
230	G.C. Ottawa	30	1	1	1		
231	G.C. Ottawa	31	1	1	1		
232	G.C. Ottawa	32	1	1	1		
233	G.C. Ottawa	33	1	1	1		
234	G.C. Ottawa	34	1	1	1		
235	G.C. Ottawa	35	1,2,6	1,2,6			
236	G.C. Ottawa	36	1	1	1		
237	G.C. Ottawa	37	1	1	1		
238	G.C. Ottawa	38	1	1	1		
239	G.C. Ottawa	39	1,2,3	2,3	1,2,3	1,2,3	
240	G.C. Ottawa	40	1	1	1		
241	G.C. Ottawa	41	1	1	1		
242	G.C. Ottawa	42	Flies not available				
243	G.C. Ottawa	43	1,2,3	3			
244	G.C. Ottawa	44	1	1	1		
245	G.C. Ottawa	45	1	1	1		
246	Canton-S	1,3	1,3	1,3	1,3		
247	V-Mauritius Island	1	1	1	1		
248	V-Sycheylles Island	1	1	1	1		
249	<u>D. Orena</u>	9					
250	<u>D. Pseudoobscura:</u>						
	Bogota	9					
	Strawberry Canyon	8					

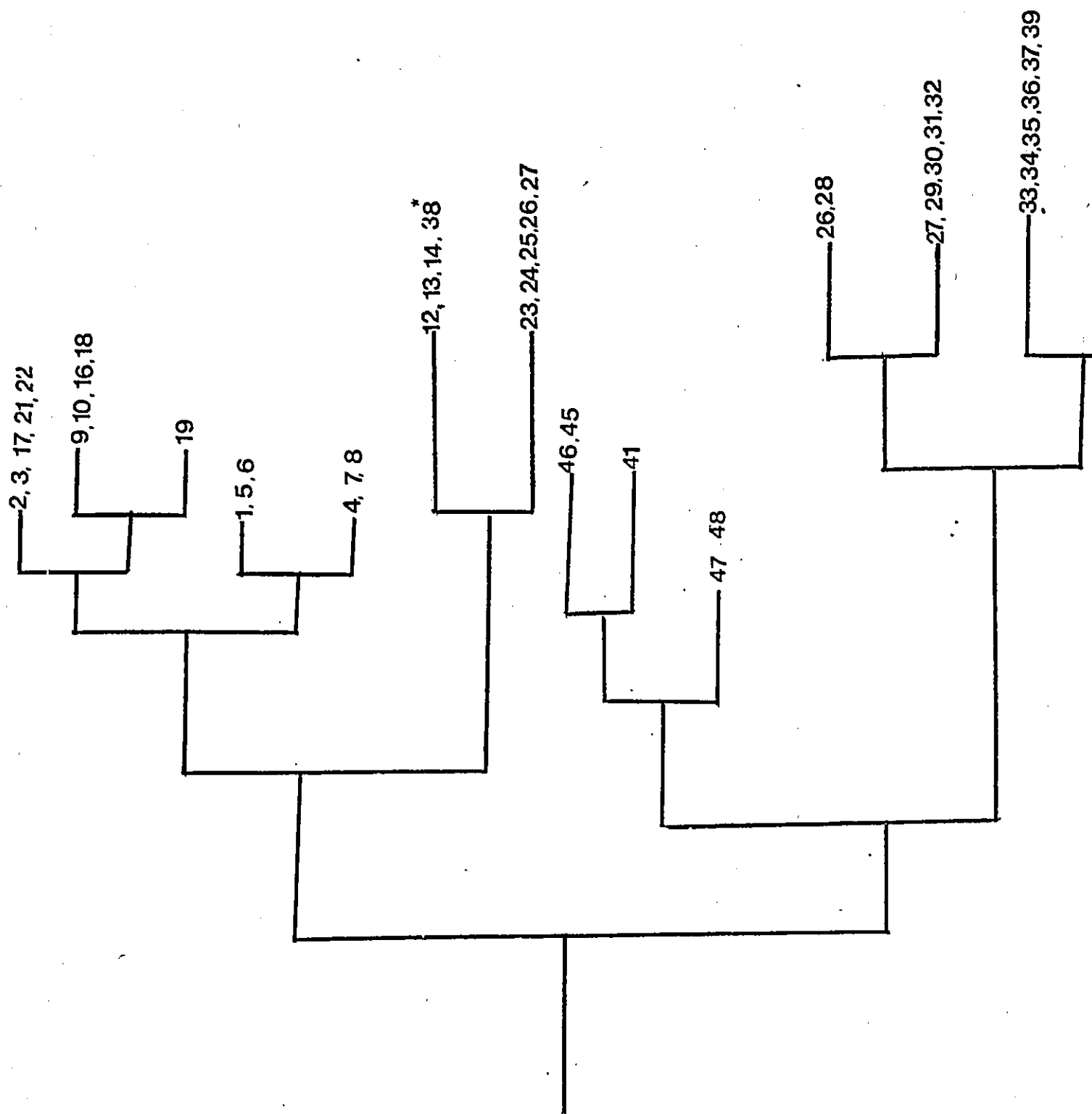
SER. NO.	ORIGINAL REFERENCE	AMYLASE PHENOTYPE 5 FLIES HOMOGENATE	AMYLASE PHENOTYPE OF SINGLE FLIES				
			#1	#2	#3	#4	#5
251	Cairo 1	1,2,3	1	1	1	1,3	1,3
252	Cairo 2	1,5	1,5	1,5	1,5	1,5	1,5

Original reference refers to the designation given by collectors. #1, #2....#5 refer to the number of trials to score amylase phenotypes of single flies.

Appendix 2.2 Mating scheme used to obtain strains homozygous for the second wild chromosome. I^+ and II^+ represent any two wild chromosome. SM5(Cy) is a multiply inverted second chromosome balancer; Sp, J, L^2 , Pin: second chromosome marked by Sternopleural, Jammed, Lobed and Pinnate.

Generation 1 :	SM5(Cy) ----- Sp J L^2 Pin (tester stock)	X	I^+ ----- II^+ (wild strain)
Generation 2 :	SM5(Cy) ----- Sp J L^2 Pin	X	SM5(Cy) ----- I^+ or II^+
Generation 3 :	SM5(Cy) ----- I or II^+	X	SM5(Cy) ----- I^+ or II^+
Generation 4 :	SELECT:		
	I^+ ----- I^+	or	II^+ ----- II^+ (FLIES ISOGENIC)

APPENDIX 2.3 Uweighted-pair group method phenogram (Sneath and Sokal, 1973) for 50 haplotypes of D.melanogaster, only main branches are shown along with the amylase alozyme allelic designation of all members of a branch. The number of members in each branch is bracketed.



APPENDIX 5.1 Hydrophilic and hydrophobic region of the predicted protein of D.virilis. Hydrophilic residues are above (positive) and hydrophobic residues are below (negative). The actual amino acids are listed below the graph using the 1-letter code.

