POLYMORPHISM IN THE *THR-GLY* REGION OF THE *PERIOD* GENE

LENGTH POLYMORPHISM IN THE *THREONINE-GLYCINE* REPEAT REGION OF THE *PERIOD* GENE IN *DROSOPHILA MELANOGASTER*

By

FAYAZ ALLADINA, B.Sc.

A Thesis

Submitted to the School of Graduate Studies

in Partial Fulfilment of the Requirements

for the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (1993)
(Biology)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: Length Polymorphism in the *Threonine-Glycine* Repeat Region of the *Period* Gene in *Drosophila melanogaster*

AUTHOR: Fayaz Alladina, B.Sc. (McMaster University)

SUPERVISOR: Professor R.S. Singh

NUMBER OF PAGES: x, 146

ABSTRACT

The *period* gene determines biological rhythmicity in *Drosophila melanogaster*. The X-linked gene is 7.4kb, containing 8 exons and 7 introns from which a 4.5kb message is translated. A striking feature of the protein encoded by *per* is a series of alternating threonine-glycine residues in the fifth exon. Moreover, this string of residues is polymorphic for length variation in natural populations, the most frequent variants having 17, 20 or 23 Thr-Gly pairs. In the present study, a geographic analysis of this polymorphism within North American populations was conducted, the results of which indicate significant variation of allele frequency with latitude. The use of spatial autocorrelation analysis and Mantel tests clearly show that the most common variant, encoding 17 Thr-Gly pairs, exhibits a clinal pattern in its distribution along a north-south axis. Furthermore, DNA sequence analysis of several variants has uncovered a novel new variant which encodes 22 Thr-Gly pairs whose nucleotide sequence differs from any published data. Similar statistical analysis conducted on seven allozymes for populations collected along the same transect shows that several have monotonic clinal patterns in their allele frequency distributions which also show correlation with latitude. A previous study of morphological traits on the same populations showed the existence of a non-monotonic clinal pattern. Comparison of the results observed for the molecular and morphological markers indicates that they are subject to different evolutionary forces. The results highlight the importance of comparing patterns of geographic variation using different genetic elements.

ACKNOWLEDGEMENTS

After what seems like an eternity, I cannot express in words the sense of satisfaction I feel in realizing that a personal goal has materialized. Armed with this newly found peace of mind, it is with great pleasure that I acknowledge the contributions of both my friends and my teachers.

I would like to thank Ling Wen Zeng for teaching me, above all, to be tolerant. I only hope that he has gained as much insight into the importance of others opinions as I have through our monthly arguements. Next, apart from fine tuning my culinary talents, Alberto Civetta has taught me the strength of the human spirit. Although I have never told him as much, I will always be in awe of his courage and will. I am (and always will be) indebted to Dr. Michael "Maniatis" Stoehr. Words cannot express my gratitude for the kindness and patience he extolled while guiding me through the intricacies of molecular biology. In the same light, I am indebted to Dr. Monika "Sambrook" Labes. Not once was I turned away when in dire need of assistance. Words cannot do justice for my appreciation of her interest towards my success. I would also like to extend my thanks to Tony Long. Never have I met an individual so dedicated to a discipline and so intent in making others see what he saw. I consider myself privileged to have had such a keen student as a lab mate and friend. To the many new faces in the lab - thanks goes out to Dr. Francisco Molina for reviewing sections of this thesis, statistical assistance and for having enough confidence in my ability to actually let me be the "teacher of the art" for a change. I am indebted to Suman Mukhopadhyay for always making himself available whether it was for technical help or merely to philosophize. I would also like to acknowledge the patience of Todd Prescott in assisting an occasionally computer illiterate Mac user on the finer aspects of programming in Hypercard. The statistics computed in this thesis would not have been possible without his unselfish assistance.

Now, for my teachers. I will always be indebted to Dr. Richard Morton. Dr. Morton has always been honest in his opinions of my work and work ethic. Although at times I may have seemed aloof, he was incessant in his desire to make a competent molecular biologist out of me. I only hope I have not dissappointed him. Thank you Dr. Morton for being my teacher. I would also like to thank Dr. Brad White and Dr. Lisle Gibbs for taking the time to help me troubleshoot. There are parts of the methods section that would not have been possible without their knowledge and generosity. Their willingness to assist me despite schedules of their own was unsurpassed. Finally, I would like to thank my supervisor, Dr. Rama Singh. Dr. Singh continued to support my project even when it seemed as though nothing would come of it. His support, both personal and financial, made my stay at McMaster both comfortable and possible. Thank you Dr. Singh for your time, concern, optimism and patience. Last but by no means least, I would like to thank my family for their love and support. Nothing I have managed to accomplish during my years at McMaster would have been possible without them and it is therefore fitting that I dedicate this thesis to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

1.1 Classification of Genetic Variation Using Visible Marker and Inbreeding Approaches

Description of the genetic variation in natural populations provides the fundamental data for studies of evolutionary genetics. It is not surprising then, that the focus of this discipline has been directed at characterizing the amounts and types of genetic variation existing in natural and laboratory populations of organisms. Prior to the advent of novel molecular techniques, studies of genetic variation consisted of two types of data. The first were data obtained from what can be called the *visible marker gene approach.* This technique studied morphological variation that was a consequence of the segregation of alleles at single loci which could be studied either by direct observation of phenotypes in nature or by test crosses in breeding species (Lewontin 1991). Although this sort of study served as a model system for the study of evolution in action, it required the occurence of large scale phenotypic changes for the detection of variation and furthermore, did not provide a clear picture of the total amount of genetic variation present. Moreover, it was difficult to measure fitness in different genotypes. The second method of measuring genetic variation was based on the *inbreeding approach* and consisted of measuring fitness differences in individuals of different genotypes using entire chromosomes as the unit of observation (Dobzhansky and Spassky, 1953). The approach made use of the well established fact that individuals homozygous or isogenic for one or more chromosomes were generally less viable and less fecund than random heterozygotes and as such, it was possible for the concealed variation present in entire chromosomes to be revealed by

making individual genotypes homozygous. The problem with this method was that this observation could not be made or interpreted at the gene level. Neither method was therefore fully able to satisfy essential criteria for studying variation, namely: 1) to be able to detect phenotypic differences caused by allelic substitutions at single loci in single individuals, 2) to distinguish allelic substitutions at one locus from substitutions at other loci, 3) to distinguish all allelic substitutions from each other and 4) to represent loci that were an unbiased sample of the genome with respect to physiological effects and degree of variation. Thus, the fundamental question regarding the total amount of variation present in a population remained unanswered (Hubby and Lewontin, 1966).

1.2 Protein Electrophoresis and Population Genetics

Investigation of protein variation pioneered by Lewontin and Hubby (1966) and Harris (1966) set the stage for what would be years of intensive study, by thousands, in the field of population genetics using gel electrophoresis (see Lewontin, 1974 and Nevo, 1978 for reviews). The reason behind the explosion in the field of electrophoretic investigation was two-fold. First, the initial studies indicated, at least topically, high levels of naturally occuring variation in natural populations thus supporting the idea of the ubiquity of genetic variation within species. Second, the study of electrophoretic mobility of enzymes and proteins satisfied the four essential requirements mentioned above and thus, phenotypic variation in the form of allozyme mobility, unlike for morphological characters, could be assigned unambiguously to variation at a single locus. Furthermore, because these variants were immune to developmental differences, they could serve as useful markers for research on species comparisons (Kreitman, 1987).

The technique, in its simplest form, seperated protein variants based on their charge and conformational characteristics but was limited to those structural genes with a stainable gene product and hence, limited the number and type of loci that could be studied. In order to resolve variation left undetected using standard gel electrophoresis, modifications to the standard technique such as sequential acrylamide gel electrophoresis (SAGE), heat sensitivity and other criteria useful in distinguishing proteins were developed, increasing detection of variation and overall resolving power (Singh *et al.*, 1976; Singh, 1979). These modifications, however, proved bittersweet in that they improved the resolving power of a technique that was a blessing to experimental population genetics but at the same time, made it clear that protein gel electrophoresis was only capable of resolving a portion of the total variation present at a structural gene locus.

1.3 Molecular Techniques and the DNA Revolution

In recent years, the advent of recombinant DNA technology has changed our ability to characterize genetic variation at the genic level. Because the bands on protein electrophoresis gels revealed phenotypes, and not genotypes, the total extent of genetic variation at structural loci remained unknown (Kreitman, 1983). Moreover, it was soon realized that the static allelic frequency data obtained from electrophoretic studies were not capable of distinguishing between the relative roles of natural selection, genetic drift, migration, mutation and recombination in generating and maintaining genetic variation within and between populations (Ewens 1977).

1.3.1 The Use of Restriction Enzyme Surveys: The Past

The ability to uncover variation at the DNA level has radically altered this situation, particularly through the use of restriction enzymes. Six base recognizing restriction endonucleases have been widely used to address population genetic questions by providing a coarse screen of nucleotide variability (see Kreitman, 1987; Leigh-Brown, 1989 for reviews). But since these enzymes theoretically cut DNA once every 4096 bp, even a battery of enzymes would not sufficiently uncover variation at a level of significance. Ten enzymes optimally chosen would sample every 400 bp of a structural gene locus. A modification of this technique employs four-cutter restriction endonucleases providing higher level resolution of a surveyed gene region (Kreitman and Aguadé, 1986a). Because these enzymes cut once every 256 bp, using 10 optimally selected enzymes would facilitate the surveying of approximately every 30 bp in a concensus sequence of a genome. Because of polymorphism, even more sites would be effectively surveyed as sequences not recognized in the consensus would be recognized in one or more members of a population of sequences. Since each recognized sequence involves 4 bp, under optimal conditions one could screen 1 in every 7 bases (Riley *et al.*, 1989). Such a technique cannot address the same questions as complete sequence information because it is ineffective at distinguishing between silent and replacement polymorphisms within coding regions unless the sequence of the gene being surveyed is available. In this case, one could ellucidate between changes occuring in coding and non-coding regions. Kreitman and Aguadé, (1986a) were the first to employ this technique in their study of haplotype variation in two populations of *D. melanogaster* at the *Adh* locus. They estimated that using this technique, approximately 25% of all single nucleotide changes within the region of interest were detected including all insertion/deletion variation and even single nucleotide length differences. Thus, high

resolution four-cutter analysis can, and has been able to provide insight into the level and pattern of nucleotide variation in nature (Kreitman and Aguadé, 1986a; Miyashita and Langley, 1988; Riley *et al.*, 1989).

1.3.2 The Use of DNA Sequencing: The Present

However useful the four-cutter technique may be, it should be realized that uncovering the absolute amount of variation present at a structural locus requires nothing short of the complete nucleotide sequence. Although the use of restriction enzymes for the study of DNA polymorphism represented an important advancement over studies using allozyme markers in population analysis, these studies were still time consuming and required large amounts of starting material. Alternately, DNA sequencing allows for allelic differences between individual genes to be identified directly at the nucleotide level. The complete resolution afforded by DNA sequencing enables one to detect all nucleotide substitutions including those that are not translated into protein differences. Sequence variation in non-coding regions such as introns or between synonymous codons, because they are free from selective forces acting through the phenotype of the encoded protein, are the best available source for estimating the frequency of polymorphism due to neutral gene substitution (Kreitman, 1983). This is crucial for understanding the role of natural selection and random genetic drift in the maintenance of genetic variation. But, traditional DNA sequencing also has its shortcomings - it is too time consuming, requiring large scale cloning strategies and is relatively expensive a task.

1.3.3 PCR Based Sequencing: The Future

The Polymerase Chain Reaction (Saiki *et al.*, 1985), an *in vitro* method for the enzymatic amplification of specific DNA segments from a complex genome has greatly simplified the study of sequence variation and has, as a result, transformed the way in which genetic analysis can be carried out (see Erlich and Arnheim, 1992 for a review). PCR amplification coupled with direct cycle sequencing will undoubtedly prove to be invaluable to molecular population genetics by revealing the absolute amount of variation in the form of detailed sequence for any number of alleles of a locus. Because the need for time consuming cloning associated with conventional sequencing is eliminated, large data sets can be accumulated in a time efficient manner. By virtue of the resolution afforded, this should enable more rigorous testing of hypotheses regarding the role of evolutionary forces in maintaining variation at the nucleotide level.

1.4 *Drosophila* and Restriction Studies

The ability to estimate levels of DNA sequence variation in various regions of the genome has opened up new doors for studying the origin and significance of genetic variation in natural populations. *Drosophila melanogaster* has been the choice organism for studies of this sort because of its short generation time, the ease by which it can be maintained in the laboratory and because its genetics is well understood. Molecular surveys of regions of genomic DNA in *Drosophila* have revealed that rates of recombination between adjacent nucleotides is small and consequently, that adjacent nucleotides share a common evolutionary history. This fact has facilitated hypotheses testing regarding the significance of observed levels of variation (Aquadro, 1992). Interspecific restriction studies in particular, have and still are leading to significant

progress in the resolution of the relative contribution of selection, drift, mutation and other factors in the maintenance of genetic variation (Kreitman, 1991).

1.5 Genomic Regions Surveyed in *Drosophila*

Numerous restriction studies have been conducted on regions of the *Drosophila* genome spanning virtually every chromosome to different degrees. Most studies have used intraspecific restriction analysis and have revealed variable levels of insertion/deletion variation possibly due to transposable elements (Aquadro *et al.*, 1986). They have also revealed extensive variation due to base pair substitution and data available for 0.5% of the euchromatic portion of the *D. melanogaster* genome indicates that nucleotide diversity (heterozygosity) averages 0.004 (Begun and Aquadro, 1992). Furthermore, this nucleotide diversity varies in different parts of the genome as is evident from table 1.5, and even within parts of the same gene (Kreitman, 1983).

TABLE 1.5

Nucleotide Diversity in *D. melanogaster*

Gene Region	Region Size(kb)	Nucleotide Diversity
Chromosome I (X)		
yellow-achaete	35	0.001
Phosphogluconate Dehydrogenase	16	0.003
zeste-tzo	20	0.004
period	30	0.001
white	45	0.009
Notch	60	0.005
vermillion	20	0.001
forked	25	0.002
G-6-PD	13	0.001
suppressor of forked	24	0.000
Chromosome II		
Gpdh	23	0.008
Adh	13	0.006
Dopa decarboxylase	65	0.005
Amylase	15	0.008
Punch	14	0.004
Chromosome III		
Esterase-6	22	0.005
Metallothionein-A	2	0.001
Hsp 70A	25	0.002
rosy	40	0.003
Chromosome IV		
cubitus interruptus	1	0.000

From Begun and Aquadro (1992)

1.6 The Causes of Genetic Variation

Natural selection and mutation-drift have been the two most important evolutionary forces posited to be accountable for observed levels of sequence variability. Because random genetic drift is a function of population size and is due to chance sampling variation from generation to generation, it has predictable consequences for genetic variation and has thus served as the null hypothesis for testing molecular data. The basic tenet of the neutral theory of molecular evolution is that genetic drift is the predominant driving force of molecular evolution. According to the theory, polymorphism is a transient phase of evolution and the observed divergence between two randomly chosen genes between two species is a function of species divergence time and mutation rate (Kimura, 1983). Consequently, a gene region with high mutation rate or low functional constraint is expected to be highly variable within a species and therefore, accumulate rather rapidly, fixed differences between species. More concisely then, variation within a species is expected to be correlated positively to divergence between species (Hudson *et al.*, 1987). It is this arguement which has been used to either support or reject the null hypothesis mentioned earlier by authors who have conducted interspecies restriction surveys of gene regions (Kreitman and Aguadé, 1986b; Schaeffer and Aquadro, 1987; Kreitman and Hudson, 1991). Selection, as mentioned, is the alternative explanation for observed levels of genetic variation. Selection at one site, unlike drift, does not affect the long term rate of accumulation of sequence divergence between species at linked neutral sites, but rather, only the specific site under selection (Hudson, 1990). Selection does, however, affect levels of intra-population variation at both selected and linked neutral sites by what is known as the "hitchhiking effect" (Maynard Smith and Haigh, 1974; Hudson, 1990). The

degree to which this occurs is dependent on the strength of selection and rate of recombination in the region under consideration. Balancing selection leads to an accumulation of variants at linked neutral sites and thus, an *excess* of variation *around* the selected site (Strobeck, 1983; Hudson and Kaplan, 1988). Alternately, directional selection results in a *deficiency* of variation *around* the selected site by sweeping a favoured variant through a population to fixation, eliminating much of the variation at linked neutral sites with levels of recombination determining the size of the region being affected (Kaplan *et al.*, 1989; Aquadro, 1992, Langley *et al.*, 1993).

1.7 Clinal Variation and Natural Selection

Clinal variation is described as the monotonic change in some morphological character such as colour or height along a geographical gradient such as altitude or latitude that is ubiquitous in natural populations. There are several textbook examples of clinal variation occuring in nature including height variation in *Potentilla glandulosa* (Claussen *et al.*, 1984) and genetic variation in the moth *Amathes glareosa* in its melanic form due to a single locus mutation (Kettlewell and Berry, 1961). Allozyme data indicate that allele frequencies of many protein coding loci vary clinally with examples including the *lactate dehydrogenase* locus in *Fundulus heteroclitus* (Powers *et al.*, 1979), the *Lap* locus in *Mytilus edulis* (Koehn *et al.*, 1980) and *glycerol-3-phosphate deydrogenase* and *Alcohol dehydrogenase* in *D. melanogaster* (Oakeshott *et al.*, 1982). The existence of clinal variation has been considered indicative of the action of natural selection since purely historical explanations seem inadequate to account for the large spread occurence of clines, particularly for those repeated across different continental regimes (Koehn *et al.*, 1983; Anderson and Oakeshott, 1984; Lachaise, 1988; Singh and Long, 1992). In some cases,

the evidence for selection has been reinforced by laboratory studies which indicate that there are biochemical differences between the protein products of different alleles at a locus (Powers *et al.*, 1979; DiMichele *et al.*, 1991). In other organisms, it has been possible to set up selection cage experiments and mimic "natural selection". These experiments have shown that selection can act to alter the allele frequency in a direction predicted by that of the naturally occuring cline (Cavener and Clegg, 1981). This method though, is problematic in that it is impossible to determine if the selection pressure applied in the artificial environment is comparable or representative of the selection pressure occuring in nature.

1.7.1 The *Adh* Locus of *Drosophila*: An Example of Clinal Variation

The *Adh* locus of *D.melanogaster* is a well documented example of clinally varying electrophoretic variation (David, 1982; Singh *et al.*, 1982). The frequencies of the two most common electromorphs, *F* and *S*, varies with latitude on three seperate continents - Eastern United States, Asia and Australia, the *S* allele being most common in Southern latitudes. Furthermore, biochemical studies indicate that the product of the *S* allele may be more thermostable than that of the *F* allele, favouring a selective explanantion. That is, the *S* allele may be favoured in hotter, more southern climates by natural selection (Van Delden, 1982). Despite this, the physiological significance of the *F/S* polymorphism remains largely unknown. In their study of clinal variation on three continents, Oakeshott *et al.*, (1982) attempted to find correlations between environmental parameters along the cline and the frequency of *S*. Their results were unequivocal in that the frequency of the *S* allele was found to vary only with one environmental factor - rainfall! Moreover, there are parts of the world where the cline does not occur at all. Yet despite these observations

which seemingly preclude selection, the strongest argument for the role of selection in maintaining the *Adh* cline is the consistency with which the cline is maintained over three continental regimes (Anderson and Oakeshott, 1984).

1.8 The *Period* Gene of *Drosophila*

The *period* gene of *Drosophila* regulates both circadian rhythms, those having a 24 hour cycle such as adult/pupal eclosion profiles and locomoter activity, and ultradian rhythms, those less than 24 hours such as male courtship song cycles (Yu *et al.*, 1987). Much is known concerning the ethological and developmental process regulated by this gene. Furthermore, the molecular genetics of *per* and the mechanisms of regulation and expression of its gene products are well understood. The *per* gene produces three separate and distinct messages and it is phenotypically variable due to the existence of three naturally occurring mutants which differ in their rhythmicity (see Hall and Kyriacou, 1990 and Kyriacou, 1990 for a complete review on the *period* locus). Yet, very little work has been done pertaining to the variability of this gene in natural populations. The few studies to date have primarily focussed on comparisons of sequence variability between species and levels of synonymous and nonsynonymous substitution in *per* relative to other characterized and variable *Drosophila* genes (Colot *et al.*, 1988; Thackeray and Kyriacou, 1990; Begun and Aquadro, 1991 ; Kliman and Hey, 1993; Peixoto *et al.*, 1993). Of late, several studies on intraspecific variability have been conducted, specifically surveys pertaining to the variability in the Threonine-Glycine repeat region of exon five (Yu *et al.*, 1987; Costa *et al.*, 1991; Costa *et al.*, 1992).

The most striking feature of the primary structure of the *per* gene product in *D. melanogaster* is a run of alternating Threonine-Glycine pairs (Jackson *et al.*, 1986; Citri *et*

al., 1987). It is the region surrounding and including this repeat that has been implicated in determining the species specific differences in the *Drosophila* lovesong cycle (Yu *et al.*, 1987; Wheeler *et al.*, 1991). Moreover, this region has also been implicated in the thermostability of circadian behaviour (Ewer *et al.*, 1990). The minisatellite DNA which encodes the Thr-Gly repeat is polymorphic in length in *Drosophila melanogaster*, in both laboratory strains (Yu *et al.*, 1987) and in natural populations (Costa *et al.*, 1991; Costa *et al.*, 1992). The most common variants encode 17, 20 or 23 Thr-Gly pairs although rare variants, such as those encoding 15, 18 and 21 Thr-Gly pairs exist. The work of Costa *et al.*, (1992) has shown that within European populations of *D. melanogaster*, populations from the extreme south show higher frequencies of the shorter Thr-Gly variant (TG17) as compared to more northern populations. Spatial analysis conducted on their data set concluded that the Thr-Gly 17 variant does in fact show a clinal pattern in its distribution in European populations and that the frequency of this allele significantly correlates with geographic distance as well as latitudinal distance. They suggest the possibility that the length polymorphism might be maintained as a consequence of thermal selection based on the findings of Ewer *et al.*, (1990) which showed that sequences surrounding and including the Threonine-Glycine repeat region are involved in providing thermostability to the circadian phenotype.

1.9 Objectives of the Present Study

Interest in geographical patterns of genetic variation derives largely from the hope that the patterns observed might provide an understanding of the population genetic processes that have given rise to them (Wright, 1978). For the most part, such hopes have remained unfulfilled as it is generally the case that such patterns have arisen in more than

one way and due to more than one factor. To further compound the problem, there is often insufficient knowledge about the past history of the populations under investigation to distinguish between alternatives. Lastly, at least until recently that is, formal procedures for the analysis of geographical variation have been lacking (Epperson, 1993). Relatively recent work by Robert Sokal and his coworkers pertaining to spatial autocorrelation analysis in biological systems has helped clarify this last problem (Sokal and Oden, 1978; Sokal, 1979; Sokal and Wartenberg, 1983). These authors have put this method forward in a general framework for the evaluation of geographical patterns of biological variation and have applied them to a variety of situations.

Of late there have been increasing numbers of studies on spatial structure that generally indicate strong spatially autocorrelated distributions of genetic variation (Sokal and Oden, 1978; Sokal and Menozzi, 1982; Esteal, 1985; Sokal *et al.*, 1987; Baker, 1992; Costa *et al.*, 1992). Natural selection can create sharp spatial patterns, and many studies have therefore sought also to detect evidence of natural selection in spatial patterns of gene frequencies among subpopulations. That is, spatial autocorrelation has been used to support the existence of underlying patterns in variation, clinal or other, thought to be present in the data. Keeping this in mind, it was decided that we would study the distribution and pattern of the Threonine-Glycine repeat in latitudinal populations of *D. melanogaster*. This was decided for several reasons. First, the *Drosophila* populations for this study had been previously used for allozyme analysis for which data were available. It would be then possible to compare the results of the pattern of protein variation along the gradient to that from a nuclear gene. The flies had also been used for a study of morphological variation in 10 characters in which a non-monotonic clinal pattern had been documented (Long, 1992). Secondly, considering the work of Costa *et al.*, (1992) who

detected a clinal pattern in the distribution of the Thr-Gly17 variant in European populations, we felt it would be interesting to see whether the trend could be repeated in North American populations. Because no significantly selective arguement could be posited for the existence of the pattern, observation of a similar pattern on a different continent might provide credence for the action of selection at this locus. Thirdly, because the analytical tools required to assess the significance of our data were available, namely the spatial autocorrelation theory, we felt that a stringent analysis could be performed on both the allozyme and nuclear data obtained. This would enable us to support or reject the existence of a clinal pattern for the Thr-Gly region as per Costa *et al.*, (1992) and compare the monotonic or non-monotonic pattern of clinal variation in molecules and morphology (Long, 1992).

Chapter 2

MATERIALS

2.1 Strains of Drosophila Used

Five populations of *D. melanogaster* were used for this study (Ottawa, ON., Windsor, ON., Louisville, KY., Cartersville, GA. and Tampa Bay, FLA.), comprising 126 lines and covering 28 to 45 degrees latitude along a North/South latitudinal gradient. The Ottawa lines were collected during the Autumn of 1992, by Dr. Donal Hickey of the University of Ottawa. The Windsor, Louisville and Cartersville lines were collected as part of a larger population set by Tony Long in the Summer of 1988. The Tampa Bay lines were provided by Dr. Bruce Cochrane of the University of Southern Florida and were collected during the Spring of 1992 (Figure 2.1).

2.1.1 Maintenance of Isofemale Lines

After establishment, all isofemale lines were maintained in 100 mL vials containing approximately 10 ml of standard banana medium at 20 °C on a 12 hour day/night cycle.

2.1.2 Establishing Isogenic Stocks

For use in the present study, each of the 126 lines was crossed to a compound X stock (C(1)DX yf/fB^3), courtesy of the Bowling Green State Drosophila Stock Centre, in order to standardize the X chromosome (Appendix Figure A1). The compound chromosome, in this case, had two markers, yellow(*y*) and forked(*f*) to facilitate sexing. Individual males from the isofemale lines were crossed via single-pair matings to virgin females of the compound X stock and reared on banana medium at 20 °C and a 12:12 L:D cycle. Males

Figure 2.1: Map of geographic locations from which *Drosophila melanogaster* lines were collected. Locations designated (°) represent populations used for the Southern analysis of *period*. Locations designated (•) represent populations for which the allozyme analysis was conducted and locations designated (†) represent populations for which the analysis of morphological variation was conducted.


The Pas
Winnipeg
Ottawa
Windsor
Dayton
Louisville
Nashville
Cartersville
Cordele
High Springs
Tampa Bay
400 MILES

from the F1 and all proceedeing generations, harbouring the isolated X chromosome, were collected for nucleic acid extraction.

2.2 Biochemicals

All Biochemicals were of the highest quality available and were obtained from major supply houses. Lysozyme, salmon sperm DNA, proteinase K, DNAase-free RNAase, restriction endonucleases, T4 DNA Ligase and ampicillin were purchased from Boehringer Mannheim Canada. Bacto-Tryptone, Bacto-Yeast and Agar were purchased from DIFCO (Detroit, Michigan). Spermine, spermidine and dextran sulfate were purchased from Sigma Chemical (St. Louis, Missouri).

2.3 Reagents

All reagents were purchased from Sigma Chemical, Fisher Scientific, BDH, BioRad or Gibco BRL. Nylon membrane was purchased from Millipore (Immobilon N) and BioRad (Zetaprobe GT+). *All solutions cited in the Methods were prepared according to Sambrook et al., (1989). Where a modification to the recipe was made or in cases where the recipe was obtained from a different source are listed below.*

Drosophila Culture Medium

Stock cultures were maintained on standard banana medium.

Agar	20g
Small bananas	2
Corn syrup	2 tbsp.
Sugar	1.5 tbsp.
Yeast	60g

Boil agar and water. Mix remaining ingredients and add to boiling waster/agar mixture. Continue to boil for 10 minutes and then allow to cool. When media is cooled to 40 °C, add

36mL tegosept (10g Methyl p-hydroxy benzoate
in 100mL 95% ethanol). Pour.

Luria Broth

Bacto-tryptone	10g
Bacto-yeast	5g
NaCl	5g
ddH_20	1L
1N NaOH	4mL

The dry ingredients were first dissolved in distilled water.
The broth was then autoclaved and if antibiotics were required,
added once the media had cooled to below 50 °C,
as antibiotics such as those used are heat labile.
Ampicillin was commonly used and was added
at a concentration of 50-100ug/mL.

Luria Broth Agar

Bacto-tryptone	10g
Bacto-yeast	5g
NaCl	5g
ddH_20	1L
1N NaOH	1mL
Agar	15g

Following the method above with the exception that
solid agar and less NaOH was added prior to autoclaving.
Antibiotics were added after allowing LB Agar
to cool to below 50 °C

STE Buffer

Nacl	0.1M
Tris-HCl (pH 8.0)	10mM
Na_2EDTA (pH 8.0)	1mM

STET Buffer

Nacl	0.1M
Tris-HCl (pH 8.0)	10mM
Na_2EDTA (pH 8.0)	1mM
Triton-X	5%

10X Ligation Buffer

Tris-HCl (pH 7.5)	200mM
$MgCl_2$	50mM
Dithiothreitol	50mM
Bovine serum albumin	50ug/mL
ATP	0.5mM

Stored in 0.5mL aliquots at -20 °C.
ATP added at the time of reaction.

Homogenization Buffer

Tris-HCl (pH 8.0)	10mM
NaCl	60mM
Na_2EDTA (pH 8.0)	10mM
Spermine	0.15mM
Spermidine	0.15mM
ddH_20	19.5mL

Stored at 4 °C for up to 4 weeks.

Nuclear Isolation Buffer (NIB)

Tris-HCl (pH 7.5)	10mM
NaCl	60mM
Na_2EDTA (pH 8.0)	10mM
Spermine	150mM
Spermidine	150mM
Triton-X	4.62mL

ddH20	to 1L

stored at 4 °C

10X TAE Agarose Gel Running Buffer (per L)

Tris base	48.4g
Glacial Acetic Acid	11.42mL
0.5M Na_2EDTA (pH 8.0)	20mL

Adjust the pH of buffer to 7.8 with glacial acetic acid, used at 1X for electrophoresis.

6X Agarose Gel Loading Buffer

Bromophenol blue	0.25%
Sucrose in water	40%(w/v)

Stored in 1mL aliquots at 4 °C.

5X TBE Acrylamide Gel Running Buffer (per L)

Tris base	54g
Boric acid	27.5g
0.5M Na_2EDTA (pH 8.0)	20mL

pH of the buffer should be 8.3, used at 1X in PAGE.

30% Acrylamide Stock Solution (per 100mL)

Acrylamide	29g
Bisacrylamide	1g

Solution was stirred until fully dissolved and then filtered through Whatman paper under a vaccuum. The solution was then stored in dark bottles at

room temperature for up to 4 weeks.

100X TNE_2 Stock Solution

Tris base	12.11g
NaCl	5.85g
Na_2EDTA	7.45g

The mixture was taken to 50mL with ddH_20, the pH adjusted to 8.0 with HCl to dissolve and the final volume adjusted to 100mL.

TNE_2 Saturated Sepharose CL6B

Sepharose CL6B	30mL

Add ddH_20 to Sepharose to 50mL in a centrifuge tube in order to make a 60% slurry. Rinse ten times with 1X TNE_2 by spinning full speed on a Dynac tabletop centrifuge for 5 minutes per spin; resuspending as a 60% slurry in TNE_2 each time.

Rinsed Glass Beads

500um glass beads	20mL
HCl	0.1N
NaOH	0.1N

Add beads to 50mL centrifuge tube and fill with 0.1N HCl. Shake and then rinse once with ddH_20, once with 0.1N NaOH, 3 times with ddH_20 and twice with 1X TNE_2.

Salmon Sperm DNA

Salmon sperm DNA	50mg/mL(w/v)

Dissolved in ddH_20 using a stirrer for 1hour,
passed through an 18g needle 20 times in order to shear,
extracted with phenol:chloroform and then
chloroform:isoamyl alcohol
followed by ethanol precipitation and finally
resuspending again in ddH_20. Stored in 0.5mL
aliqouts at -20 °C. Denatured by boiling
for 10 minutes prior to use.

Prehybridization Solution (per 10mL)

---

ddH_20	7mL
50% Dextran sulfate	2mL
10% SDS	1mL
NaCl	0.58g

Heated to 63 °C to facilitate dissolution.

Hybridization Solution (per 10mL)

---

ddH_20	7mL
50% Dextran sulfate	2mL
10% SDS	1mL
NaCl	0.58g
Salmon sperm DNA	15uL

Salmon sperm was denatured prior to addition.
Solution warmed to 63 °C prior to addition

Chapter 2
METHODS

2.4 High Molecular Weight Nucleic Acid Purification

Various methods for extraction of nucleic acids were used of which two methods yeilded the best quality nucleic acid generally free of RNA contamination.

2.4.1 Extraction Using SDS Lysis

The first method was a modification of that of Ashburner (1989). Approximately 200 flies were placed in 2.5 mL homogenization buffer and ground in a glass mortar using a pestle fixed to a drill press. The pestle was passed through 10 - 12 times until the presence of red eye pigment was no longer visible. The supernatant was then centrifuged at 8000 r.p.m. for 5 minutes and the resulting pellet resuspended in 500 uL homogenization buffer. 50 uL 10% SDS and proteinase K at a final concentration of 100 ug/mL was added and the entire mixture was mixed thoroughly yet gently prior to incubation at 37 °C for 45 minutes in a heated water bath. This was then extracted twice with a mixture of phenol:chloroform and twice with chloroform:isoamyl alcohol. The nucleic acids were then precipitated with 100% ethanol in the presence of 5M NaCl and the pellet, which appeared as a very viscous mass at the interface, was pulled out with a glass capillary tube. Following a 30 minute 70% ethanol wash, the pellet was dried and then resuspended in 200 uL $T_{10}E_1$(pH 7.5), RNased (100ug/mL) at 37 °C for 25 minutes and reprecipitated with ethanol.

2.4.2 Extraction From CsCl Gradients

The second method was a modification of Bingham *et al.*, (1981). Approximately 500 flies were added to 5 mL Nuclear Isolation Buffer (NIB) and homogenized as above using a pestle attached to a drill press. The slurry was centrifuged for 1 minute at room temperature at 4000 r.p.m. in a Dynac tabletop centrifuge to remove debris and the supernatant was collected and recentrifuged at 7000 r.p.m. at 0 °C for 7.5 minutes using an SS34 Sorvall rotor. The resulting pellet was collected and resuspended in 1 mL fresh NIB and recentrifuged as above. This pellet was then resuspended in 1.9 mL NIB and 480 uL of 10% sarkosyl was added and the slurry mixed gently by rocking. To this, 3.69 g of solid cesium chloride was added and mixed gently until disolved. A mixture of NIB/2% sarcosyl was added to the slurry to bring the volume to 4 mL and this was then added to a 3.9 mL Beckman Quikseal pollyallomer tube using a pasteur pipette. The tubes were then sealed using a quikseal tube heat sealer. The samples were then spun at 55 000 r.p.m. for 24-36 hours at 15 °C in a fixed angle vTi.65.2 rotor. Following spinning, the tubes were removed, taking care to avoid agitation, and the gradients were dripped using a 25 gauge needle for a vent hole near the top and a drain hole 3 mm from the bottom (Appendix Figure A2). The DNA was collected using this *side puncture* method and was dialysed exhaustively against sterile $T_{10}E_1$ (pH 7.5) for 3 changes of 1L each at 4 °C. The obvious viscosity of the DNA fraction relative to the other fractions of the drip, negated the use of ethidium bromide in the gradient as a means of facilitating collection.

2.5 Analysis of Nucleic Acids Following Extraction

2.5.1 Quantification Against Standard

For both methods, agarose gel electrophoresis against a quantified standard was performed on the native samples obtained in order to obtain a crude estimate of both yield and purity. For more accurate estimates of yield and purity, spectrophotometry was performed as outlined in Sambrook *et al.*, (1989).

2.5.2 Spectrophotometric Quantification of Nucleic Acids

DNA was quantified by examining UV absorbance at 260 nm and 280 nm on a Perkin Elmer Lambda 3 UV/VIS Spectrophotometer. The absorbance at 260 nm was used to calculate nucleic acid concentrations under the assumption that 1 OD unit is equivalent to 50 ug/mL dsDNA. The ratio of absorbances at 260 and 280 nm was used to assess the purity of the samples. Pure DNA preparations result in OD_{260}:$OD_{280} \geq 1.8$.

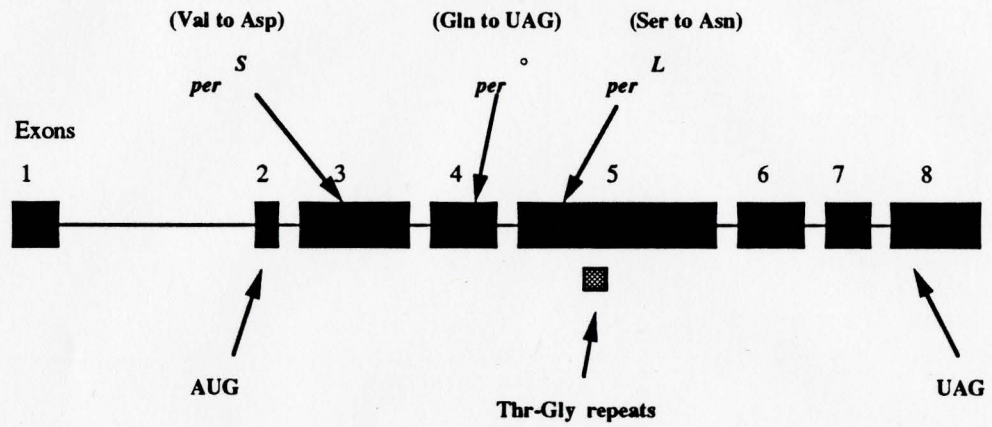
2.6 Preparation of Clones

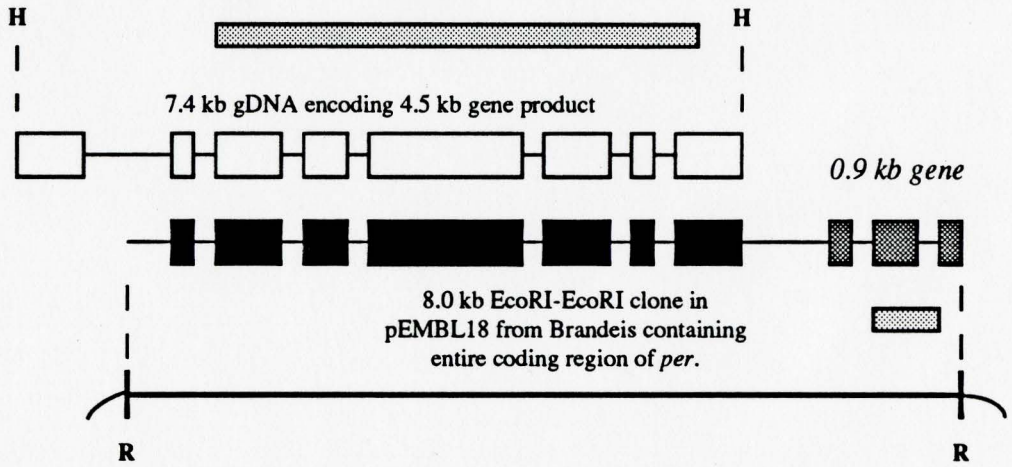
The clone required for this study was obtained from Michael Rosbash of Brandeis University in January 1991. An 8.0 k.b. *EcoRI-EcoRI* fragment containing the *period* gene cloned into the *pEMBL18* phasmid was transformed and used to obtain all other subclones for this study (Figure 2.6).

2.6.1 Preparation of Competent Cells

Transformation of competent *E. coli* was performed as in Chung *et al.*, (1989). *E. coli* DH5a were cultured until early exponential phase in 5 mL of LB broth or until an OD_{600} of approximately 0.4, transferred to 15 mL polypropylene tubes, pelleted by centrifugation at

Figure 2.6: Schematic representation of the intron/exon structure of the *period* gene along with the *EcoRI-EcoRI* clone received from Brandeis University containing the entire coding region of *per*. All subclones were constructed from this 8.0kb clone.


(Val to Asp)
(Gln to UAG)
(Ser to Asn)
S
per
per
per
L
Exons
1
2
3
4
5
6
7
8
AUG
Thr-Gly repeats
UAG


H
H
7.4 kb gDNA encoding 4.5 kb gene product
0.9 kb gene
8.0 kb EcoRI-EcoRI clone in
pEMBL18 from Brandeis containing
entire coding region of per.
R
R

1000g for 10 minutes at 4 °C and then resuspended at one-tenth their original volume in ice cold Transformation and Storage Solution (TSS).

2.6.2 Genetic Transformation of Competent Cells

A 100 uL aliquot of competent cells was mixed with 1 uL of the complete *per/pEMBL* construct and incubated at 4 °C for 30 minutes. 900 uL of TSS augmented with 20 mM glucose was then added and the cells were grown at 37 °C at 225 r.p.m. in a shaker for 60 minutes in order to allow expression of the antibiotic-resistance gene. Transformants were selected by plating cells in triplicate on LB agar plates containing 50 ug/mL ampicillin using the aseptic technique. Plates were incubated for 16-24 hours at 37 °C.

2.6.3 Transformant Screening

To demonstrate that transformation of bacterial cells was mediated by the input *per/pEMBL* DNA, transformants were screened for the presence of the construct DNA via alkaline lysis of minipreps prior to restriction and gel electrophoresis. Unique single colonies (*E. coli* harbouring the *per/pEMBL* construct) were used to innoculate 5 mL of LB broth containing 50 ug/mL ampicillin. The innoculum was incubated overnight at 37 °C in a rotary-wheel incubator. 1.5 mL of each culture was then transferred to a microfuge tube and the cultures were centrifuged for 20 seconds in order to pellet the cells. The supernatant was discarded and the pellets were resuspended in 350 uL of STET by vortexing. 10 uL of freshly prepared lysozyme (30 mg/mL) was added to each tube prior to boiling in water bath for exactly 3 minutes. The tubes were then centrifuged at 12 000g for 15 minutes at room temperature and the pellets removed and discarded with sterile toothpicks. To the resulting supernatant, 330 uL of isopropanol was added and the mixture

was centrifuged for 10 minutes at 12 000g at 4 °C. The new pellet was washed with 70% ethanol, dried *in vacuuo* and resuspended in 100 uL $T_{10}E_1$ (pH 8.0) buffer at 65 °C in order to completely dissolve. 10 uL of each miniprep was then restricted, according to manufacturers instructions, using the appropriate endonuclease and screened for the presence of plasmid and insert DNA via agarose gel electrophoresis.

2.6.4 Permanent Storage of Transformants

Following gel electrophoresis, a single miniprep was chosen to be converted into a permanent stock. Approximately 500 uL of the miniprep was added to 500 uL of a mixture of LB broth containing 14% DMSO (freezing solution) and the mixture stored at -70 °C in a cryovial.

2.6.5 Large Scale Preparation of Plasmid DNA

In order to obtain a greater yield and purity of the plasmid/insert construct, large scale plasmid preps were carried out essentially as outlined in Sambrook *et al.*, (1989). A sterile wooden stick was used to scrape the permanent stock which was then used to streak an LB-amp plate. After incubation at 37 °C for 24 hours, a single colony was picked and used to innoculate in duplicate, 10 mL of LB broth containing 50 ug/mL of ampicillin. The culture was then incubated overnight at 37 °C in a rotary wheel incubator. The following day, the 10 mL cultures were added to two 1L flasks containing 250 mL of LB with ampicillin and incubated for a further 24 hours at 37 °C in a shaking incuator set at 300 cycles per minute.

2.6.6 Harvesting of Bacteria

The bacterial cultures were transferred to sterile 250 mL reinforced plastic bottles and harvested by centrifugation at 4000 r.p.m. for 15 minutes at 4 °C in a Sorvall GSA rotor. The supernatant was removed and the bottles were inverted and let to dry on the benchtop. The pellets were then resuspended in ice cold STE by vortexing and recentrifuged as above in order to collect the bacterial cells.

2.6.7 Lysis of Bacteria

The pellets obtained from the harvest were then resuspended in 10 mL of TEG (solution I in Sambrook *et al.*, 1989). 1 mL of freshly prepared lysozyme (10mg/mL) was then added followed by 20 mL of solution II. The contents of the plastic bottle were then mixed thoroughly by inversion and the bottle was stored at room temperature for 10 minutes. 15 mL of ice cold solution III was added and the contents were mixed by shaking vigourously several times. The bottles were then stored on ice for 10 minutes during which time a white precipitate, containing chromosomal DNA and high molecular weight RNA, formed. The plastic bottles containing the bacterial lysate were centrifuged at 5000 r.p.m. for 15 minutes at 4 °C and the rotor allowed to stop without braking in order not to disturb the packed pellet. The supernatants were transferred to new sterile plastic bottles, filtering through four layers of cheesecloth and taking care not to disrupt the pellet. 0.6 volume of isopropanol was added to the supernatant containing the plasmd DNA and mixed thoroughly prior to storing the bottles at room temperature for 10 minutes. The nucleic acids were then recovered by centrifugation at 5000 r.p.m. for 15 minutes at room temperature in a Sorvall GSA rotor. The supernatant was removed, the pellet was washed in 70% ethanol, benchtop dried and dissolved in 1.5 mL of $T_{10}E_1$ (pH 8.0).

2.6.8 Purification of Plasmid DNA Using PEG

The plasmid DNA was purified by precipitation with polyethylene glycol. The 1.5 mL nucleic acid solution was transferred to a sterile 15 mL polypropylene tube and 1.5 mL of ice cold 5M lithium chloride was added in order to precipitate high molecular weight RNA. The solution was mixed well and centrifuged at 10 000 r.p.m. for 10 minutes in a Sorvall SS34 rotor. The supernatant was transferred to a new 15 mL tube and an equal volume of isopropanol was added followed by centrifugation at 10 000 r.p.m. for 10 minutes at room temperature. The resulting pellet was washed with 70% ethanol, dissolved in 250 uL of $T_{10}E_1$ (pH8.0) containing DNAase-free RNAase (20 ug/mL), then transferred to an Eppendorf microfuge tube and stored at room temperature for 30 minutes. 250 uL of 1.6 M sodium chloride containing 13% weight/volume of polyethylene glycol (PEG 8000) was added and the plasmid DNA was recovered by centrifugation at 12 000g for 5 minutes at 4°C in a microfuge. The pellet was then dissolved in 200 uL of $T_{10}E_1$ (pH8.0) and extracted once with phenol:chloroform and then with chloroform:isoamyl alcohol. 50 uL of 10 M ammonium acetate and two volumes of 100% ethanol were added to the resulting aqueous phase in order to precipitate the plasmid DNA, which was recovered by centrifugation at 12 000g for 5 minutes at 4 °C in a tabletop microfuge. The pellet was then washed, dried *in vacuuo* and resuspended in 250 uL of $T_{10}E_1$ (pH 8.0). The concentration of the plasmid DNA in solution was determined via spectrophotometry at OD_{260}.

2.7 Preparation of Subclones

Various subclones were constructed from the original 8.0 kb *per* gene (p8.0), including a 5.7 kb fragment containing the entire coding region of *per* cloned into *pEMBL18* (p5.7)

as well as a 1.3 kb fragment corresponding to the fifth exon of the gene (containing the variable Threonine-Glycine repeat region) cloned into *pUC19* (p1.3) (Figure 2.7).

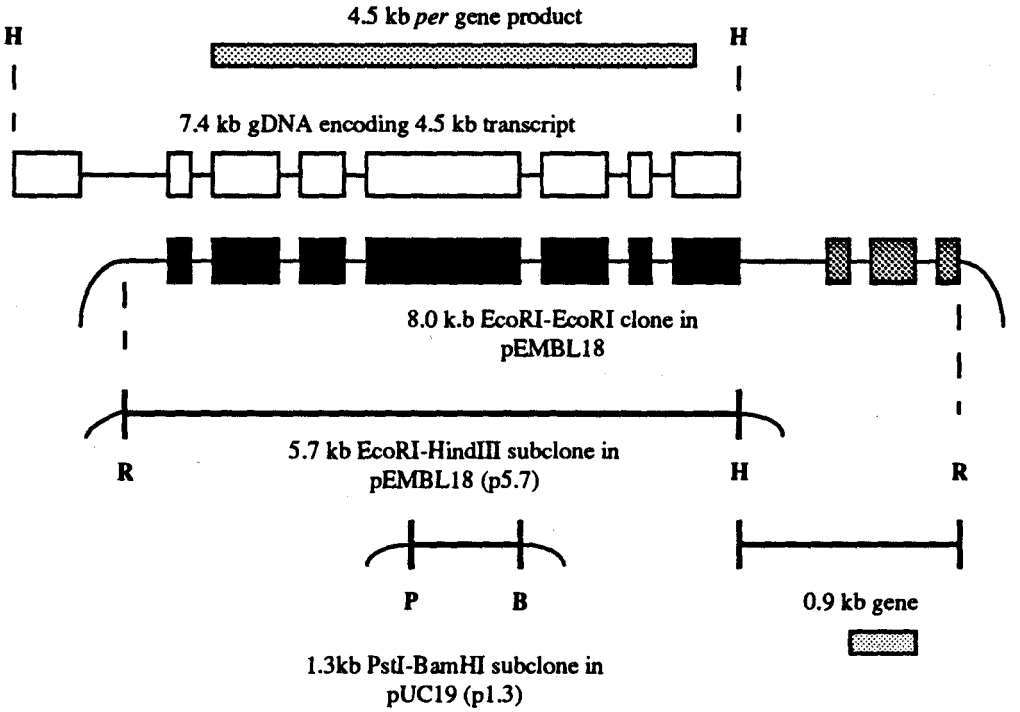
2.7.1 Ligation of Insert DNA

Construct p8.0 was double digested using the restriction endonucleases *EcoRI* and *HindIII*, run on a 0.8% TAE agarose gel, stained with 5 uL of ethidium bromide (0.5ug/mL), visualized under short wave ultraviolet light and the band corresponding to the 5.7 kb fragment excised. The fragment was purified using the Gene Clean II Kit from Bio 101. The insert DNA was then mixed with linearized *pEMBL*, also restricted with *EcoRI* and *HindIII*, at a 1:3 dilution. One-tenth the volume of 10mM ATP was added to the mixture followed by one-fifth the volume of 5X ligation buffer and 1 unit of T4 DNA ligase. The microfuge tube containing the ligation reaction was placed in an ice bucket and left in a 4 °C cold room overnight. The following day, 100 uL of fresh competent cells were aliquoted into three seperate microfuge tubes and three seperate dilutions of the ligation mixtures plus $T_{10}E_1$ (pH 8.0), were added to the cells. The solutions were mixed and left at 4 °C for 30 minutes to allow for incorporation of insert and vector DNA into the competent cells following which 900 uL of LB broth augmented with 20 mM glucose was added to each of the three microfuge tubes. The cultures were then incubated for 60 minutes at 37 °C with shaking. A control was also included in the above scheme that contained everything listed except for the ligation reaction mixture.

2.7.2 Screening Ligated Transformants

Each of the dilutions was plated onto LB agar plates containing ampicillin and X-gal in triplicate. 10 uL of the first dilution was spread on a plate, followed by 100 uL of the same

Figure 2.7: Schematic representation of the 8.0kb *EcoRI-EcoRI* clone along with the two subclones constructed for use in the Southern work. The 1.3kb *PstI-BamHI* clone was used for surveying the length variation of the Thr-Gly repeat region.


4.5 kb *per* gene product
H
H
7.4 kb gDNA encoding 4.5 kb transcript
8.0 k.b EcoRI-EcoRI clone in
pEMBL18
5.7 kb EcoRI-HindIII subclone in
pEMBL18 (p5.7)
R
H
R
P
B
0.9 kb gene
1.3kb PstI-BamHI subclone in
pUC19 (p1.3)

dilution on a seperate plate, followed by the remaining 880 uL on yet another plate. This was repeated for the second and third dilutions as well as the control reaction for a total of twelve plates, which were subsequently incubated overnight at 37 °C. After 17 - 20 hours, single white colonies were picked, cultured and harvested for mini-prep analysis followed by large scale plasmid preparation. In the case of p1.3, a similar scheme was followed. The p5.7 subclone was cut with the restriction endonucleases PstI and BamHI and the band corresponding to 1.3 kb excised and purified. Because pUC 19 has compatible cohesive ends in its multiple cloning site, it was restricted and linearized with the same endonucleases and used in the ligation reaction to the 1.3 k.b. fragment.

2.8 Restriction Digestion of gDNA

All digestions were carried out according to manufacturer's instructions (Boehringer Mannheim Canada). In the case of CsCl extracted DNA, genomic DNA was first concentrated prior to restriction. Based on UV spectrophotometric analysis, the required quantity was first precipitated using two volumes of 100 ethanol and 1/25 volume 5M NaCl, then washed in 70% ethanol followed by drying *in vacuuo*. The pellet was then resuspended in 15uL $T_{10}E_1$ (pH 7.5) and placed in a 37 °C water bath for 15 minutes in order to dissolve the pellet. Approximately 5 units of enzyme were used per microgram of gDNA in order to ensure complete digestion. Samples were mixed by flicking the microfuge tube several times followed by a short spin in a microfuge. They were then placed at 37 °C in a water bath for 4-6 hours. For gDNA extracted using the SDS lysis method, the same procedure was followed with the exception that there was no concentration step. Following incubation, the reaction was stopped by adding 0.5M EDTA (pH 8.0) to a final concentration of 0.01M. The samples were mixed with the appropriate

loading buffer, depending on whether agarose or acrylamide was being used, and the samples then loaded.

2.9 Agarose Gel Electrophoresis

DNA was separated on horizontal agarose gels using electrophoresis grade agarose from either Pharmacia LKB or Gibco BRL. Gels varied in concentration from 0.7% to 1.0% depending on desired fractionation size. Agarose was dissolved in either 1X Tris Acetate EDTA (TAE pH 7.5) or in 1X Tris Borate EDTA (TBE pH 8.3) with heating in a microwave oven to boiling. The agarose was allowed to cool to 50 °C prior to pouring. The corners of the electrophoresis rig were first sealed by dripping agrose to fill the seams followed by pouring to fill the tray (Appendix Figure A3). The comb was then inserted and the gel was allowed to solidify prior to filling the reservoirs with the appropriate 1X buffer, either 1X TAE or TBE depending on the buffer used for the gel itself. The gel was covered with buffer to a depth of 1mm above the gel and the wells flushed with buffer to clear them. Samples of gDNA, marker DNA or plasmid DNA were mixed with one tenth volume agarose gel loading buffer, loaded and run between 2-5 V/cm. Quantified lambda DNA digested with HindIII was used as a molecular weight marker on all agarose gels. Gels were removed after running the dye to the required distance (bromophenol blue runs at the same speed as dsDNA that is approximately 500 bp in length), placed in a tupperware container, covered with the buffer from the electrophoresis rig reservoir and stained with 0.5 ug/ml of ethidium bromide for approximately 30 minutes with gentle rotation on a shaking platform. The gel was then destained by soaking in water for 10 minutes and visualized using a Fotodyne short wave ultraviolet transilluminator and was then

photographed, if required, using a polaroid MP4 land camera, Polaroid 545 land film holder, Toshiba R60 filter and Polaroid type 57 land film.

2.10 Polyacrylamide Gel Electrophoresis

2.10.1 Gel Casting

Because agarose gels generally cannot reliably resolve fragments below 500 bp, polyacrylamide gel electrophoresis was performed in order to separate gDNA varying in size from 50 to 1500 bp, or more specifically, DNA restricted with four base recognizing endonucleases or 4 cutters. Depending on the quantity and concentration of acrylamide required for the gel, varying volumes of a 30% acrylamide stock solution were mixed with 5X TBE buffer (pH 8.3) and double distilled water and filtered through a 0.45 um syringe filter from Gelman Sciences. 10mL of 30% stock, 10mL of 5X TBE, 30mL of double distilled water were added together and filtered in order to pour a single 6% gel with 1.5mm spacers. 250uL of 10% ammonium persulfate was then added followed by 25uL of TEMED for a 6% acrylamide gel and the solution was mixed thoroughly but carefully as acrylamide is a potent neurotoxin . 20cm X 22cm outer plates and 20cm X 20cm inner plates were taped together along the sides and bottom using 1.5 mm spacers between the plates. The plates were thoroughly washed prior to taping, with mild detergent, followed by a rinse in 95% ethanol and then double distilled water. An 0.8% agarose solution was heated and dripped along the inner sides of the taped plates in order to facilitate sealing of the bottom as it was still prone to leaking despite taping. The acrylamide solution was loaded into a 60cc syringe and was carefully unloaded between the glass plates. When the plates were almost full, keeping the sealed end on the benchtop, the open end was gently lowered onto a piece of solid styrofoam so as to form a 10° angle with benchtop. At this

point, a 15 well comb from BioRad was inserted and bulldog clamps were placed across the length of the comb securing the comb between the plates (Appendix Figure A4). For our purpose, 50 mL of acrylamide was sufficient to pour a single gel given the size of the mentioned plates and the thickness of the spacers and all gels used for Southern analysis contained 6% acrylamide, enabling one to reliably score fragments ranging in size from 50 to 1500 bp on a single gel. The gel mold tended to polymerize within 30 minutes; if the gel had not set within this time period, it was discarded. After a period of two hours, the tape holding the plates together was removed as was the comb. The wells were carefully and thoroughly rinsed with double distilled water to remove any traces of unpolymerized acrylamide.

2.10.2 Vertical Electrophoresis

The plates were secured between brackets provided with the Protean Electrophoresis chamber from BioRad and the gel sandwich snapped onto the cooling core of the unit. The gel was then immersed into the vertical chamber containing 1X TBE (pH 8.3) buffer and was prerun at 50V for 30 minutes using a BioRad model 1000/500 power supply (Appendix Figure A5). Sample volumes ranging between 20 - 30 uL were then loaded using a standard micropipettor, after adding one fifth the volume acrylamide gel loading buffer, and the gel run at 2V/cm for 16 hours for genomic separations or until the bromophenol blue indicator dye in the loading buffer had migrated to 5cm from the bottom of the gel. It is possible to run such non-denaturing polyacrylamide gels at up to 10V/cm without significant loss in resolution. In fact, mini gels used for diagnostic purposes (ie. for testing low molecular weight markers), run in a similar, but smaller, apparatus from BioRad, where gel dimensions were 7cm X 10cm were generally run at this voltage. Low

molecular weight markers were constructed for use in vertical polyacrylamide gel electrophoresis and consisted of either *HinfI* cut *pUC18* (65-1550bp) or *HaeIII* cut *pUC18* (50-550bp).

2.11 Southern Blotting

2.11.1 Preparation of Agarose Gels for Transfer

Besides its use for genetic analysis, Southern blotting was also used a diagnostic tool in order to assess the quality of genomic preparations, completeness of restriction as well as the quality of the probe used. Methodology was essentially as outlined in Southern (1975) following many various modifications of Sambrook *et al.*, (1989). After running agarose gels restricted with six base recognizing endonucleases, gels were stained with ethidium bromide, and photographed as described above. The gels were then carefully placed in a larger tupperware container and covered with 0.25N HCl in order to fragment larger fragments of DNA facilitating their even transfer. This was performed with gentle agitation until the bromophenol blue indicator dye, still present within the gel matrix, turned green. The gel was then rinsed in double distilled water for two minutes and then transferred to a new tupperware container and soaked in a 1.0M NaCl/0.5N NaOH solution with agitation for two changes of 15 minutes. Treatment of the gel with an alkaline solution denatures the DNA for subsequent hybridization to the probe and also cleaves the molecule at the HCl depurinated sites. The gel was then neutralized by agitation in a 0.5M Tris (pH 7.5)/1.5M NaCl solution for two changes of 15 minutes. The DNA was then transferred out of the matrix onto a support via the capillary method.

2.11.2 Capillary Method

A large rectangular Corning baking dish was filled with 20X SSC (pH 7.0) and a rectangular piece of plexiglass, to be used as a platform, was placed perpendicularly over the glass dish. Three pieces of Whatman filter paper, slightly larger than the gel, were then placed over the platform and their ends allowed to drape into the reservoir of buffer in the dish. The gel was then placed onto the saturated filter paper and a piece of buffer equilibrated Nylon membrane was placed on top of the gel. A glass pipette was then rolled over the surface to remove any bubbles that may have formed between the filter paper and the gel and the gel and the Nylon membrane. Three more pieces of Whatman filter paper, slightly smaller than the dimensions of the nylon, were then placed on top of the nylon and bubbles removed as described. The sides of the gel were then surrounded by strips of Parafilm in order to prevent contact between the two layers of filter paper. A two inch stack of Kraft paper towels, cut to the same size as the top layer of fiter paper, was then placed on top followed by a 700g 20cm X 20cm square acrylic plate in order to secure the assembly (Appendix Figure A6). Transfer buffer was then added, if required, to the glass dish such that the lowest layer of filter paper was completely saturated. The open sides of the dish were then covered with plastic wrap in order to prevent evaporation of the buffer in the course of the overnight transfer. Transfer was allowed to proceed for 16 hours with one replacement of the paper towels as they became saturated. The mass flow of transfer buffer through the agarose gel to the absorbent layer of paper facilitates the elution of nucleic acids from the agarose gel onto the immobilizing matrix.

2.12 Electrophoretic Transfer

2.12.1 Background to the Technique

Smaller fragments of DNA (less than 500 bp) are inefficiently resolved on agarose gels. For better resolution of small fragments, polyacrylamide gel electrophoresis is used following transfer by electroblotting. Unlike in the capillary transfer where DNA fragments are carried from the gel in a flow of liquid and deposited on the surface of a solid support, in electroblotting, nucleic acids migrate from electrophoretic gels onto a solid support under the driving force of an electric field. This facilitates the elution of smaller fragments as well as a more even transfer. Disadvantages of this method include the requirement of large volumes of buffer to ensure that the buffering power of the system is not depleted due to electrolysis. Moreover, ohmic heating, which is a result of the current passed through the system, requires that electrophoretic transfer be carried out using extensive external cooling or else, transfer is significantly impaired. Nonetheless, the use of positively charged Nylon membranes allows fragments as small as 10 bp to be bound in buffers of very low ionic strength using electroblotting (Church and Gilbert, 1984). Many different types of positively charged nylon membranes were used in order to determine their efficacy for the electroblotting technique, not all of which proved successful. Of Schleicher & Schuell Nytran, Millipore Immobilon N PVDF, BioRad Zetaprobe GT+, Pall Biodyne A and ICN Biotrans, both Immobilon N PVDF and Zetaprobe GT+ gave consistent results using the protocol described and were thus used for all transfers.

2.12.2 Preparation of Gel for Electrophoretic Transfer

Following genomic separation, the plates were placed on a flat surface and one of the side spacers was removed. The plates were then pryed open using the removed spacer for

leverage taking care not to detach the gel from the bottom plate. The gel, attached to the glass plate, was then immersed in a large tupperware container containing 0.5 ug/mL ethidium bromide and allowed to shake gently for 45 minutes as polyacrylamide, unlike agarose, tends to quench the signal of ethidium bromide and therefore requires a longer period of incubation. Following staining, the gel was carefully slid off the plate and placed on an ultraviolet transilluminator and photographed as described previously. A piece of Whatman paper, cut to the exact dimensions of the gel was then placed on the gel and the gel was lifted and placed, Whatman side down, on the glass plate. The gel was then saturated in transfer buffer, which in the case of electroblotting is 0.5X TBE (pH 8.3). The apparatus used for transfer was either a BioRad Trans-Blot cell, or a custom designed electroblotter built by Charles Barbagallo of Polytech Products (Somerville, MA), depending on the size or number of gels being blotted. The gel was placed Whatman side down on a another piece of Whatman paper of the same dimensions, and this was then placed on a Scotch-Brite porous pad that was saturated with 0.5X TBE. A pasteur pipette was carefully rolled over the entire surface of the gel taking care not to tear the gel but ensuring that there were no bubbles trapped between the two layers of Whatman and between the gel/Whatman layer. It was found that when rolling the pipette over the surface of the gel, it was essential to do so lightly, in order not to contort or stretch the gel as this resulted in band shifting following hybridization. A piece of positively charged nylon, pre-equilibrated in transfer buffer, slightly larger than the gel itself, was then placed on top of the gel, followed by two pieces of saturated Whatman the same size as the two pieces underneath the gel. Bubbles were eliminated using a pipette at each stage to ensure that transfer would not be impaired. Finally, another porous pad, also saturated in buffer, was used to top the assembly (Appendix Figure A7).

2.12.3 Setting up the Electroblotting Chamber

The porous pad sandwich, containing the gel and membrane, was placed flat on a hinged mounting tray that was part of the blotter. The mounting tray was closed, taking care not to shift the assembly and this was then mounted between parallel plate electrodes in a large tank of buffer. A coated aluminum cooling core from a BioRad Protean II Multi Cell was modified and attached to an external refrigerated circulator from Canlab and placed within extra brackets of the transfer chamber and was allowed to circulate for approximately one hour allowing the transfer buffer to cool to approximately 2 °C. The chamber with cooling core was then placed on a magnetic stirrer and a stir bar was placed in the bottom of the chamber allowing for circulation of the buffer. It should be noted that this was generally done approximately one hour *prior* to removing the gel from the electrophoresis chamber. After mounting the gel between the plate electrodes, transfer was conducted at 75V/770mA for exactly 1 hour using a BioRad model 250/2.5 power supply (Appendix Figure A7). The time required for transfer is dependant on several factors including thickness and concentration of the poyacrylamide matrix as well as the strength of the electric field and as in the present study, the choice or lot of immobilizing matrix used. Because even high molecular weight nucleis acids migrate relatively rapidly from the gel, depurination/hydrolysis, prior to transfer, is unnecessary.

2.12.4 *In situ* Denaturation of Filter Bound DNA

Unlike in agarose, denatured DNA migrates out of polyacrylamide with reduced efficiency and thus denaturing is performed after transfer as opposed to prior to transfer as in the capillary method. Following electrophoretic transfer, the tray containing the gel sandwich was removed and disassembled and the DNA denatured *in situ* by placing the

membrane on top of three sheets of Whatman saturated with 0.4N NaOH for 10 minutes. The xylene cyanol FF marker dye that was also transferred from gel to membrane and was used to assess efficiency of transfer, turned yellow within seconds after placing the membrane on the NaOH containing filter paper. The membrane was then neutralized for 10 minutes on filter paper saturated with 2X SSC (pH 7.0), the xylene cyanol FF indicator turning turqouise again (Appendix Figure A8).

2.12.5 UV Crosslinking DNA to Nylon Filters

UV crosslinking ensures that DNA is retained on nylon filters during hybridization and when the filters are stripped if required for reuse. For charge modified nylon, UV crosslinking, or baking at 80 °C for that matter, as a means of covalently attaching nucleic acids to a solid support is not necessary. But for very small fragments of DNA or for reprobing of the molecule, UV crosslinking is recommended. UV crosslinking facilitates immobilization of the DNA molecule by covalent binding between light activated thymine bases and amine groups on the membrane surface. But, the binding of the molecule is dose dependant and overexposure of the blot results in decreased hybridization capacity as the DNA is sufficiently damaged. Following neutralization of the membrane, it was placed on a glass plate and covered with a layer of Saran Wrap. It is essential to use this type of plastic wrap because it is static free and because it does not stick to the membrane, allowing easy removal. The membane was then put on a platform approximately 23.5cm underneath a ballast containing two 15W germicidal light sources and irradiated at 800uW/cm^2 for exactly 2.5 minutes (Appendix Figure A8). This gives an effective UV dose of 2000 uW/cm^2 which was found to be the optimum from a series of experiments in which signal to noise ratio, following hybridization, was tested. This applies to a damp filter as the

optimal dose for dry membranes is much lower and was not determined. The UV flux at 23.5cm was determined using a Blak-Ray 260nm UV meter.

2.13 Probe Radiolabelling and Purification

2.13.1 Random Primer Extension

Probes to be used for hybridization were radiolabelled essentially according to Fineberg and Vogelstein (1984). All radiolabelling was performed using the Oligolabelling Kit from Pharmacia and P^{32} α d-CTP, specific activity 3000 Ci/mmol (NEN Dupont). Fragments to be used as probes were linearized, run on agarose, visualized and the appropriate band cut out of the gel. The band was purified and excess agarose removed using the Gene Clean II Kit and quantified using ethidium bromide staining against a prequantified control. Labelling was then performed according to manufacturers instructions using 50ng of template DNA at 37 °C for 2-3 hours. To the 50uL labelling reaction, 50uL of 1X TNE_2 was added to stop the reaction at the end of the time period specified and the tubes spun for 10 seconds in a designated radioactive microfuge.

2.13.2 Preparation of Spin Columns

Considering the period of extension and purity of input DNA, full incorporation of the radionuclide is impossible. Consequently, in order to remove unincorporated radiolabel, which contributed to high levels of background following hybridization in other experiments, labelling reactions were put through spin columns. The spin columns used were a modification of those comercially available from Pharmacia. A microfuge tube was pierced with an 18 gauge needle near the 1.5uL marking on the side followed by a piercing at the bottom of the tube with a 26 gauge needle. The side piercing was a vent hole

whereas the bottom piercing was a dripping hole. 500 micron rinsed glass beads from PolySciences Inc., were added to the microfuge tube until they occupied one third the distance of the slanting portion of the tube. To this, 1mL of TNE_2 saturated Sepharose CL6B from Pharmacia was added and the microfuge tube was then pushed through a pre-cut cap of an orange capped 15mL polystyrene Falcon centrifuge tube. The 15mL plastic tube was then placed in a clinical tabletop centrifuge from IEC and based on quantitative experiments performed to empirically determine optimum recovery based on spinning time, spun for exactly 1 minute and 45 seconds in order to compact the Sepharose slurry. The microfuge tube and cap were removed and screwed on to a new 15mL tube prior to adding the 100uL labelling reaction. After addition of the labelling reaction, the tubes were spun again for 1 minute and 45 seconds. The labelled probe was recovered and transferred to a fresh sterile microfuge tube and either stored at -20 °C for up to one week or quantified using scintillation counting and used immediately (Appendix Figure A9).

2.13.3 Scintillation Counting

In order to determine the specific activity of the probe, scintillation counting was performed. To a glass scintillation vial from Fisher Scientific, 1uL of the labelled probe was dropped on to a 1cmX1cm precut piece of Whatman filter paper sitting inside the vial. To this, 5mL of Scintillation fluid from Fisher Scientific was added and the vial placed within a Beckman Model LS1801 Liquid Scintillation System. Once auto-counted, the appropriate conversion was made in order to calculate the the specific activity of the entire sample. Probes were consistently labelled to specific activities ranging between 5.0×10^8-2.0×10^9 dpm/ug.

2.14 Nucleic Acid Hybridization

Nucleic acid hybridizations were carried out according to protocols prescribed with Gene Screen Plus hybridization membrane from New England Nuclear. Both hybridization and prehybridization were carried out in plastic hybridization bags from Gibco BRL or using Dazey Seal-a-Meal bags as conventional plastic bags often leaked during hybridization and were more difficult to free of air bubbles. Using Gene Jockey from Biosoft software for the Macintosh, it was determined that for linearized p1.3, a hybridizing temperature of 63 °C was optimum. Consequently, all prehybridizations and hybridizations were carried out at this temperature. Prehybridizations, hybridizations and washes were carried out in tupperware containers in a shaking incubator water bath from Techne.

2.14.1 Prehybridization of Filter Bound Nucleic Acids

The prehybridization step serves to block sites on the membrane that are free of hybridization sensitive sequences. Agents such as SDS, Blotto and Denhardt's Reagent serve as effective blocking agents. For the prehybridization step, approximately 0.0375 mL of Gene Screen hybridization solution were used per cm^2 of membrane. Any pockets of air were removed by rolling a pipette over the surface of the bag, draining any trapped solution to the bottom of the bag by lifting followed by expulsion of the bubbles to the outside of the bag. The bag was then sealed using a Meals-in-Minutes heat sandwich sealer from Sears. The bag was then submerged in a tupperware box containing water and allowed to incubate at 63 °C with shaking for 4-6 hours. Prehybridization can be carried out overnight but it was found that for single copy sequence detection, periods of incubation longer than the time specified resulted in diminished signals.

2.14.2 Hybridization of Filter Bound Nucleic Acids

Prior to hybridizing, the probe to be used was made single stranded by heat denaturation. The probe was boiled for 10 minutes followed by submerging in an ethanol/ice slurry for 5 minutes. Similarly, 20uL of salmon sperm DNA (50mg/mL) was heat denatured and was used as a carrier in the hybridization. Both the salmon sperm carrier and probe were then added to a plastic tube containing prewarmed hybridization solution and the solution mixed thoroughly. The prehybridization solution was completely removed from the bag by cutting an opening at the corner of the bag and the fresh hybridization solution added at 0.025mL per cm^2 of membrane. The bubbles were removed and the bag resealed prior to submerging in the tupperware container. The hybridization was allowed to proceed for 16-18 hours with agitation. Insufficient agitation often results in increased levels of background and is therefore a critical component of the hybridization step.

2.14.3 Washing Radioactive Filters

Following hybridization, the hybridization bags were cut open and the filters removed. The filters were placed in a tupperware container containing 250mL 2XSSC/0.1%SDS and were placed on shaker for 5 minutes at room temperature. The hot liquid waste was discarded and fresh wash solution was added and the filter washed for a further 5 minutes. Following the first set of room temperature washes, the filter was submerged in 250mL of a 0.2XSSC/0.1%SDS solution and placed back in the water bath shaker and incubated with agitation at 57 °C for 20 minutes. This wash was repeated another two times for 20 minutes each at 57 °C. Following the last 57 °C wash, the filter was removed and placed on top of a sheet of Whatman filter paper for 5 minutes in order to partially dry it. A

Geiger counter was run over the filter at both, the 1X and 0.1X setting in order to determine the radioactivity remaining on the filter. If any area on the filter not containing recognizable sequence was still hot, the filter was washed again at 57 °C in 0.2XSSC/0.1%SDS until the signal had significantly attenuated. Single copy and low abundance sequences are not expected to give signals at the 1X setting and consequently, only the pUC marker and *per* control tended to activate the Geiger counter.

2.14.4 Exposing Radioactive Filters

The damp filter was then placed inside Glad freezer bag, air pockets removed and sealed using the heat sealer. The sealed filter was then secured, using tape, to the inside of a 23cmX26cm autoradiographic cassette from Wolf containing two Cronex intensifying screens from Dupont. The cassette was taken to a dark room where a sheet of Kodak X-AR 5 imaging film was placed on top of the filter under a Kodak red filter light. The cassette was sealed, placed inside a black plastic garbage bag and left in the dark at room temperature. Exposure may also take place at -70 °C as it is supposed to stabilize the decay, but from experience, no difference was detected between the room temperature exposure and the -70 °C exposure. Filters were allowed to expose for 12-48 hours prior to developing in a Picker Cadi image developer.

2.15 Polymerase Chain Reaction Amplification

2.15.1 Reaction Conditions and Cycling Parameters

PCR amplification was carried out in 100uL reaction volumes using AmpliTaq polymerase. Reactions were carried out according to manufacturer's instructions using the GeneAmp Kit from Perkin-Elmer with minor modifications in a Perkin Elmer 480 thermal

cycler. Approximately, 61.5uL of double distilled water, 10uL of 10X Reaction buffer containing 10mM Tris-HCl (pH 8.3), 50mM KCl, 1.5mM $MgCl_2$ and 0.001% (w/v) gelatin to a final reaction concentration of 1X and 2uL of each dNTP to a final concentration of 200uM were added to a thin walled GeneAmp tube. To this, 0.5uL of AmpliTaq polymerase (2.5 units) were added and the mixture thoroughly agitated in order to disperse the Taq polymerase. Two 24-mer primers (Figure 2.15.1), chosen using Biosoft's Gene Jockey software for the Macintosh given the sequence of the *period* gene from Citri et al. (1987) were synthesized at the Mobix Central Facility at McMaster University and added at a final concentration of 1.0uM along with 200ng of CsCl purified genomic DNA template. The gDNA was heated at 95 °C for 5 minutes prior to adding it to the reaction in order to inactivate any proteases present. The 100uL reactions were then overlayed with 50uL of mineral oil and cycled for 1minute at 95 °C, 1 minute at 65 °C and then 1minute at 72 °C for 35 cycles. Due to the high G+C content of the primers, a high annealing temperature (65 °C) was employed. Approximately 5uL aliquots were removed following amplification and run on 2% agarose gels followed by ethidium bromide staining in order to visualize products.

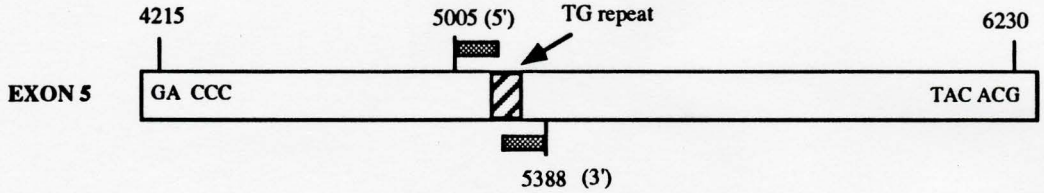
2.15.2 Purification of PCR Products

Desired fragments from the PCR reaction were re-amplified using the above conditions and then either run on 2% agarose gels followed by purification through sterile glass wool columns or directly ethanol precipitated in the presence of ammonium acetate, depending on the intensity and purity of the ethidium bromide visualized band. Bands that had to be cut out of the gel were placed inside a 500uL microfuge tube containing 2-3mm of sterilized glass wool which had been pierced at the bottom with a 20 gauge needle. This was then

Figure 2.15.1: Schematic representation of the fifth exon of *period* containing the Thr-Gly repeat region along with primers used for PCR amplification. The translated amino acids amplified along with the repeat are shown in the bottom panel.

Primer 1 (per 5'): 5'-CCC GTC CAC GAG GGC AGC GGG GGC-3' (5005-5029)

Primer 2 (per 3'): 5'-CCG CGC GAC TCG CGG TGC TTC TTC-3' (5364-5387)



PVHEGSGGSGSSGNFTTASNIHMSSVTNTSIAGTGG TG REPEAT NGTNSGTGTGASSSKGGSAAIPPVTLTESLLNKHNDEMEKFMLKKHRESRG

per 5' per (3')

placed within a 1.5mL microfuge tube, forming a two-tiered system. The wool was washed twice with $T_{10}E_1$ prior to placing the gel slice in side the smaller tube. The gel slice was then spun at 6000rpm for ten minutes and the samples were pooled prior to extraction with phenol:chloroform. The nucleic acids were then precipitated with ethanol in the presence of NH_4OAc at -20 °C. Single, intense bands, not requiring gel purification, were directly precipitated with ethanol using 100uL of the PCR reaction. The precipitate was allowed to form at -20 °C for 1 hour and followed by a wash in 70% ethanol. The product was then dried on the benchtop, resuspended in 10uL $T_{10}E_1$ (pH 7.5) and 2 uL were quantified using spectrophotometry.

2.15.3 Sequencing of PCR Products

All sequencing of PCR products was conducted in the MOBIX Central facility. Essentially, a cycle sequencing procedure was used which employed a two step cycling profile in which annealing was carried out at 50 °C, extension or Taq reaction at 60 °C and denaturation at 96 °C. Approximately 200 - 250 nanograms of DNA was required as starting template and 3.2 picomoles of primer with a final concentration of 0.8 picomoles/uL. All subsequent analysis of sequence chromatograms was performed on a Macintosh IIvx computer using Applied Biosystems Incorporated's Sequence Editor 1.03 software.

2.16 Statistical Analysis

Three statistical tests specifically designed to evaluate spatial structure were applied to the Thr-Gly frequency distributions. Specifically, they were employed in order to detect the presence of an underlying clinal pattern in the data, whether monotonic or non-

monotonic. The same tests were further applied to data on seven allozymes (*G6-pd, Adh, Fuc, 6-Pgd, Gpd, Odh Pgm)* surveyed in 10 latitudinal populations encompassing the cline. The purpose of this comparison was to acertain whether a similarity existed in the degree and pattern of spatial structuring in two classes of molecular data - nuclear versus protein.

2.16.1 The Royaltey-Astrachan-Sokal Test

The first test applied was the Royaltey-Astrachan-Sokal non-parametric test (Royaltey *et al.*, 1975). Specifically this test was designed to assess whether observed patterns of geographic variation can be considered significant departures from the null hypothesis of random variation over an area. The detailed methodology of this test is described in preceding reference but essentially, each locality along a transect, for example, is considered a vertice in a graph whose edges are connections between the localities based on the criteria of geographic contiguity. That is to say, the localities are connected in space and these connections are considered edges of the graph. In the case of the present North American transect, a simple chain or connected graph results, which negates the use of more complex Gabriel-connected networks which must satisfy the conditions of triangular inequality, traditionally constructed when employing this test. The variables at each locality, (in our case individual allele frequencies), are then ranked and the distributions of absolute differences in rank along edges between vertices yields various statistics of edge lengths that can be compared to expectations. Specifically, if we consider a graph of n vertices (5 for the Thr-Gly data and 10 for the allozyme data) and e edges (4 for the Thr-Gly data and 9 for the allozyme data) with incidence I_j at vertex j, where the incidence is the number of edges joining the vertex j (equivalent to 2 in all cases except for populations at

the termini of our chain graphs in which the value equals 1), then the number of edges $e = \sum I_j$. If the variables assigned to the n localities are ranked such that locality i is given rank R_i, we can then compute a rank difference $L_{ij} = |R_i - R_j|$, referred to as edge length, which corresponds to each edge between points i and j in the graph. In other words, if we have n points with e number of edges joining theses points, the length of an edge joining a and b will be defined as $L(a, b) = |f(a) - f(b)|$. The mean edge length will then be $L = \sum L_j / e$.

2.16.2 Spatial Autocorrelation Analysis

The second test conducted was spatial autocorrelation analysis. The purpose of this test was to determine whether the observed allele frequency at one locality was dependant on the values at another locality. The test provides a detailed description of gene frequency variation in space, and is independant of preliminary assumptions about underlying population structure. The technique has been described in a number of publications (Sokal and Oden,1978; Sokal, 1979; Cliff and Ord, 1981) and has been applied to various populations (Sokal and Menozzi, 1982; Epperson and Clegg, 1986; Costa *et al.*, 1992). In spatial autocorrelation analysis, all possible pairs of localities are grouped into arbitrary distance classes and spatial patterns of allele frequencies are summarized by a set of similarity coefficients, each one calculated within a distance class. As such, it is essential to first define which localities are neighbours. The statistic used in this case, was Moran's ***I*** as opposed to Geary's ***c***, which is also, but less commonly, used in spatial autocorrelation analysis and represents the degree of genetic similarity between populations as a function of their distance apart. Moran's ***I*** is expected to be very close to zero under the null hypothesis of no spatial structure and positive or negative, respectively, in distance classes within which pairs of allele frequency values fall on the same side or opposite side of the

mean. Thus, a significant positive autocorrelation indicates that at the distance considered, allele frequencies are similar, ie. both deviate from the mean in the same direction. Alternately, a significant negative autocorrelation represents dissimilarity and a nonsignificant value means that there is no consistent relationship between pairs of allele frequencies at that distance. A cline is considered to be present when Moran's *I* decreases continuously with increasing distances between populations (Sokal, 1979) from significantly positive to significantly negative.

In order to tabulate the spatial autocorrelation statistic, several expressions must be computed. The analysis can be conducted on nominal, ordinal and interval data and the reader is urged to consult Sokal and Oden, (1978) for greater detail pertaining to the mathematics as the derivations presented in the following are intended to give merely an outline of the basic statistic. The Moran coefficient *I* is represented as $I = n\sum w_{ij}z_iz_j / W\sum z_i^2$. In this case, the expression $z_i = x_i - x$, where x_i is the value of the variable x (allele frequency in our case) at locality i, x is the mean of x across all localities. Similarily z_j is the difference between the observed value and average value at locality j. w_{ij} is merely the weight associated with the two localities. That is, if locality i and locality j are neighbours within a prescribed distance class, $w_{ij} = 1$. If they are not neighbours within a given distance class, $w_{ij} = 0$. For example, Cartersville and Louisville would be neighbours in the 450 mile class (described in the Results) and subsequently, the value of w in this case would be 1. They would not however, be neighbours in the 900 or 1250 mile classes and would therefore have a w equal to 0 in both of these comparisons. The quantity W is the sum of the adjacency matrix weights, that is, the total sum of "1's" in a given comparison for a distance class or more simply, twice the number of edges in the

connecting graph. z_i^2 is the sum squares of the deviates and n is the total number of localities (5 for the Thr-Gly data and 10 for the allozyme data).

2.16.3 Mantel Tests of Matrix Association

As an alternative way of relating genetic distance between these populations to their geographic distance a third test was employed, namely Mantel tests (Mantel, 1969; Sokal *et al.*, 1987; Costa *et al.*, 1992). Unlike spatial autocorrelation analysis, Mantel tests serve to indicate the strength of the association between genetic distance and space, latitude or any environmental parameter. As such, it allows one to predict possible extrinsic causes for allelic frequency distributions. In the present study, symmetric distance or similarity matrices were compared by Mantel tests using a **Z** statistic. The **Z** statistic is calculated as $Z = \Sigma\Sigma\ (X_{ij}\ Y_{ij})$ where X_{ij} and Y_{ij} are the elements of two square, not necessarily symmetrical matrices. X_{ij} is the measure between points i and j of matrix X (which was genetic distance or allele frequency difference in this case) and Y_{ij} is the measure between the same points in matrix Y (which was spatial or latitudinal distance). The value of **Z** depends on the type and degree of association between the corresponding elements of the constituting matrices, X and Y; generally, the more positive the value of **Z**, the greater the association.

Chapter 3

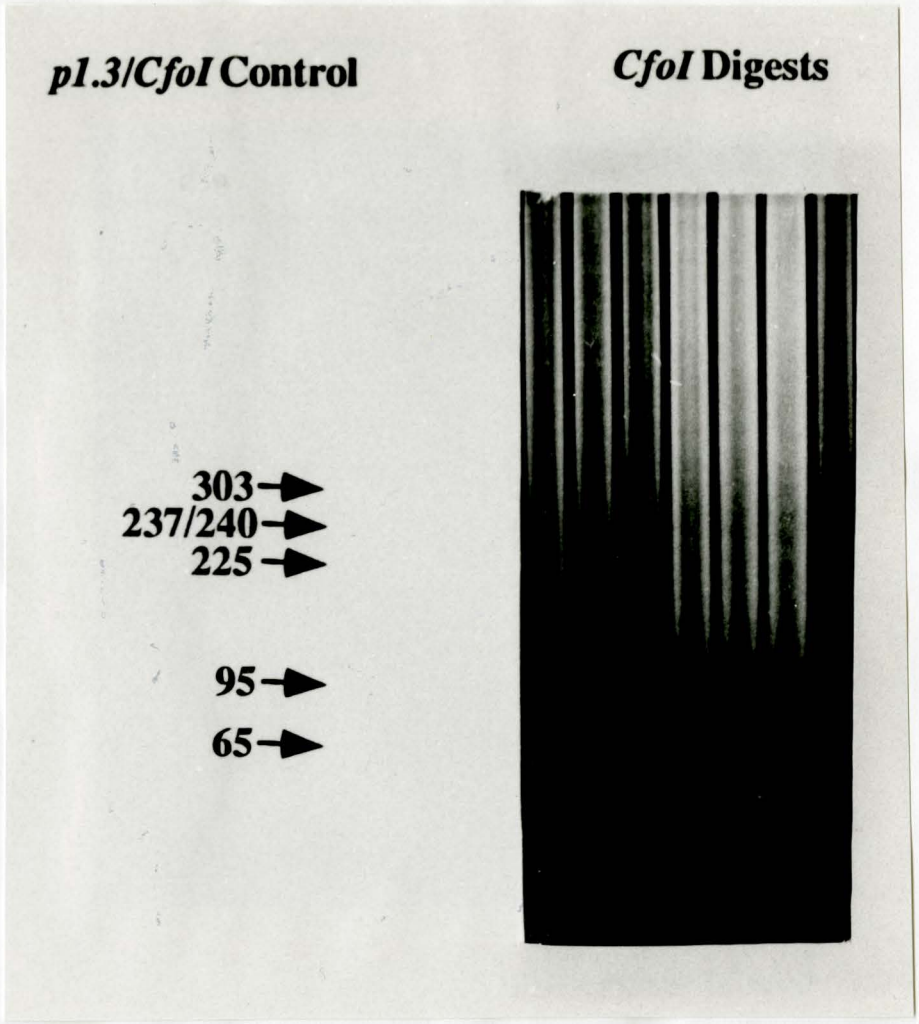
RESULTS

3.1 Variation in the Thr-Gly Repeat Region

Variation in the Threonine-Glycine repeat region (hereafter referred to as the Thr-Gly or TG repeat) was surveyed amongst the five populations. From the original 8.0kb *EcoRI-EcoRI* clone containing the entire *per* coding region, two subclones were constructed, a 5.7kb *EcoRI-HindIII/pEMBL18* construct (p5.7) and a 1.3kb *PstI-BamHI/pUC19* construct (p1.3) (Figure 2.7). Using the p1.3 subclone of the period gene, polymorphism in the Thr-Gly region was scored . From the work of Yu *et al.*, (1987), it was determined that *CfoI* digested gDNA produced a series of bands which when hybridized with p1.3, would reveal the polymorphism in the region of exon 5 of *period* containing the Thr-Gly minisatellite (Figure 3.1a).

The frequencies of the different Thr-Gly length variants in the different populations are shown in Table 3.1. The most frequent alleles are represented by the designations TG17, TG20 and TG23. These are the variants originally described by Yu *et al.*, (1987) and represent 17TG repeats, 20TG repeats or 23TG repeats, respectively, within the region described. Essentially, an 18bp "cassette" is responsible for the difference between the TG17 and TG20 alleles and two such "cassettes" result in the difference between the TG17 and TG23 alleles (Figure 3.1b). In our work, another variant was detected using the Southern technique that was intermediate in mobility between the TG20 and TG23 variants (Figure 3.1c). Initially, it appeared as though this variant might be an artifact or be associated with incorrect scoring of autoradiograms since no other group working with *period* had detected this variant, despite TG15 and TG18 variants being discovered in rare

Figure 3.1a: A 6% non-denaturing acrylamide gel containing *CfoI* digested gDNA following ethidium bromide staining. An autoradiogram of the same gel following electrophoretic transfer and hybridization with the p1.3 probe is shown below.


p1.3/CfoI Control
CfoI Digests
303
237/240
225
95
65

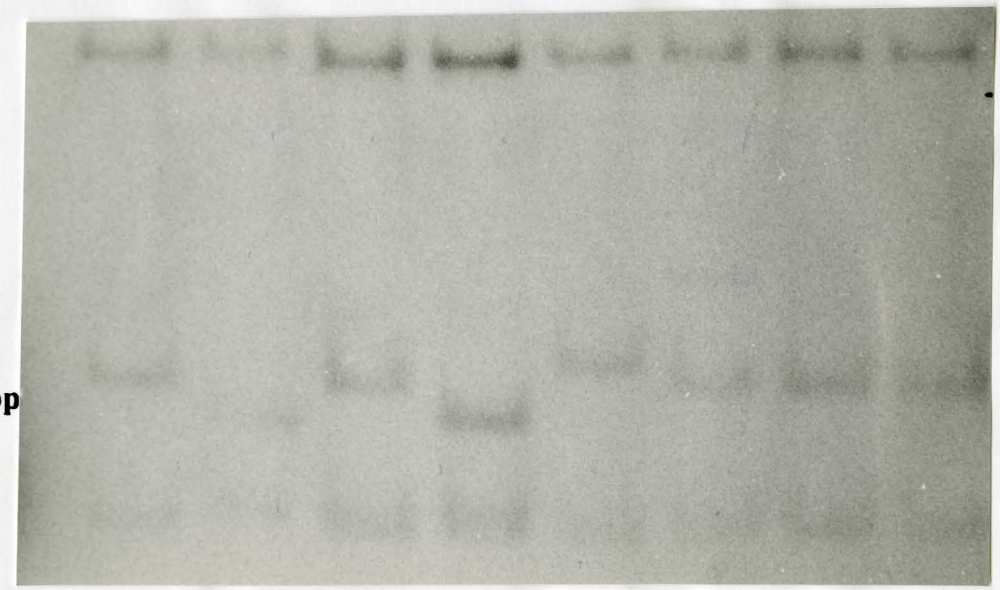

303bp

Table 3.1
Thr-Gly Allele Classes for Lines Across Populations

Tampa, FLA.		Cartersville, GA.		Louisville, KY.		Windsor, ON.		Ottawa, ON.	
TA02	3	CA1	2a	LA3	2b	WA4	3	OT1	2a
TA05	2a	CA2	2a	LA5	2a	WA7	2a	OT2	1
TA13	1	CA3	2a	LA7	2a	WB1	3	OT3	2a
TA15	2a	CA6	1	LA9	2a	WB6	2a	OT4	2a
TA17	1	CB1	1	LA10	2a	WB8	1	OT5	2a
TA18	2a	CB3	3	LB1	1	WB11	1	OT6	2a
TA20	2a	CB4	2a	LB3	2a	WC2	2b	OT7A	2a
TA23	1	CB5	2b	LB4	1	WC3	3	OT7B	3
TA25	1	CC2	2a	LB7	2a	WC5	2a	OT8	1
TA26	1	CC3	1	LB8	2a	WC8	2a	OT9	1
TA29	1	CC13	1	LB9	1	WC11	2a	OT10	2a
TA30	1	CC7	1	LC1	2a	WC13	1	OT11	1
TA34	1	CC8	1	LC2	2a	WC15	2a	OT12	2a
TA36	2a	CD1	2a	LC3	2a	WD5	1	OT13	3
TA37	1	CD3	1	LC5	3	WD6	1	OT14	1
TA40	1	CD5	2a	LC7	2b	WD8	1	OT15	1
TA51	1	CD8	2a	LC8	1	WD14	1	OT16	2a
TA52	1	CD15	2b	LD1	1	WE1	2a	OT17	1
TA62	1	CE1	1	LE2	1	WE3	2a	OT18	2a
TA76	1	CE3	1	LE3	2a	WE6	1	OT19	2b
TA80	1	CE4	1	LE5	2b	WE8	1	OT20	3
TA82	2a	CF1	1	LF1	2b	PHA1	1	OT21	2a
TA84	2a	CF2	1	LF4	2a	PHA2	2a	OT22	2a
TA93	2a	CF3	2a	LF5	2a	PHA3	2a	OT23	3
TA99	1	CF7	1	LF7	2a	PHB1	2a	OT34	3
						PHB2	2a		
Counts	*1=16*		*1=13*		*1=6*		*1=10*		*1=7*
	2a=8		*2a=9*		*2a=14*		*2a=12*		*2a=12*
	2b=0		*2b=2*		*2b=4*		*2b=1*		*2b=1*
	3=1		*3=1*		*3=1*		*3=3*		*3=5*
Total:	***25***		***25***		***25***		***26***		***25***
Frequency									
1=17 TG	**0.64**		**0.52**		**0.24**		**0.385**		**0.28**
2a=20 TG	**0.32**		**0.36**		**0.56**		**0.462**		**0.48**
2b=21 TG	**0.00**		**0.08**		**0.16**		**0.038**		**0.04**
3=23 TG	**0.04**		**0.04**		**0.04**		**0.115**		**0.20**

Figure 3.1b: A schematic representation of the Thr-Gly region sequence showing the 18bp cassette responsible for generating the three major size classes of repeats, TG17, TG20 and TG23. The cassette is represented in bold type and the remaining Thr-Gly pairs in normal type.


5'- GGCACTGGTGGCACGGGCACTGGTACAGGTACAGGTACTGGAACTGGAACTGGA
18 bp CASSETTE
ACGGGAACAGGTACAGGCACAGGCACAGGCACTGGAACAGGCAATGGAACA-3'

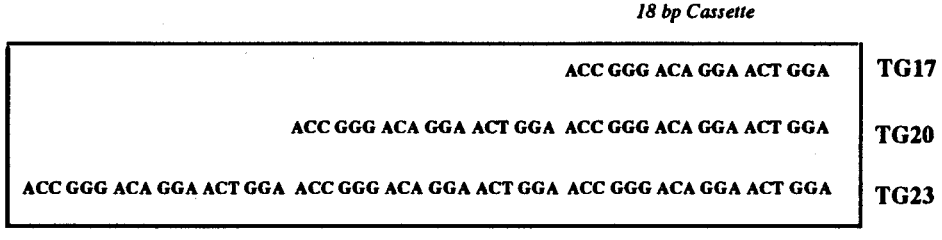
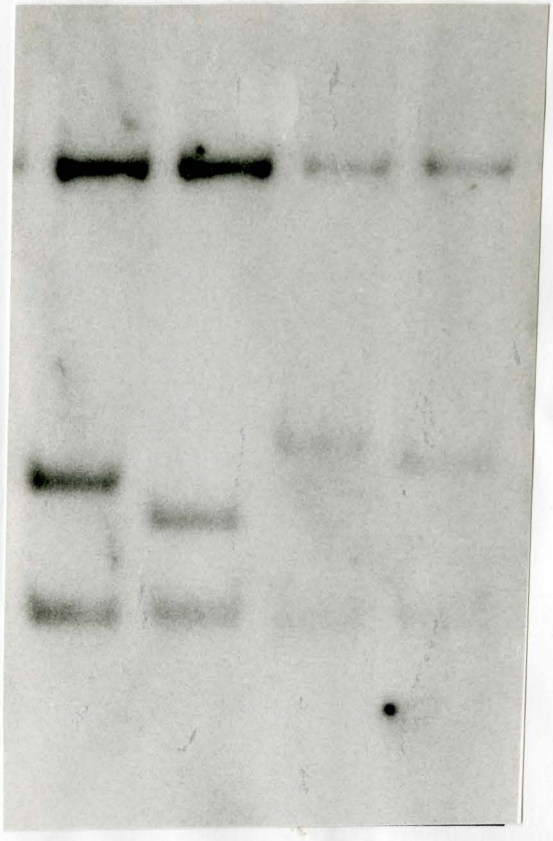

18 bp Cassette
ACC GGG ACA GGA ACT GGA
TG17
ACC GGG ACA GGA ACT GGA ACC GGG ACA GGA ACT GGA
TG20
ACC GGG ACA GGA ACT GGA ACC GGG ACA GGA ACT GGA ACC GGG ACA GGA ACT GGA
TG23

Figure 3.1c: An autoradiogram showing the Thr-Gly variant demonstrating the mobility shift.

TG20 TG17 TG23 TG22



frequency in European populations (Costa *et al.*, 1991). When individual lines carrying this repeat were run along side lines carrying the other modal class repeats, it was found that this was in fact a true variant and not the result of some artifact or mobility shift (Figure 3.1d). PCR amplication of lines carrying this repeat followed by subsequent sequencing of the product revealed this as true strain polymorphism. The sequence analysis conducted on various Thr-Gly alleles as part of this study is presented in section 3.6.

3.2 Latitudinal Variation in TG17 and TG20 Allele Frequency

The work of Costa *et al.*, (1992), on the Thr-Gly repeat region in latitudinal European populations of *D. melanogaster* showed certain trends in allele frequency distributions. Specifically, the TG17 allele tended to occur much more frequently in Southern latitudes as compared to the North, reaching frequencies of upto 0.762 in Casablanca and as low as 0.379 in Bristol (see Table 3.2). Conversely, the TG20 allele showed the opposite trend, being present in greater frequency in Northern populations (0.517 in Bristol) and less so in the Southern (0.096 in Casablanca); the TG22 and TG23 variants occuring in low frequency and being relatively rare. In our study, similar trends were evident. From Table 3.1, it can be seen that the frequency of the TG17 allele is high in Southern populations and low in more Northern populations and that an opposite trend is observed for the TG20 allele, although not as pronounced as for the TG17 variant. The frequency of the TG17 allele reaches 0.64 in Tampa Bay and drops to 0.28 in populations from from Ottawa. Conversely, the frequency of the TG20 allele is 0.32 in Tampa Bay and increases to 0.48 in Ottawa. The data from the Louisville population seem to confound these trends as is evident from Figure 3.2. The standard error associated with the frequencies of the TG17 and TG20 alleles shows that this difference is significant. The

Figure 3.1d: A 2% agarose gel showing the four different size classes of Thr-Gly repeat regions observed in the present study following PCR amplification. The allele described as demonstrating the shift is clearly intermediate in mobility between the TG20 and TG23 alleles.

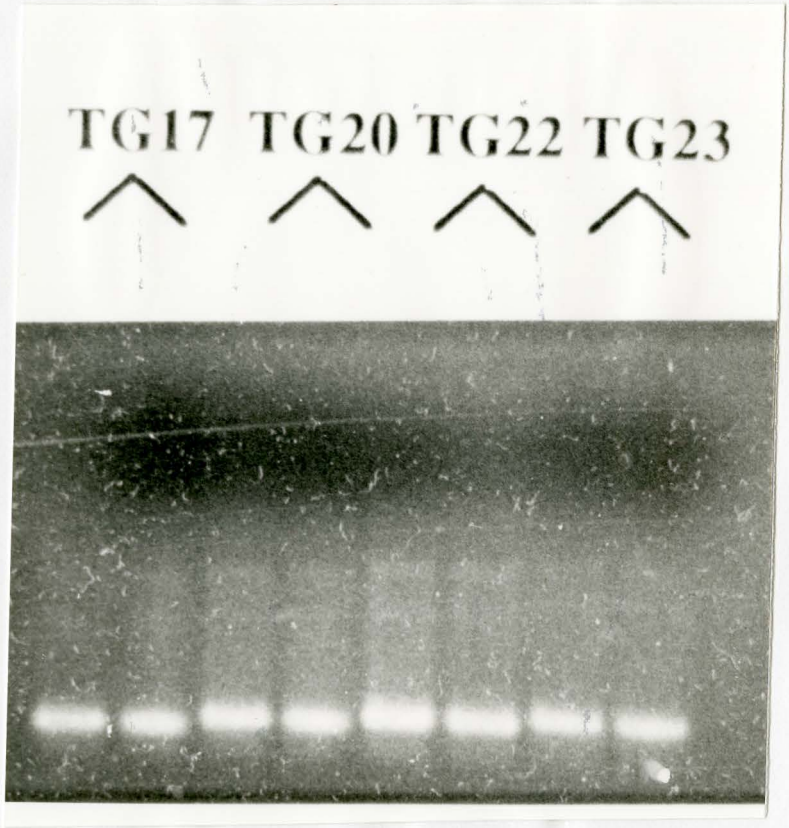
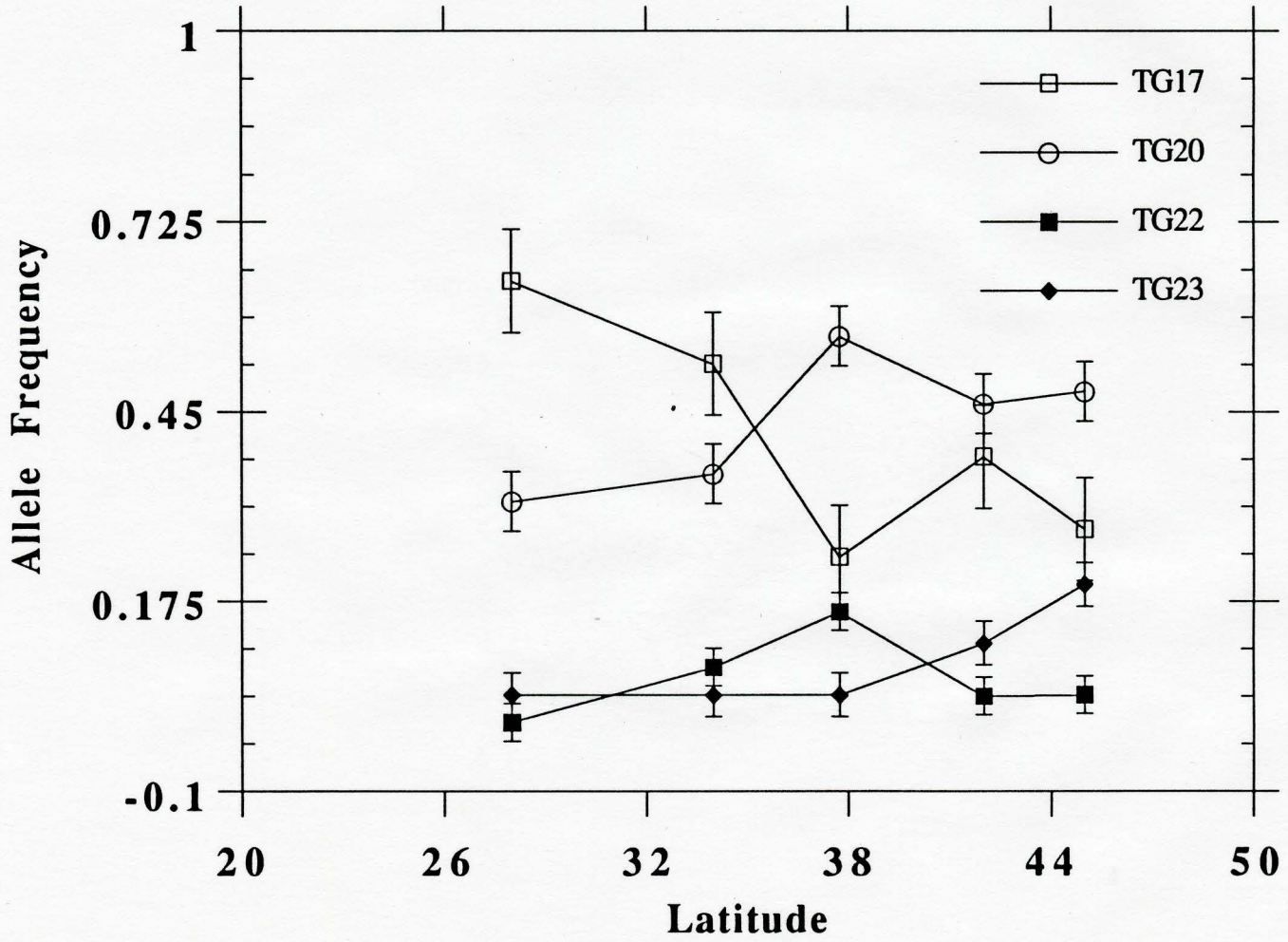

TG17
TG20
TG22
TG23

Table 3.2
Frequencies of Thr-Gly Variants From European *Drosophila* Populations[1]

	Thr-Gly Length Variants						
Population	**14**	**15**	**17**	**18**	**20**	**21**	**23**
Casablanca			0.762	0.047	0.096		0.096
Rethimnon	0.040		0.680		0.240		0.040
Ravanusa			0.655		0.172		0.172
Zakynthos			1.000				
Andros			0.571		0.285		0.143
Trapani			0.647		0.294		0.059
Sibari	0.077		0.538		0.307		0.077
Albaida	0.033		0.466	0.033	0.400		0.067
San Mateo	0.100		0.466	0.033	0.366	0.033	
Pietrastornina	0.021		0.566		0.326		0.087
St Tropez			0.325		0.650		0.025
Bologna	0.040		0.640		0.200		0.120
Bordeaux			0.522		0.348		0.130
Cognac	0.028	0.028	0.277		0.444	0.028	0.194
Merano	0.033		0.400	0.033	0.333		0.200
Tiszafured			0.451		0.290		0.258
Bristol			0.379		0.517		0.103
Leiden			0.286		0.714		

[1] Table was reproduced from Costa *et al.*, (1992).

Figure 3.2: Frequency plot of Thr-Gly allele frequencies across the various populations surveyed.


1
0.725
0.45
0.175
-0.1
Allele Frequency
20
26
32
38
44
50
Latitude
TG17
TG20
TG22
TG23

frequency of the TG17 allele drops rapidly from Cartersville to Louisville and then increases again in the Windsor population (Figure 3.2). Similarly, the TG20 allele increases rapidly in Louisville populations relative to Cartersville and then drops again in the Windsor lines (Figure 3.2). Fluctuations of this sort are expected due to random variation and are even evident in the work of Costa *et al.*, (1992).

3.3 Statistical Analysis of Thr-Gly Frequency Data

Because the TG17 and TG20 alleles were by far the most prevalent and because they seemed to vary latitudinally, they were the focus of all subsequent statistical analyses. The reader is asked to refer to the Methods (sections 2.16.1 - 2.16.3) for more information regarding the purpose and methodology behind each of the statistical tests performed.

3.3.1 Departure From Randomness of Thr-Gly Frequencies

All populations surveyed showed polymorphism to varying degrees in the frequencies of Thr-Gly alleles. Because the TG17 allele displayed the strongest clinally varying trend, its frequency was compared with the cumulative allele frequencies of all other alleles across localities using a G-test (Sokal and Rohlf, 1981). The purpose of this preliminary test was to check if the observed allele frequencies showed a significant departure from randomness. By means of a 5 x 2 contingency table, a G = 11.66 was obtained which is significant at $p < 0.025$ using a $\chi^2_{(4d.f.)}$, showing that the observed gene frequencies do not represent random samples drawn from a panmictic or genetically uniform population.

3.3.2 The Royaltey-Astrachan-Sokal Non-Parametric Test

To further assess if the observed pattern of geographic variation departed significantly from random variation over an area, the Royaltey-Astrachan-Sokal non-parametric test was applied (Royaltey *et al.*, 1975). When the variables, in this case allele frequencies for TG17 and TG20, were ranked and absolute differences between ranks between localities were determined, no significant value for mean edge lengths were found. For the TG17 allele, an average edge length, $L = 1.75$ was obtained and for the TG20 allele, an edge length of $L = 1.75$, neither of which were significant, $p < 0.15$ on a one-tailed exact test (Figure 3.3.2 a & b). Our case was special in that our data formed a simple chain (a tree with only two terminal vertices). As such, the number of edges was easily computed by $e = n - 1$, where n equals the number of populations, and a maximum value of $I_j = 2$. This method for the analysis of spatial variation is sensitive to the number of localities sampled and works optimally for $n > 9$, although there are several publications in which values of $n < 9$ have been employed for this test (Royaltey *et al.*, 1975 and references therein). With $n < 5$ however, no significant trend can be established using the prescribed statistics. Our data fall on the threshold of sensitivity of this test and as a result, the insignificant pattern observed is not all that surprising considering that even the slightest departure from a perfectly monotonic trend in allele frequency will result in reduced significance. As an example of how sensitive this technique becomes to small numbers of n and deviations from a perfectly monotonic trend in allele frequency, allele frequency distributions were manipulated so that the Louisville and Windsor TG17 allele frequencies were switched around then ranked. When the test was performed with the modified data

Figure 3.3.2 a & b: Royaltey-Astrachan-Sokal test performed on TG17 and TG20 alleles. Panels show each population arranged in a simple chain as they would be in space. Each population is ranked with rank number in bold. Values in parenthesis next to each population represent allele frequencies observed and mean edge length computed from the graph is shown below each chain. Panel A represents the test for the TG17 allele and panel B shows the test on the TG20 allele.

A.

4 • OT (0.28)

3 • WI (0.385)

5 • LO (0.24)

2 • CA (0.52)

1 • TB (0.64)

TG17 *L = 1.75*

B.

2 • OT (0.48)

3 • WI (0.462)

1 • LO (0.56)

4 • CA (0.36)

5 • TB (0.32)

TG20 *L = 1.75*

(rank order 1-2-3-5-4 from Tampa Bay to Ottawa), a mean edge length $L = 1.25$ was obtained which was significant ($p < 0.05$) by a one-tailed exact test. Furthermore, when the Louisville data was dropped all together, a perfect rank order was obtained resulting in a mean edge length of $L = 1.0$ which was highly significant ($p < 0.001$). Both examples were included in order to illustrate the sensitivity of this test to small sample sizes and the requirement of a perfectly monotonic trend in allele frequency required for significance in instances in which a very small number of localities are sampled.

3.3.3 Spatial Autocorrelation Analysis on Thr-Gly Alleles

Because the Royaltey-Astrachan-Sokal non-parametric test uses unweighted data, that is, variables are ranked but equally so in space, spatial autocorrelation analysis was conducted in order to determine whether the observed allele frequency at one locality was dependant on the values at another locality (Sokal and Oden,1978; Sokal, 1979; Cliff and Ord, 1981). If such a correlation exists, the allele frequency is said to exhibit spatial autocorrelation. Recalling from the Methods (section 2.16.2), it is first essential to partition the data into arbitrary distance classes. The data for the TG repeat frequencies were divided into three distance classes - 450, 900 and 1250 miles, and as is customary in spatial autocorrelation analysis, the least common of the three frequently occuring alleles, TG23, was excluded from the analysis (Sokal *et al.*, 1989; Costa *et al.*, 1992). All pairs of localities that were neighbours within a given distance class were clumped together and used to determine a value of Moran's I specific for that distance class (ie. all pairs of localities within 450 miles of one another were clumped together as were all pairs of localities that were neighbours between 451 and 900 miles of each other, etc.). For the TG17 allele, a steady decline, from positive within 450 miles to highly negative at large

distances, was observed for Moran's *I* indicating a pattern of spatial structuring. For the TG20 allele, a similar pattern was observed with Moran's *I* decreasing from positive at close distances to negative at farther distances (table 3.3.3). The significance of the *I* values obtained could not be determined using the conventional methods prescribed by Sokal and Oden (1978) because the number of populations used to conduct the analysis was lower than that required to use these methods. For large numbers of populations, the distribution of *I* tends towards normality and thus, the prescribed methodology for determining levels of significance can be applied. For small populations, or $n < 8$, the authors suggest working out the distribution of *I* under random permutations. Consequently, an algorithm that permutes the observed frequency distribution of a particular allele across the various localities and then repeatedly recalculates values of the *I* statistic for each set of permutated data was made using Hypercard 6.1 for the Macintosh (see Appendix C for algorithm). The permuted data were used as the null distribution and used to evaluate the significance of the observed *I* statistics for the various distance classes. The simulation results indicate that the observed *I* value or autocorrelation statistic for all neighbours within 450 miles for the TG17 allele was marginally significant ($I_{450} = 0.50128$, $p < 0.06$) and significant for the third distance class ($I_{1250} = -1.2970$, $p < 0.05$), indicating, at least at the level of individual autocorrelation statistics, significant spatial structuring for this allele. The result of the permutations for the TG20 allele revealed that the observed autocorrelation statistics were not significant for any distance class, or rather, did not vary from those *I* values expected under the null hypothesis of no spatial structuring. A correlogram or a plot showing how spatial autocorrelation varies as a function of distance between neighbours was constructed for the TG17 and TG20 alleles

Table 3.3.3
Autocorrelation Statistics for TG17 and TG20 Alleles
(overall *p* values for correlograms are given)

	Distance Class (in miles)			
	450	900	1250	*p*
TG17	0.501279	-0.477801	-1.2970	< 0.1
TG20	0.12530	-0.47976	-0.5410	< 0.3

and the overall significance of the correlograms was determined using the Bonferroni procedure as described in Sakai and Oden, (1983) and Sokal *et al.*, (1987). The results indicate that although the overall correlograms are not significant, the directional trend in autocorrelation is indicative of a significant clinal pattern for the TG17 allele (Figure 3.3.3).

3.3.4 Mantel Tests and Regression Analysis on Distance Matrices

As was described in the Methods (section 2.16.3) as an alternative way of relating genetic distance between these populations to their geographic distance, Mantel statistics were computed (Mantel, 1969; Sokal *et al.*, 1987; Costa *et al.*, 1992). Prior to conducting the Mantel tests of matrix association, simple linear regression analysis was performed on the parameters of the matrices used for comparison. Genetic distances were computed for all alleles across localities according to the method of Nei (1972) as well as matrices for geographical distances between the populations and latitudinal distances (Table 3.3.4). The result of regressing genetic distance on geographical distance, using a simple linear model resulted in a regression coefficient of $b = 0.00005127$ ($p < 0.2$) which is clearly not significant. Similarly, regression of genetic distance on latitudinal distance resulted in a regression coefficient of $b = 0.004368$ which is also not significant ($p < 0.08$). Because of the small sample sizes used for the regression analysis, even a single outlier can distort the regression. In our case, the poorly fitting Louisville population discussed earlier, is clearly responsible for the negative result. When the Louisville-Cartersville and Louisville-Tampa Bay genetic, geographical and latitudinal distances were removed from the analysis, highly significant regression coefficients were observed ($b = 0.0000818$, $p < 0.003$ and $b = 0.005161$, $p < 0.005$) for geographical and latitudinal distance, respectively, regressed on genetic distance. Mantel tests were then conducted by obtaining a null distribution of Z

Figure 3.3.3: Correlograms for the TG17 and TG20 alleles. Overall p values of the correlograms using the Bonferroni criterion are $p < 0.1$ and $p < 0.3$ respectively despite significant values for several individual autocorrelation statistics.

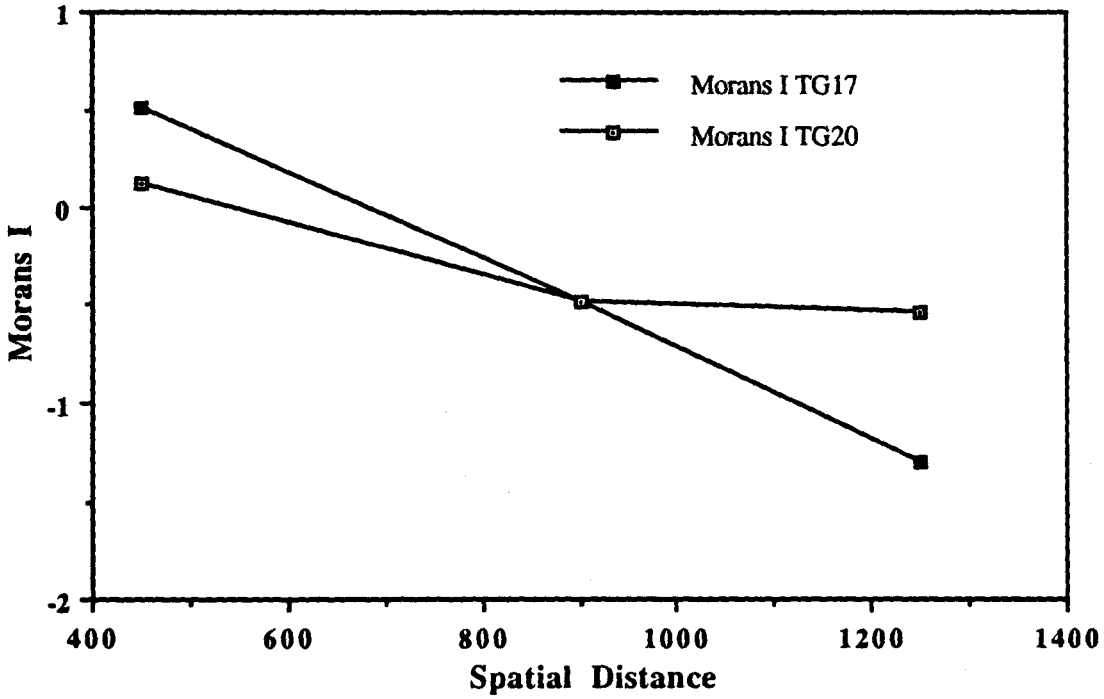

Morans I TG17
Morans I TG20
Morans I
1
0
-1
-2
400
600
800
1000
1200
1400
Spatial Distance

Table 3.3.4

Genetic Distance Matrix

	Tampa Bay	Cartersville	Louisville	Windsor	Ottawa
Tampa Bay		0.0078	0.1112	0.0422	0.0951
Cartersville			0.0630	0.0200	0.0596
Louisville				0.0295	0.0279
Windsor					0.0109
Ottawa					

Spatial Distance Matrix (miles)

Tampa Bay	420	790	1130	1250
Cartersville		370	700	900
Louisville			330	760
Windsor				430
Ottawa				

Latitudinal Distance Matrix

Tampa Bay	6.00	9.75	14.00	17.00
Cartersville		3.75	8.00	11.00
Louisville			4.25	7.25
Windsor				3.00
Ottawa				

under the hypothesis of no association between genetic distance and geographical distance and, genetic distance and latitudinal distance, through a Monte Carlo approach. This approach was taken because in order to obtain the null distribution of Z , $n!$ permutations of our data would have to be carried out. Because $10!$ yields a very large number of permutations, a Monte Carlo approach was more feasible in which the randomization scheme was simulated a sufficient number of times in order to obtain an empirical distribution of Z adequate for significance testing purposes. Rows and columns of one matrix were repeatedly and randomly permutated while keeping the other matrix constant and each time recalculating the test statistic. The algorithm designed for this purpose was again written in Hypercard 6.1 for the Macintosh and is also available on request (see Appendix C). In this case, genetic distance was kept constant and the null distribution thus obtained was used to evaluate the significance of the observed Z values. The Z-value for the product of the genetic distance matrix and geographical distance matrix, $Z = 350.654$ was compared to the null distribution of Z-values obtained through 2000 permutations. Similarly, The Z-value of the product of the genetic distance matrix by the latitudinal distance matrix, $Z = 4.751$ was compared to its null distribution from 2000 permutations. The results indicate that the product of the genetic distance by geographical distance matrix was not significant ($p = 0.087$) whereas the product of the genetic distance matrix by latitudinal distance matrix was significant ($p = 0.042$). Therefore, it seems that genetic distances between localities correlate significantly, or are associated, with latitude but not geographical distance indicating that the north-south structuring is climatic and not historical.

3.3.5 Mantel Tests on Individual Thr-Gly Allele Frequencies

Nei's genetic distance takes into consideration all alleles at a given locus. In order to assess the relationship between individual Thr-Gly alleles and gegraphical and latitudinal distances, Mantel tests of matrix association were also conducted using inter-population allele frequency differences. The results were then compared to the null distribution obtained from 2000 permutations. The allele frequency difference matrix for the TG17 allele by geographical distance matrix yielded a $Z = 1510.1$ which was not significant, $p < 0.1$ and that against latitudinal distance, a $Z = 20.30$ which was significant ($p = 0.037$). The results for the allele frequency difference matrix for TG20 by geographical distance was not significant ($Z = 823.88$, $p < 0.3$) nor against latitudinal distance ($Z = 11.22$, $p < 0.15$). Thus, even at the level of individual allele frequency differences, the TG17 allele appears to be associated with latitudinal distance but not geograhical distance indicating that some spatial structuring does exist and that this is climate or environmentally driven and not due to historical factors. Moreover, the pattern seems to be indicative of a monotonic clinal change in allele frequency for the TG17 allele. A summary of all the preceding tests performed with pertinent statistics is presented in Table 3.3.5.

3.4 Analysis of Allozyme Frequency Data

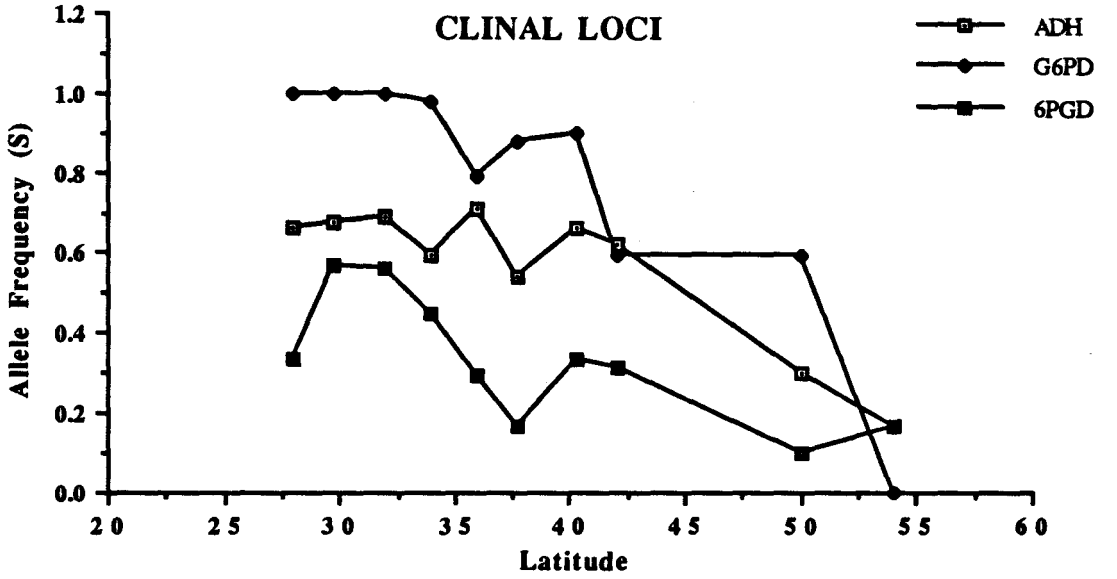
Similar tests were conducted on previously accumulated allozyme data in order to ascertain the degree and pattern of spatial structuring for a different type of molecular marker. The allozyme frequency data were compiled by Ozgur Huner during the summer of 1990 and the North American populations from which the data were obtained are shown in Figure 2.1. The seven allozymes studied were *Adh, G6-pd, Adh, Odh, Pgm, 6-Pgd,*

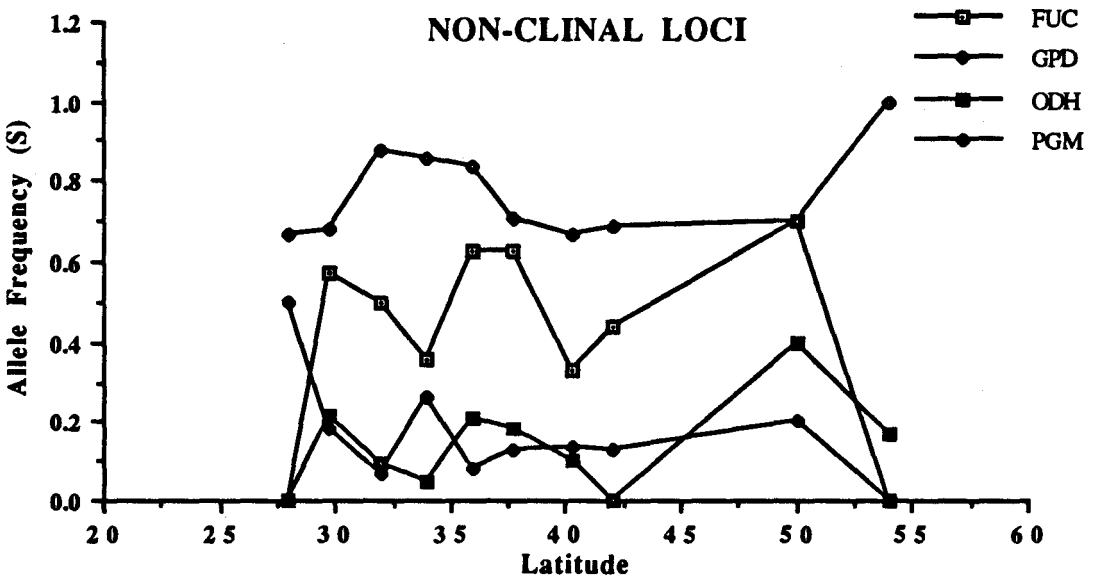
Table 3.3.5
Summary of Statistical Analyses Performed on Thr-Gly Repeat Data

G-test of Allele Frequency Heterogeneity		
G-value = 11.66 (5 x 2 contingency)	Departure From Randomness	$p < 0.025$
Royaltey-Astrachan-Sokal Test for Departure From Randomness		
TG17	L = 1.75	$p < 0.15$
TG20	L = 1.75	$p < 0.15$
Spatial Autocorrelation Analysis		
TG17 *I450* = 0.5013	Positive Association at Short Distances	$p < 0.06$
TG17 *I1250* = -1.297	Negative Association at Large Distances	$p < 0.05$
TG20 *I450* = 0.1253	No Significant Association	$p < 0.30$
TG20 *I1250* = -0.5410	No Significant Association	$p < 0.20$
Regression Analysis		
Genetic on Spatial Distance	b = 0.0000513	$p < 0.2$
Genetic on Latitudinal Distance	b = 0.00437	$p < 0.08$
Mantel Tests of Matrix Association		
Genetic x Spatial Distance	Z = 350.65	$p < 0.1$
Genetic x Latitudinal Distance	Z = 4.75	$p < 0.05$
TG17 x Spatial Distance	Z = 1510.1	$p < 0.1$
TG17 x Latitudinal Distance	Z = 20.30	$p < 0.05$
TG20 x Spatial Distance	Z = 823.88	$p < 0.3$
TG20 x Latitudinal Distance	Z = 11.22	$p < 0.15$

Exact probabilties where applicable are presented in the text

Figure 3.4: Frequency plot for allozymes along latitudinal transect surveyed. Top panel shows those loci displaying a strongly clinal pattern while bottom panel shows those loci displaying a non-clinal pattern.


CLINAL LOCI
ADH
G6PD
6PGD
Allele Frequency (S)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
20
25
30
35
40
45
50
55
60
Latitude


NON-CLINAL LOCI
FUC
GPD
ODH
PGM
Allele Frequency (S)
1.2
1.0
0.8
0.6
0.4
0.2
0.0
20
25
30
35
40
45
50
55
60
Latitude

Fuc and *Gpd* and their frequency distributions as a function of latitude are presented in graphical form in Figure 3.4. It is clear from the figure that two classes of loci exist, those showing geographic structure (clinal pattern) and those not showing structure (non-clinal pattern). The clinally varying loci include *Adh, G6-pd* and *6-Pgd* with the remaining four loci not elliciting a discernable pattern.

3.5 Statistical Analysis of Allozyme Frequencies

3.5.1 Preliminary Test for Departure From Randomness

As for the Thr-Gly repeat frequency data, G-tests were conducted on all allozymes to assess, as a preliminary step, significant departures from randomness. The results of the G-test were not surprising, in that four of the seven allozymes showed significant levels of departure. These loci were *Fuc* ($27.14_{9\ \text{d.f.}}$, $p < 0.005$), *G6-pd* ($61.18_{9\ \text{d.f.}}$, $p < 0.001$), *6-Pgd* ($18.78_{9\ \text{d.f.}}$, $p < 0.05$) and *Odh* ($17.37_{9\ \text{d.f.}}$, $p < 0.05$).

The Royaltey-Astrachan-Sokal test of departure from random geographic variation was also applied to the localities along the transect. The 10 localities were connected by means of a simple chain and the frequencies of each allele for a specific allozyme were ranked among localities (Table 3.5.1). When mean edge length was determined, it was found that only two allozymes, *G6-pd* and *6-Pgd* gave significant mean edge lengths ($L = 1.186$, $p < 0.005$; $L = 2.001$, $p < 0.005$ respectively). All other allozymes failed to give significant mean edge lengths and therefore did not represent departures from random geographic variation at the level of resolution afforded by this test. Therefore, for the two allozymes giving significant values, a definite pattern of geographic structuring exists. Whether or not this pattern is clinal requires further testing.

Table 3.5.1
Royaltey-Astrachan-Sokal and G-Tests for Departure From Randomness

Allele	Mean Edge Length	*p*
Adh	2.722	ns
Fuc	3.111	ns
G6-pd	1.186	< 0.005
Gpd	3.555	ns
Odh	4.167	ns
6-Pgd	2.000	< 0.005
Pgm	2.611	ns

Allele	G-Test	*p*
Adh	11.84	ns
Fuc	27.14	< 0.005
G6-pd	61.18	< 0.001
Gpd	13.21	ns
Odh	17.37	< 0.05
6-Pgd	18.78	< 0.05
Pgm	12.71	ns

3.5.2 Spatial Autocorrelation Analysis on Allozymes

For spatial autocorrelation analysis, because of the larger latitudinal transect surveyed for the allozymes, the data were partitioned into five distance classes - 450, 900, 1250, 1700 and 2150 miles each consisting of any where from 32 to 10 comparisons. Moran's *I* statistic was then computed for each pair of neighbours within a specified distance class and because the number of populations (n = 10) was larger than that employed for the TG frequency data, significance of each autocorrelation statistic was computed without the need of random permutations. In the analysis of spatial patterns, all loci except *Fuc* showed significant autocorrelation over some distance classes. Table 3.5.2 shows the individual autocorrelation statistics at each distance class for each of the allozyme surveyed.

Adh, G6-pd and *6-Pgd* all show significant positive autocorrelation at short distances (I_{450} = 0.4868, $p < 0.001$; 0.3049, $p < 0.01$; 0.2826, $p < 0.05$ respectively) and significantly negative autocorrelation at far distances (I_{1700} = -0.8856, $p < 0.05$; -0.8708, $p < 0.05$; -0.9866, $p < 0.05$ respectively). All other allozymes except for *Fuc* show significant autocorrelation at distances between 1701 and 2150 miles. However, the small number of comparisons made between neighbours within this distance class makes the value of I_{1750} a more reliable estimate of the degree of negative autocorrelation at farther distances. Thus, at the level of spatial analysis, three of the seven allozymes, *Adh, G6-pd* and *6-Pgd* show varying clinal patterns with *Adh and G6-pd* showing clinal patterns that are strongly monotonic based on their correlograms (Figure 3.5.2). *6-Pgd* shows directional positive to negative autocorrelation indicative of a clinal pattern yet the correlogram indicates that this change is not monotonic. However, the results do indicate that significant spatial structuring of this allele exist along our transect. The correlogram

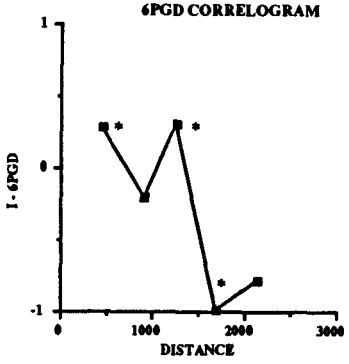
Table 3.5.2
Autocorrelation Coefficients (Morans *I*) Across Distance Classes

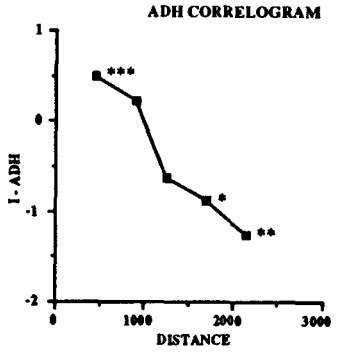
Allele	I_{450}	I_{900}	I_{1250}	I_{1700}	I_{2150}
Adh	**0.4868**	0.2222	-0.6344	**-0.8856**	**-1.2529**
Fuc	-0.0665	-0.0844	0.1378	-0.2027	-0.0029
G6-pd	**0.3049**	0.0444	-0.0225	**-0.8708**	**-1.4729**
Gpd	-0.0074	-0.2714	-0.0659	0.1873	**-0.9128**
Odh	-0.0065	-0.0235	-0.0454	-0.2002	**-0.9709**
6-Pgd	**0.2826**	-0.2077	**0.0322**	**-0.9866**	-0.7794
Pgm	-0.1515	-0.0760	-0.2629	0.4745	**-1.0288**

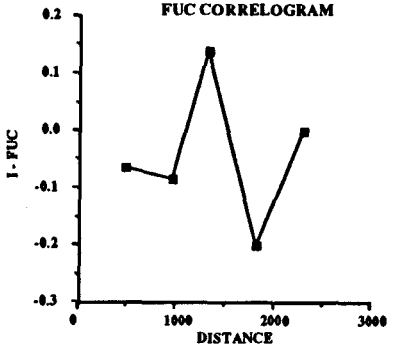
Values in **bold** represent significant autocorrelation coefficients. Refer to Figure 3.5.2 for individual significance levels.

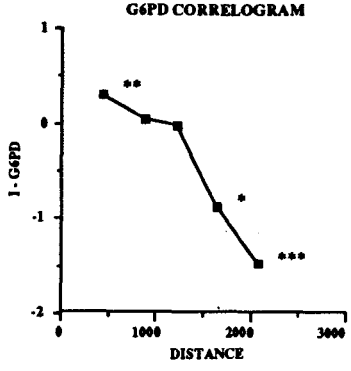
Figure 3.5.2: Correlogram surfaces for individual allozymes. Significance levels for individual autocorrelation statistics are provided where applicable.

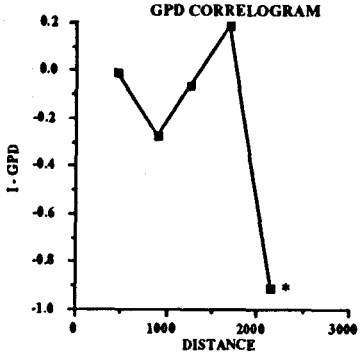
Summary of Correlograms with Significance Levels of Autocorrelation Statistics

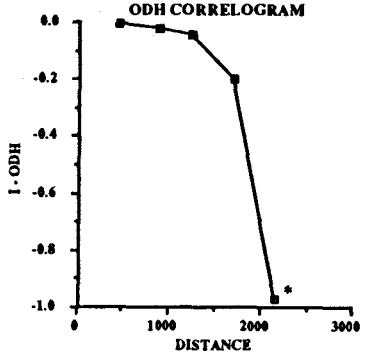


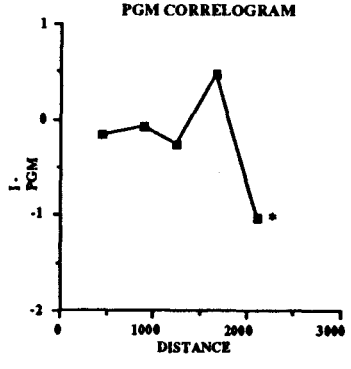












* $0.01 < P \leq 0.05$
** $0.001 < P \leq 0.01$
*** $P < 0.001$

for *Odh* is interesting insofar as it depicts a somewhat monotonic change. However, the correlogram only steeps towards greater distances and otherwise has an almost flat slope for distances up to 1250 miles and is therefore, not significantly indicative of a clinal pattern. All other correlograms are "patchy" and exhibit patterns that more closely resemble the depressions, intrusions and crazy quilts described in Royaltey *et al.*, (1975) and Sokal and Oden (1978). It is interesting to note that *Adh*, which did not show a departure from randomness in allele frequency using the G-test nor the Royaltey-Astrachan-Sokal non-parametric test, seems to exhibit based on spatial autocorrelation analysis, a clinal pattern in its distribution along our transect. Using the Bonferroni criteria to adjust significance levels, none of the correlograms as a whole is significant despite significant spatial structuring and clinal patterns existing for several loci.

3.5.3 Allele Frequency Variation Estimated by F_{ST}

The organization of genetic variation was further examined via *F*-statistics (Wright, 1951). The results for weighted allele frequencies are presented in Table 3.5.3 along with individual allele frequencies and indicate that the extent of genic differentiation among the populations varies considerably but is quite high for several loci. This is not surprising since several of the loci were specifically chosen because they have been reported as varying clinally in past literature. Furthermore, the variance in F_{ST}, although high, is not as high as for loci chosen at random as demonstrated by Singh and Rhomberg (1987a). The variance in frequency at the *G6-pd* locus among the populations ($F_{ST} = 0.440$) is the highest and that at the *Adh* locus ($F_{ST} = 0.187$) is the lowest. It follows then, that the clinal pattern observed for *G6-pd* is steeper as is also evident from its correlogram.

Table 3.5.3

Allele Frequencies[1] and F_{ST}

	Adh	*Fuc*	*G6-pd*	*Gpd*	*Odh*	*6-Pgd*	*Pgm*
Tampa Bay (6)	0.667	0.000	1.000	0.500	0.000	0.333	0.667
High Springs (28)	0.679	0.571	1.000	0.179	0.214	0.571	0.679
Cordele (32)	0.688	0.500	1.000	0.067	0.094	0.563	0.875
Cartersville (42)	0.595	0.357	0.976	0.262	0.048	0.450	0.857
Nashville (24)	0.708	0.625	0.792	0.083	0.208	0.292	0.833
Louisville (12)	0.542	0.625	0.875	0.125	0.182	0.167	0.708
Dayton (30)	0.667	0.333	0.900	0.133	0.100	0.333	0.667
Windsor (16)	0.625	0.438	0.600	0.125	0.000	0.313	0.688
Winnipeg (10)	0.300	0.700	0.600	0.200	0.400	0.100	0.700
The Pas (6)	0.167	0.000	0.000	0.000	0.167	0.167	1.000
F_{ST}	0.187	0.319	0.440	0.292	0.291	0.265	0.199

[1] Frequencies of the faster migrating alleles are given with the exception of *6-Pgd* and *Pgm*, where middle allele frequency is given. Numbers in parenthesis represent sample sizes for a given population.

3.5.4 Linear Regression Analysis on Distance Matrices

Genetic distance (Nei, 1972), latitudinal distance and geographic distance matrices were computed for the allozyme frequency data along the transect (Table 3.5.4). Regression analysis of genetic distance by geographic distance and by latitudinal distance yielded highly significant regression coefficients. For geographic distance, $b = 0.00017$, $p << 0.0001$ and for latitudinal distance, $b = 0.0131$, $p < 0.0001$ (Figure 3.5.4 a & b). Moreover, the Pearson product-moment correlation coefficient for geographic distance is $R^2 = 0.8083$ and for latitudinal distance is $R^2 = 0.8149$ indicating the strength of the association between genetic distance and both latitude and geographical distance.

3.5.5 Mantel Tests on Allozyme Frequency Matrices

The Mantel tests of matrix correlation revealed trends that were not surprising and further serve to confirm the existence of clinal patterns for several of the allozymes. The *Adh, G6-pd* and *6-Pgd* allele frequencies correlate significantly with both geographic distances between populations and are also highly significantly associated with latitude. None of the allele frequencies are associated with longitude (data not shown) indicating that the major directional component of the observed clinal pattern is north-south. Of the allozymes mentioned above, *Adh, G6-pd* and *6-Pgd* showed highly positive associations ($p << 0.001$). The Z values for the Frequency X Geographic Distance matrix are $Z_{Adh} = 9659.9$, $Z_{G6\text{-}pd} = 16303$ and $Z_{6\text{-}Pgd} = 7049.7$ and for the Frequency X Latitudinal Distance matrix are $Z_{Adh} = 121.73$, $Z_{G6\text{-}pd} = 212.35$ and $Z_{6\text{-}Pgd} = 55.525$. This result is not surprising considering that same three allozymes, *Adh, G6-pd* and *6-Pgd* showed significant spatial autocorrelation and a clinal pattern in their underlying spatial distribution. Tests of association between observed allele frequency and climatic parameters to ellucidate

Table 3.5.4

Genetic Distance Matrix

	Tam	HiSp	Cord	Cart	Nash	Lou	Day	Wind	Winn	Pas
Tam		0.122	0.114	0.054	0.153	0.136	0.060	0.106	0.272	0.344
HiSp			0.015	0.038	0.044	0.058	0.041	0.077	0.171	0.448
Cord				0.017	0.029	0.051	0.026	0.057	0.160	0.355
Cart					0.044	0.051	0.015	0.049	0.145	0.297
Nash						0.016	0.029	0.034	0.073	0.280
Lou							0.033	0.039	0.050	0.288
Day								0.027	0.122	0.267
Wind									0.098	0.174
Winn										0.205
Pas										

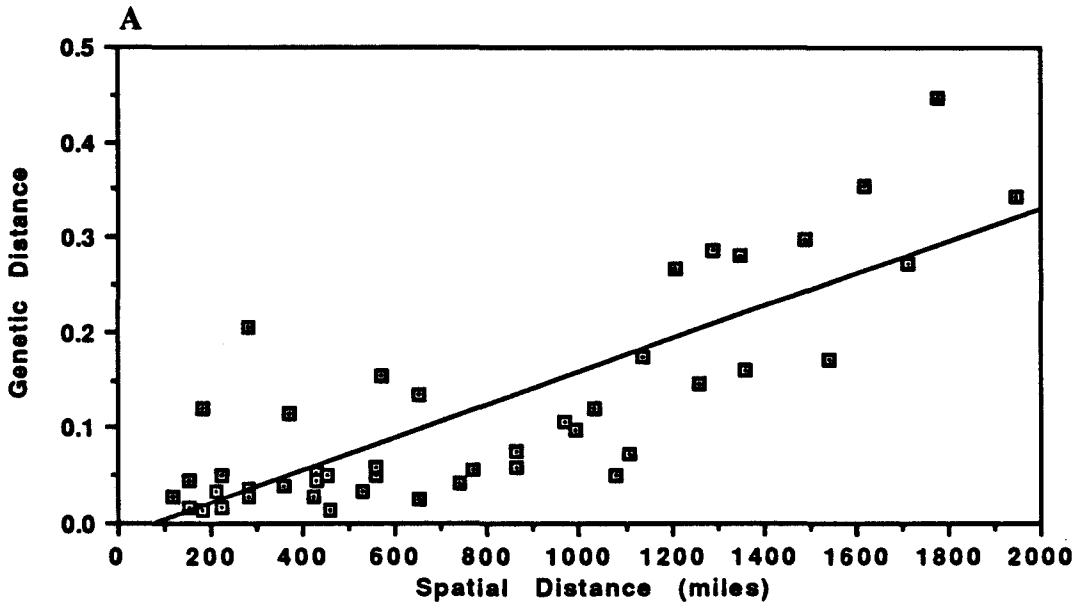
Spatial Distance Matrix

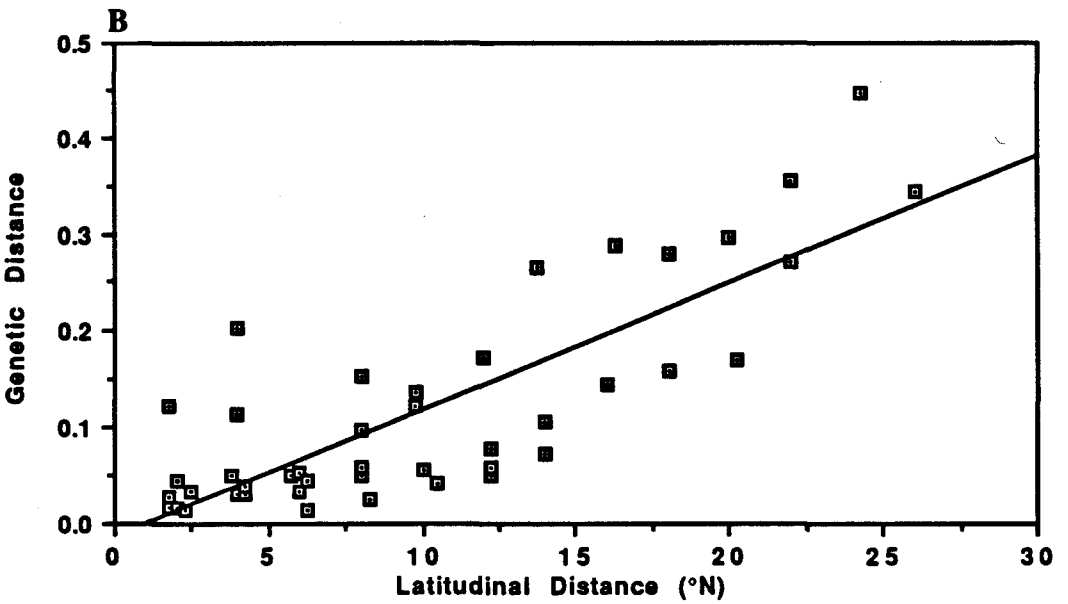
Tam	180	370	420	570	790	1010	1130	1970	2150
HiSp		190	240	390	610	830	950	1790	2070
Cord			150	300	520	730	850	1690	1970
Cart				150	370	580	700	1540	1820
Nash					220	430	550	1390	1670
Lou						210	330	1170	1450
Day							120	960	1240
Wind								840	1120
Winn									280
Pas									

Latitudinal Distance Matrix

Tam	1.75	4.00	6.00	8.00	9.75	12.25	14.00	22.00	26.00
HiSp		2.25	4.25	6.25	8.00	10.50	12.25	20.25	24.25
Cord			2.00	4.00	5.75	8.25	10.00	18.00	22.00
Cart				2.00	3.75	6.25	8.00	16.00	20.00
Nash					1.75	4.25	6.00	14.00	18.00
Lou						2.50	4.25	12.25	16.25
Day							1.75	9.75	13.75
Wind								8.00	12.00
Winn									4.00
Pas									

Figure 3.5.4 a & b: Linear regression analysis of genetic distance by spatial distance and genetic distance by latitudinal distance for allozyme loci.


A
Genetic Distance
0.0
0.1
0.2
0.3
0.4
0.5
0
200
400
600
800
1000
1200
1400
1600
1800
2000
Spatial Distance (miles)


B
Genetic Distance
0.0
0.1
0.2
0.3
0.4
0.5
0
5
10
15
20
25
30
Latitudinal Distance (°N)

factors that could possibly acount for the observed gradient were not conducted but would prove insightful considering the strength of the association between frequencies and latitudinal and geographic distance as well as the strong autocorrelation observed for several of the allozymes. A summary of the results of the Mantel tests is presented in Table 3.5.5 and a summary of the results of all preceding statistical analyses conducted on the allozyme data is presented in Table 3.5.5.1

3.6 Variation in Primary Structure of Thr-Gly Alleles

As was mentioned previously, several Thr-Gly alleles were sequenced in order to ascertain repeat size class and to determine whether the sequences were different from those published by Citri *et al.*, (1987) and Costa *et al.*, (1991). By far, the most frequently occuring variants, based on the Southern work, were the TG17 allele followed by the TG20 variant. Both these variants were sequenced in five lines, one from each of the five populations. None of the sequences obtained for these two variants differed relative to the published sequences (see Appendix B). Sequence analysis of the rare variant showing the mobility shift indicated that this modal class contained 22 Threonine-Glycine pairs (TG22). Interestingly, the sequence of this variant was markedly different from that of the published sequences of Citri *et al.*, (1987) and those of Costa *et al.*, (1991). Since this variant was novel, there was no published sequence of the same length to compare it to. However, certain differences were immediately obvious. The most obvious difference was in the region containing the 18bp cassette motif which in the TG22 variant occured only once as opposed to twice as would be expected assuming it was derived from either the TG20 or TG23 variants described in Yu *et al.*, (1987). The second difference was in the nucleotides encoding the threonine and glycine residues. Although the alternating pattern of amino

Table 3.5.5

Mantel Tests for Association Between Allele Frequency and Geographical and Latitudinal Distance Matrices

Allele	Dissimilarity Matrix	
	Geographic Distance	Latitudinal Distance
Adh	**9659.9**	**121.73**
Fuc	-8930.5	-114.46
G6-pd	**16303**	**212.35**
Gpd	4071.5	55.525
Odh	-3503.2	-41.841
6-Pgd	**7049.7**	**92.046**
Pgm	-2317.4	-29.816

Values in **bold** represent significant *Z* values based on a Monte Carlo randomization approach.
Values are not transformed and represent sums of product matrices as per the text.

Table 3.5.5.1
Summary of Statistical Analyses Performed on Allozyme Data

G-Test of Allele Frequency Heterogeneity		
Adh	G = 11.84	ns
Fuc	G = 27.14	p < 0.005
G6-pd	G = 61.18	p < 0.001
Gpd	G = 13.21	ns
Odh	G = 17.37	p < 0.05
6-Pgd	G = 18.78	p < 0.05
Pgm	G = 12.71	ns
Royaltey-Astrachan-Sokal Test of Departure From Randomness		
Adh	L = 2.722	ns
Fuc	L = 3.111	ns
G6-pd	L = 1.186	p < 0.005
Gpd	L = 3.555	ns
Odh	L = 4.167	ns
6-Pgd	L = 2.000	p < 0.005
Pgm	L = 2.611	ns
Spatial Autocorrelation Analysis		
Adh I450 = 0.4868	Positive Association at Short Distances	p < 0.001
Adh I2150= -1.2529	Negative Association at Large Distances	p < 0.01
G6-pd I450 = 0.3049	Positive Association at Short Distances	p < 0.01
G6-pd I2150= -1.4729	Negative Association at Large Distances	p < 0.001
6-Pgd I450 = 0.2826	Positive Association at Short Distances	p < 0.05
6-Pgd I1700 = -0.9866	Negative Association at Large Distances	p < 0.05
All other loci, except *Fuc*, showed only significant associations at large distances		
Interlocality Differentiation (F_{ST})		
Adh	*FST* = 0.187	relatively low
Fuc	*FST* = 0.319	relatively high
G6-pd	*FST* = 0.440	relatively high
Gpd	*FST* = 0.292	moderate
Odh	*FST* = 0.291	moderate
6-Pgd	*FST* = 0.265	moderate
Pgm	*FST* = 0.199	relatively low
Regression Analysis		
Genetic on Spatial Distance	b = 0.0002	p << 0.0001
Genetic on Latitudinal Distance	b = 0.0131	p < 0.0001
Mantel Tests of Matrix Association		
Adh x Spatial Distance	Z = 9659.9	p < 0.001
Adh x Latitudinal Distance	Z = 121.73	p < 0.001
G6-pd x Spatial Distance	Z = 16303	p < 0.001
G6-pd x Latitudinal Distance	Z = 212.35	p < 0.001
6-Pgd x Spatial Distance	Z = 7049.7	p < 0.001
6-Pgd x Latitudinal Distance	Z = 92.046	p < 0.001
All other loci did not show significant associations with either parameter		

acids was maintained, numerous synonymous substitutions were present. Furthermore, several synonymous substitutions occured in the sequences flanking the repeat on both 5' and 3' sides (Appendix B). This was confirmed in a total of four alleles that were sequenced - one from each of the populations in which this variant appeared. As mentioned, this variant was extremely rare, occuring only in 8 of 126 lines sampled.

The TG23 allele, present in several of the lines, was also different from that predicted by the published sequence of Citri *et al.*, (1987). The 5 alleles sequenced were more like the TG23 variants sequenced by Costa *et al.*, (1991) in the pattern of nucleotide substitutions (Appendix B). Moreover, careful examination of the pattern of substitutions seemed to indicate that it was also very similar to the sequence of the previously discussed TG22 allele particularly with respect to the obliteration of the 18bp cassette motif. Based on the primary structure of this region, it is possible that the TG22 variant may have arisen via a series of slippage/deletion events from the ancestral TG23 variant. Its rarity and the similarity of its sequence to that of our TG23 variant lends support to this hypothesis. Moreover, the sequence of the TG23 variant reported here supports the hypothesis of Costa *et al.*, (1991) and Peixoto *et al.*, (1993) that different Thr-Gly variants of the same length may have arisen through distinct seperate lineages.

Chapter 4

DISCUSSION

4.1 High Mutability of the Repeat Encoding Motif

As an aside to the Southern analysis of the Thr-Gly region of *per*, sequencing of several Thr-Gly alleles was also conducted and further work is currently under way. The Thr-Gly encoding motif appears to behave very much like non-coding minisatellites which are under high mutational pressure. The disparity between sequences published by Citri *et al.*, (1987) and those by Costa *et al.*, (1991) plus the fact that new variants are still being uncovered as in the present study, serves to illustrate this point. The present observation of a new length variant whose sequence varies at more than 20 synonymous sites from that of the published sequences, suggests the complex evolutionary history of this region. Furthermore, it is clear that even Thr-Gly variants of the same length can arise from quite different mutational processes as is the case for the TG23 variant. The different origins of such isolength Thr-Gly alleles has important implications for the discussion of whether selection is acting on the Thr-Gly region itself or on adjacent sequences that are in linkage disequilibrium. If similar sized variants can be derived from different lineages as has been hypothesized by Costa *et al.*, (1991) and Peixoto *et al.*, (1993), then they would not be expected to share the same flanking sequences. Further sequencing of naturally occuring variants as well as regions flanking the repeat would certainly provide insight into the role of selection in this region of *period.* It may be that similar sized repeat units make lineages that are not related (by virtue of differences in flanking sequences), appear as though they were infact related. To date, no data on sequences immediately flanking the repeat region have been published and this is a critical next step. Sequence data on introns would enable

us to select, *a priori,* a neutral gene region and facilitate comparisons of differences in patterns of substitutions in coding and non-coding sequences between lineages.

4.1.1 Is the Motif Encoding the Repeat Under Selection?

There are some lines of "evidence" which suggest that selection, if it is acting on *per*, is acting on the repeat itself. Could the length polymorphism therefore have adaptive significance? A recent clue comes from the work of Ewer *et al.*, (1990). In their study, the Thr-Gly encoding motif was removed from the *per* gene and the mutated construct was transduced into per^{01} arrhythmic mutants. They observed that the rescue of the arrhythmic locomoter phenotype was strong at 25 °C but significantly weaker at 29 °C suggesting that the Thr-Gly repeat may be important for the thermostablity of the circadian phenotype. The ideal absence of such temperature sensitivity is a cardinal feature of a true biological clock and is termed temperature compensation (Pittendrigh, 1954). Consequently, perhaps different length variants have altered temperature compensation properties when challenged with different temperature regimes. That is to say, perhaps different Thr-Gly alleles confer on their carriers different adaptive circadian responses to latitudinal temperature variation. This, coupled with the observation of a similar clinal pattern in frequency distribution of the TG17 allele on separate continents, makes the role of selection in maintaining the gradient a very tempting hypothesis. As an aside, it is interesting to note that of the allozymes surveyed, *G6-pd* was the only one demonstrating a significant departure from randomness, showed highly significant directional spatial autocorrelation, had correlogram surfaces that were monotonic and lastly, showed strong significant associations with latitude. The *G6-pd* gene *(Zw)* is located on the X chromosome of *Drosophila* and it is a somewhat strange coincidence that the *period* gene is also sex-linked. It is tempting to hypothesize,

therefore, that some form of selection is operating on both of these sex-linked loci, the result of which is the underlying clinal pattern in their frequency distributions.

4.2 Mechanisms Generating Allele Frequency Gradients

It has been suggested that Thr-Gly allele frequency variation is structured along an approximately north-south axis in European populations of *D. melanogaster* (Costa *et al.*, 1992). In the present study, use of several statistical tests designed to assess spatial distributions of allele frequencies have shown that the same trend exists in North American populations. Specifically, the decline of spatial autocorrelation with distance for the TG17 allele and positive Mantel correlations observed between allele frequency differences and latitude serve to validate the claim that the most commonly occuring of the Thr-Gly alleles is spatially structured in North America as well.

According to Costa *et al.*, (1992) at least three processes may determine allele frequency gradients on a large scale. The first is demic diffusion as has been demonstrated in humans (Menozzi *et al.*, 1978). In this case, a cline can persist as a result of population growth, dispersal and subsequent intermixing of formerly disjunct, isolated and differentiated populations. However for North American *D. melanogaster*, because it is generally believed that they have been introduced from multiple sources and have spread relatively rapidly, this process does not seem tenable. Another process that can potentially account for geographic structure is range expansion. If new territories are colonized by a few individuals from distinct geographic regions, founder effects may occur. Random genetic differences established as a consequence may persist and lead to clines in the presence of successive gene flow (Costa *et. al.*, 1992). This type of explanation has been proposed by Esteal (1985) to account for the clines observed in vertebrates such as the

giant toad *Bufo marinus*. However, this type of cline can result by chance and is not expected to be repeated between different localities. By far, the most traditional explanation proposed for the existence of latitudinal clines is an adaptive response to climatic variation (Huxley, 1938; Endler, 1977). Climate related selection, for example, is thought to be responsible for the latitudinal clines at the *alcohol dehydrogenase* and *esterase-6* loci in *D. melanogaster* (Oakeshott *et al.*, 1982; Anderson and Oakeshott, 1984). Of course, for such equilibrium clines of *Drosophila* to be adaptive, one has to assume that allele frequencies at these loci are tracking an environmental gradient of some sort that also varies in a north-south direction (Singh and Rhomberg, 1987a). However, in the present study, no environmental parameters were used in the matrix comparisons for either the Thr-Gly data or the allozyme data so no hypothesis can be put forward regarding the role of environmental parameters acting to maintain the observed gradients.

In order to make the statement that certain loci vary clinally and that this pattern is due to climate related selection for example, it is important to look at those loci that do not vary clinally (*i.e.* those that are presumably neutral). This is because even those loci that vary clinally must be under a balance of forces involving both selection and gene flow. The pattern of variation of neutral loci can therefore be used as a filter to understand the nature of the strength of selection. This is why in the present study, both presumably neutral and non-neutral molecular markers were used. For the *period* gene, as was discussed, a neutral component remains to be surveyed. For the allozymes however, both selected and neutral loci were surveyed, the choice of allozymes being determined *a priori* based on the results of previous work (Singh and Rhomberg, 1987a).

4.2.1 Inference of Drift and Gene Flow From Autocorrelation Statistics

Theory and computer simulations (Sokal and Wartenberg, 1983; Barbujani, 1987; Epperson, 1993) have shown that the interaction of genetic drift and individual dispersal (isolation by distance) can cause significant genetic resemblance between populations that are close in distance but no overall large scale pattern over larger distances. Beyond this small scale, the autocorrelation coefficient approximates zero, indicative of no significant spatial structuring. It follows then, that pure isolation by distance may therefore result in an asymptotic decline in Moran's *I*, but only over very limited distances. Consequently, genetic drift and dispersal can, and presumably do, contribute to the positive short range autocorrelation observed for the TG17 allele, but they cannot account for the significant negative autocorrelation observed between those localities separated by more than 1250 miles. The correlogram for the TG17 allele is clearly monotonic and based on its surface, is indicative of a clinal pattern. Because of the small number of populations surveyed, the same claim cannot be made for the TG20 allele, despite a monotonic correlogram surface. In this case, the individual autocorrelation statistics are not significant based on the result of randomization tests.

The results of the various statistical tests conducted on the allozymes indicate a complex pattern in allele frequency distribution with several loci showing a clearly clinal pattern. Based on the spatial autocorrelation analysis, four of the seven allozymes surveyed either lack or have uninterpretable spatial patterns. The spatial variation in these allele frequencies could result from genetic drift not counteracted by migration or selection. However, genetic drift alone, resulting from small effective sizes of populations sampled, is not a tenable interpretation for all alleles because at least three of the loci, *Adh*, *G6-pd* and *6-Pgd*, show continent wide structural patterns based on their correlograms and have

also been reported as varying clinally in the literature. If drift were operating, we would expect all loci to be subject to it, even under moderately strong selection (Barker and Mulley, 1976; Barbujani, 1987; Sokal *et al.*, 1987). Two of the correlograms appear to be very similar, namely *G6-pd* and *Adh* and to the extent that this a real phenomenon, suggests the possibility that these two loci might be tracking the same environmental gradient. On the other hand, the two loci, despite appearing similar in one dimensional correlograms, may in fact be heading in different compass directions and therefore be tracking different environmental gradients.

4.3 Difficulties in Interpreting Causation for Observed Patterns

Based on the Thr-Gly allele frequency data alone, it is impossible to judge whether or not the approximately clinal pattern observed is due to selection as this would require surveying regions outside the repeat encoding motif. Moreover, it has been shown that spatially varying selection and gene flow such as that occuring during range expansion processes do not necessarily result in dissimilar geographical patterns of allele frequencies (Slatkin and Maruyama, 1975). The fact that only a small number of populations were sampled has repercussions even at the level of determining whether the frequencies significantly depart from randomness as was evident in the Royaltey-Astrachan-Sokal test. The test is extremely sensitive to departures from a a perfectly monotonic change in allele frequency as was shown for both the TG17 and TG20 variants. It is also more sensitive to "fringe" allele frequencies occuring in populations occupying middle positions in simple chain graphs as opposed to such frequencies towards the termini of such graphs as was illustrated in the hypothetical examples. Sampling a greater number of populations, even those separated by smaller distances, would have greatly improved the resolving power of

this and other tests. The analysis of the allozyme data serves to illustrate this. Both *G6-pd* and *6-Pgd* showed significant departures from randomness based on the Royaltey-Astrachan-Sokal test. This, however, was expected due to the fact that both these loci were part of group of loci already known to exhibit spatial structuring (Singh and Rhomberg, 1987a). Unexpectedly though, *Adh*, which showed significant directional autocorrelation indicative of a clinal pattern, did not give a significant mean edge length once again illustrating the weakness of this test to more complex frequency distribution patterns.

For the Thr-Gly data, it is therefore necessary to rely on other lines of evidence to understand if selection may really be operating and whether or not the effects are directed at the Thr-Gly length polymorphism or other closely linked sequences. For the allozyme data, the pattern or lack of pattern observed in allele frequencies may be the result of selection possibly on different scales. That is to say, for those loci previously and presently showing a clinal pattern (*Adh, G6-pd* and *6-Pgd*), selection may be operating on a continental scale. For the remaining loci (*Fuc, Pgm, Gpd* and *Odh*), that differ among localities yet show no clear spatial pattern on the scale assessed in this study, selection may be acting on a strictly local scale or perhaps not at all.

Despite the fact that the clinal pattern observed for the Thr-Gly length polymorphism cannot be undisputably attributed to selective forces, the results of the Mantel tests of matrix correlation clearly show that the Thr-Gly alleles are structured latitudinally. That is, genetic distance, based on all alleles, is significantly correlated with latitude as is the frequency of the TG17 variant. Conversely, neither genetic distance nor individual allele frequency is associated with changes in spatial distance. It is interesting to note that the frequency of the TG20 allele is not associated with latitudinal distance in light of the fact that the correlogram surface shows a monotonic transition from positive to

negative, indicative of a clinal pattern (albeit insignificant for individual autocorrelation statistics). It should be noted however, that a cline may exist even if latitude itself is irrelevant . An appropriately placed "east-west" transect, altitudinal transect or sampling along some environmental gradient could serve to validate the existence of a cline where latitude does not (Singh and Rhomberg, 1987a). Taking this into consideration, Mantel tests using genetic distance matrices and TG17 and TG20 frequency matrices against longitudinal matrices were performed. In all cases, highly insignificant ***Z*** values were found indicating that for the TG17 allele, the major directional component of the cline is north-south and that the TG20 allele is not structured along any directional axes. For the allozyme data, the loci found to exhibit significant spatial autocorrelation, namely *Adh, G6-pd* and *6-Pgd*, also had frequencies that were significantly associated with latidudinal and spatial distance indicating north-south structuring of allele frequencies. Again, this was expected based on the fact that these loci were part of the subset of loci known to vary clinally and to presumably be under selection.

4.4 Interpreting Trends in Molecular and Morphological Data

An indirect goal of this study was to compare the results obtained from the analysis of geographic variation in two types of molecular markers to those obtained from a study of spatial variation in morphological characters. The basis for this stems primarily from an ongoing objective of this lab, namely to compare patterns of variation in molecules and morphology in an attempt to understand the role of selection. Long (1992) studied the same latitudinal transect for 10 different morphometrical characters and found that clinal patterns were non-linear, with larger flies occupying middle latitudes while smaller flies were more prevalent in the north and south. It also appeared as though selection was acting

primarily on head characters in the south and on body characters in the north. Two non-exclusive points are clearly evident from these results that pertain to our objective. First, the pattern of clinal variation observed for certain traits is non-monotonic and second, not all characters show selection, a point exemplified in the allozyme analysis. Long (1992) postulated that for those morphological characters presumed to be under selection, the non-monotonic clinal pattern observed was a consequence of selection for optimal body size. Alternately, the peaks observed in intermediate latitudes may have been the result of the populations response to environmental pressures in order to adapt to a phenotypic optimum that was seasonal or temporal. This result stands in contrast to that obtained for both the nuclear and allozyme data. Specifically, for the TG17 allele and for the allozyme loci showing clinal variation, the pattern was monotonic. In this case, the north-south clinal structuring invokes a climate or environmentally driven explanation to account for the observed pattern. Thus, the observation of different clinal patterns in both molecules and morphology serves to highlight the fact that the different genetic markers are subject to different types of selection.

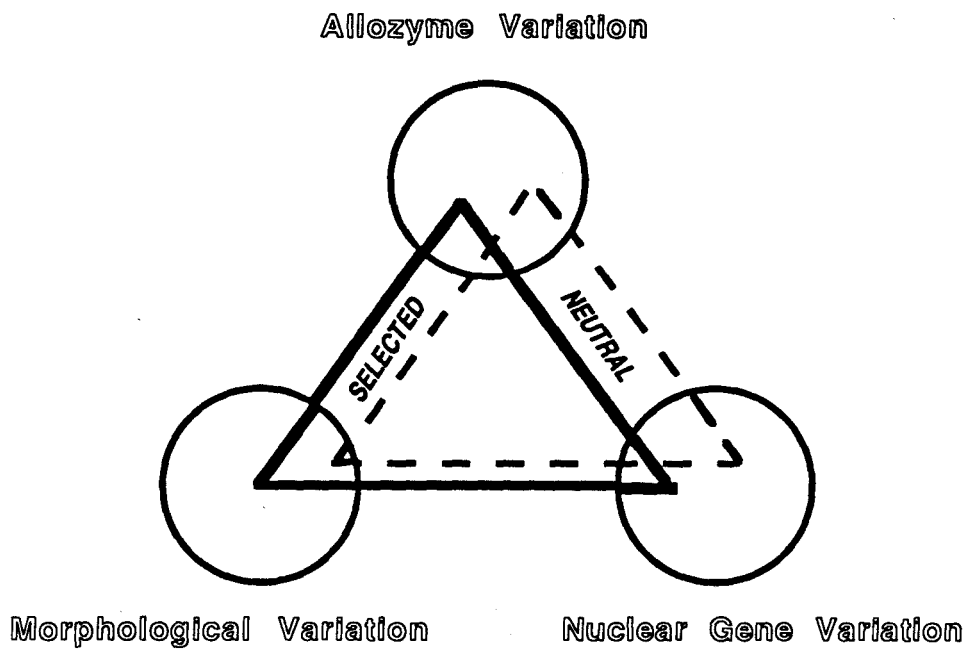
The use of both molecular and morphological markers enables us to understand the causes of geographic variation in greater detail than either does individually. Patterns of variation uncovered using molecular data allow inferences to be made regarding the patterns observed in morphological traits. This is feasible provided that the molecular markers, whether allozymes or gDNA, are neutral because only then can they provide insight into the history of the population in which they are surveyed. This stands in contrast to allozymes such as *G6-pd* and DNA markers such as the Thr-Gly region which are presumably under selection and can therefore only describe the recent history of the same population. It is therefore critical to assess variation in a neutral region of *period*, as this will make the

molecules versus morphology story complete and will enable us to draw inferences regarding the true role of selection in maintaining the similarities and differences in observed patterns of variation in the three types of genetic markers.

4.4.1 The Importance of Historical Factors

Even though the trends of the Long study indicate that selection acts primarily on head characters in the south and more on body size in the north, this sort of picture can give misleading results as characters possibly of potential importance may not have been accounted for. Molecular characterization of the same populations could in principle enable one to determine for example, if the steepness of the observed cline at various positions is due to increased selection differentials or due to constant selection with decreased gene flow as this differentiation cannot be made with phenotypic data alone (Long, 1992). This would be possible assuming the molecular markers used were neutral and therefore enabled one to determine migration rates between populations. While it is known that gene flow is high in *D. melanogaster* (Coyne and Milstead, 1987 and Singh and Rhomberg, 1987b), we need estimates of gene flow in neighbouring populations. This can presumably be obtained from the F_{ST} estimates of neutral molecular loci (Prout and Barker, 1993). Such work is presently underway and should shed greater light on the basis for the difference between latitudinal variation in molecules and morphology.

Figure 4.4: A diagramatic representation of the simultaneous use of molecular and morphological traits to understand the role of selection.


Allozyme Variation
SELECTED
NEUTRAL
Morphological Variation
Nuclear Gene Variation

Chapter 5
SUMMARY

The purpose of the present study was to assess the level and pattern of variation at the *period* gene, in particular, the frequency distribution of the threonine-glycine repeat in latitudinal populations of *Drosophila melanogaster*.. The results indicate that the spatial distribution of the TG17 variant is structured latitudinally in an approximately clinal pattern. The observation of the same pattern in European populations plus the fact that the sequences surrounding and including the motif are implicated in maintaining the thermal stability of the circadian phenotype invoke a selective explanation to account for the frequency distribution.

Analysis of allozyme frequency data was also conducted using similar statistical methods in order to assess geographical patterns. The analysis clearly indicates that several allozymes display clinal patterns in their frequency distributions. It was concluded from the results of the statistics that although drift plays an important role in short distance associations, it cannot account for the differences at large distances, where large reflects distances not capable of being traversed by *Drosophila* during a single lifetime. Instead, it is postulated, based on similarities of correlogram surfaces, that many of the allozymes may infact be tracking similar environmental gradients with selection acting at different spatial scales for the various enzymes.

The observation of rare Thr-Gly size variants prompted sequencing of the region in order to evaluate repeat class size as well as the level of variability in regions immediately flanking the repeat. Although this work is not yet complete and will be the focus of discussion elsewhere, some interesting patterns have emerged. Comparisons of inter-region sequence variability, particularly in the intron regions, should give us information

regarding neutral gene substitution and is currently under way. Furthermore, it will enable us to obtain estimates of migration and determine the role of selection in maintaining variation.

Finally, estimates of molecular variation between populations can greatly complement phenotypic studies of geographic variation. This requires molecular markers which are neutral as they can provide information regarding the history of populations. However, no precise information on migration is readily available for the populations surveyed here, although it is abundantly clear that gene flow is high in *D. melanogaster* . Despite this, comparisons of geographic structure between the molecular markers from this study and morphological markers previously studied in this lab, reveal a disparity between geographic distribution patterns. Both molecular markers clearly indicate monotonic clinal variation in several instances supporting a climate or environmentally driven explanation to account for the structuring. In contrast, the overall trend from the morphological analysis yields a clearly non-monotonic pattern with selection for optimal body size being hypothesized to account for the pattern. Further empirical work involving neutral molecular markers must be undertaken in order to better understand the role of selection in determining the latitudinal variation in both molecules and morphology.

Chapter 6

APPENDICES

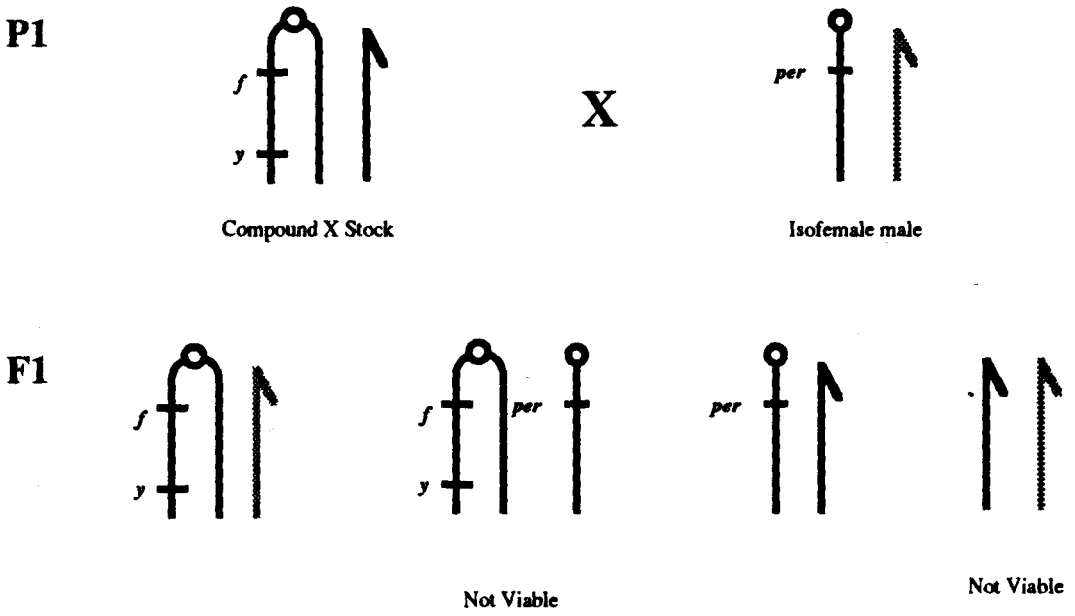


Figure A1: Cross conducted between wildtype male and Compound X female in order to standardize the X chromosome.

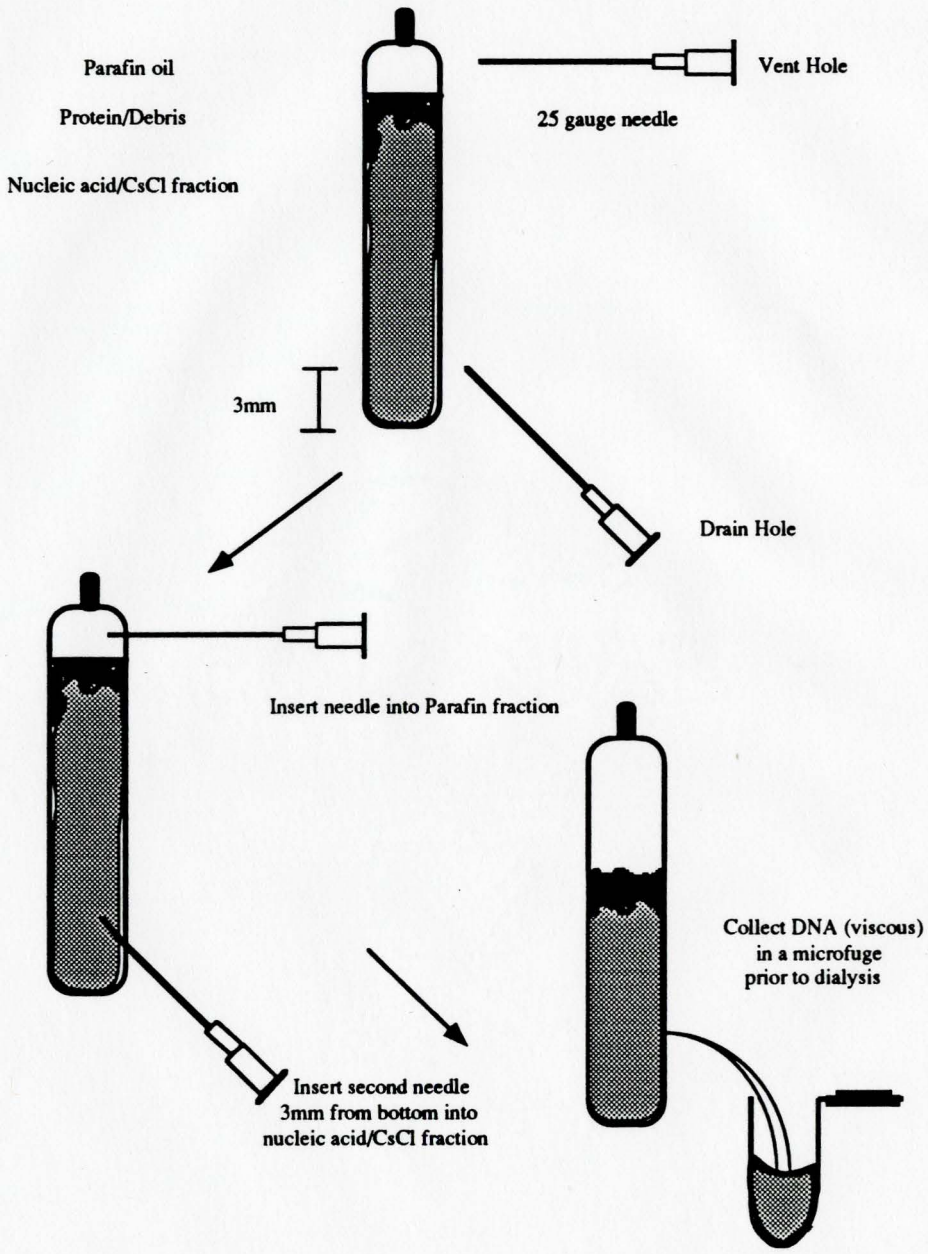


Figure A2: Methodology employed to drip CsCl gradients

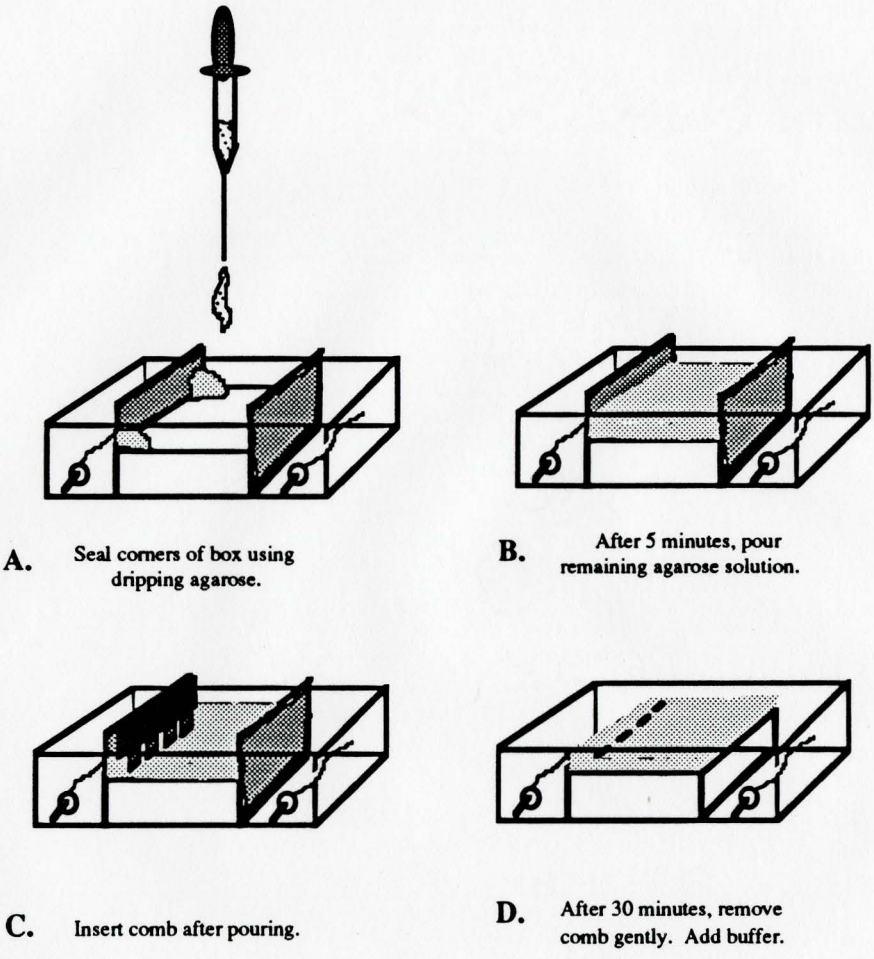


Figure A3: Casting of agarose gels and edge sealing.

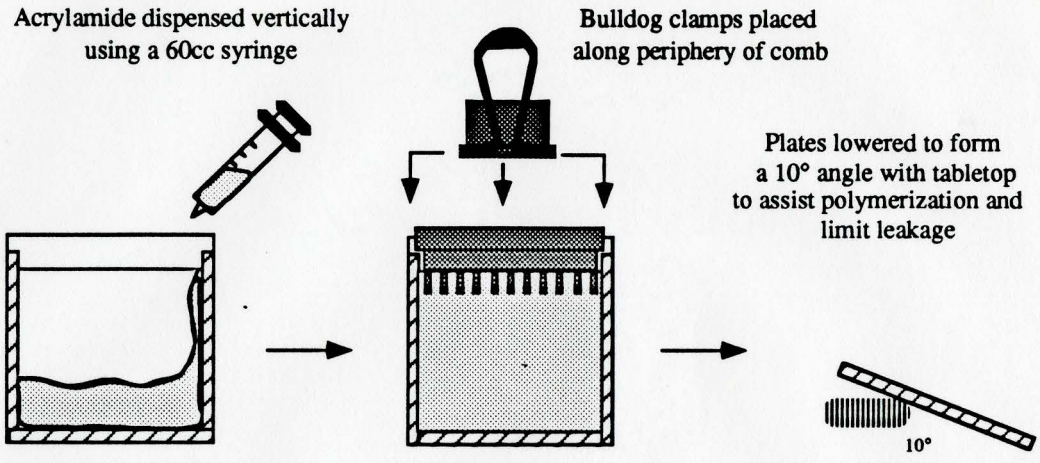


Figure A4: Methodology employed in casting acrylamide gels.

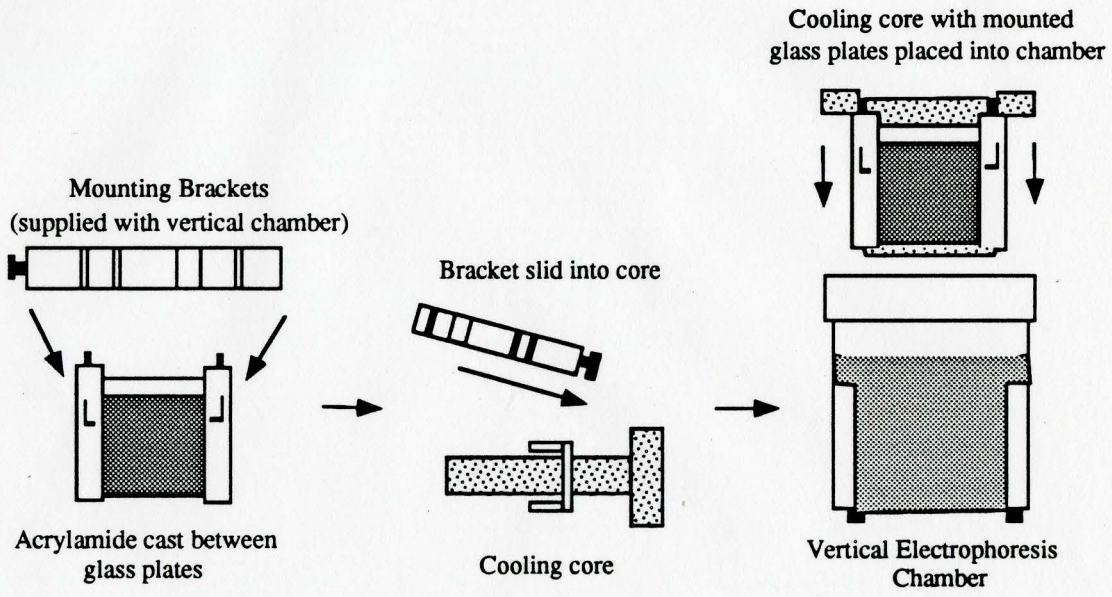


Figure A5: Securing glass plates and vertical chamber set-up.

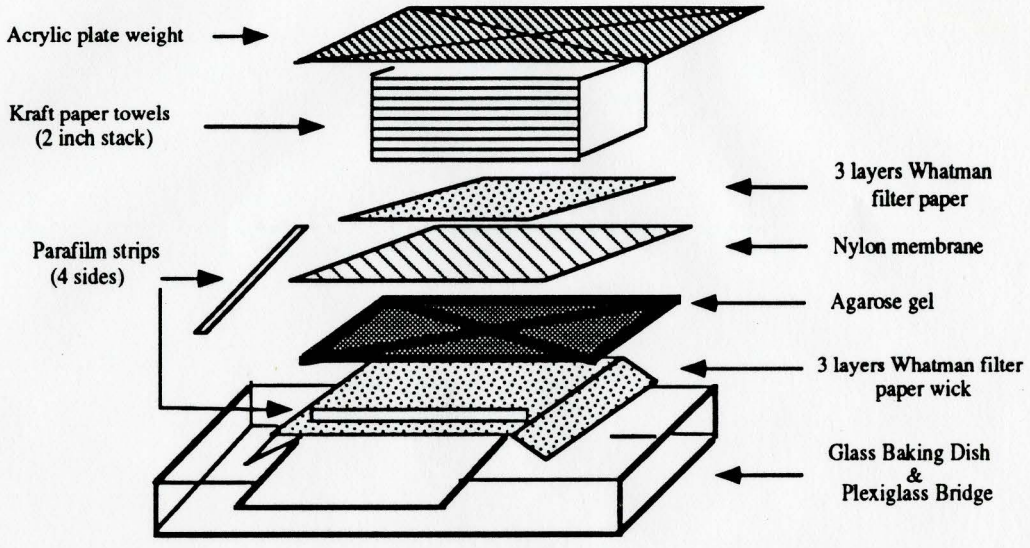


Figure A7: Setting up a Southern blot via the capillary method.

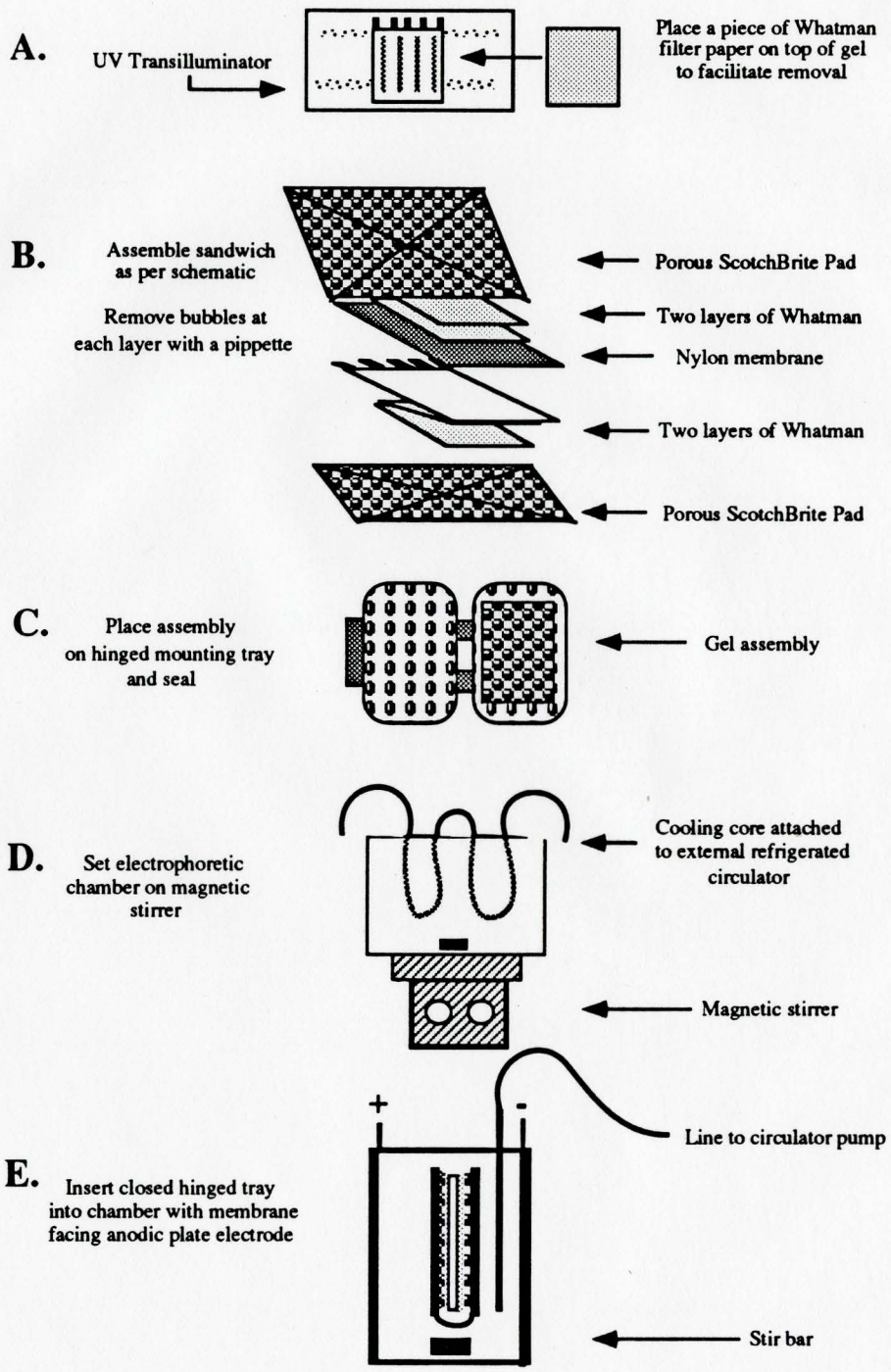


Figure A7: Methodology employed for electrophoretic transfer of DNA to nylon.

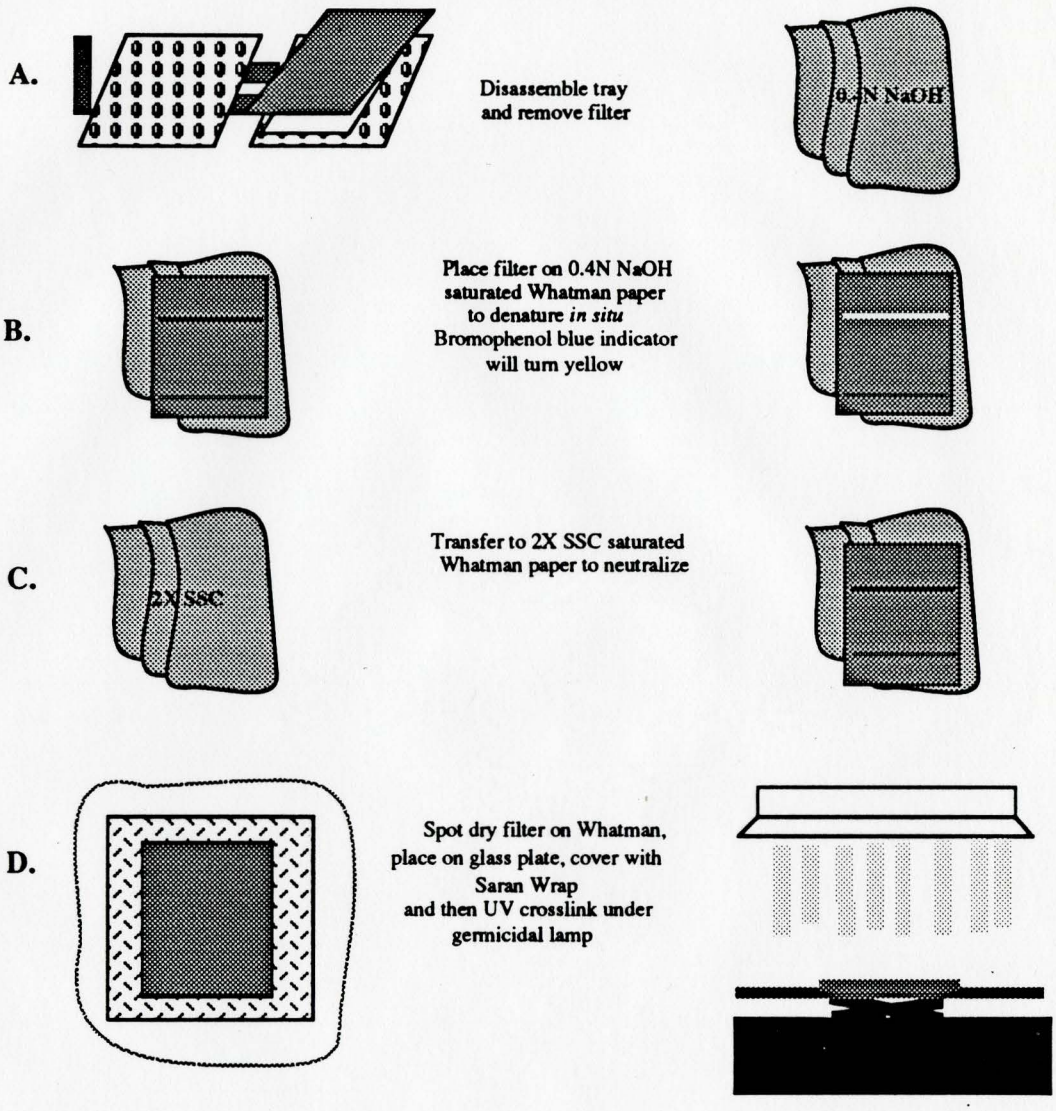
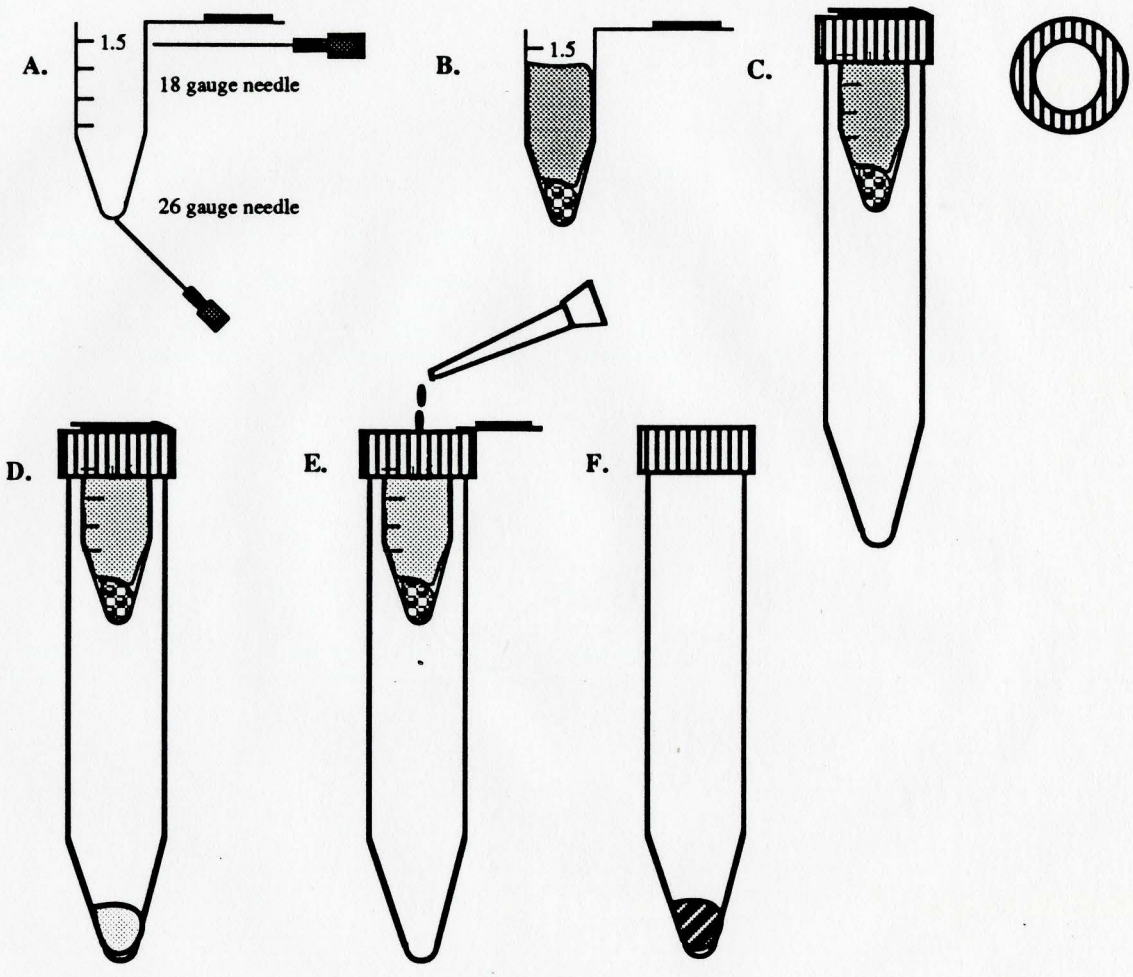


Figure A8: Methodology employed for *in situ* denaturation of filter bound nucleic acids and subsequent UV crosslinking.



A. Pierce a microfuge tube with the appropriate gauge needles.
B. Fill 1/3 up the slant with glass beads and then add 1mL TNE_2 saturated Sepharose CL6B.
C. Place microfuge inside an orange capped centrifuge tube with its top cut out.
D. Spin column inside a clinical centrifuge at full speed for 1m45s to compact Sepharose.
E. Transfer column to new centrifuge tube, add labelled reaction and spin again for 1m45s.
F. Remove column, collect hot eluate and transfer to new microfuge tube if desired.

Figure A9: Preparation of spin columns to remove unincorporated radioactive label.

AGTGGC TCCTCG GGCAAC TTCACC ACCGCC AGTAAC ATACAC ATGAGC AGTGTG ACAAAT

Thr-Gly
ACGAGC ATTGCC GGCACT GGTGGC ACGGGC ACTGGT ACAGGT ACAGGT ACTGGA ACTGGA

ACTGGA ACCGGG ACAGGA ACTGGA ACGGGA ACAGGT ACAGGC ACAGGC ACAGGC

ACTGGA ACAGGC AATGGA ACAAAT TCCGGC ACCGGA ACCGGC ACCGCC AGCTCA
Thr-Gly

TCCAAA GGCGGA AGCGCC GCCATA CCGCCA GTCACG CTGACC GAATCC CTGCTC AATAAG

CACAAC GACGAG ATGGAG AAGTTC ATGCTG

Figure B1: Comparison of the TG17 variant sequence observed in lines WE6, LC8, OT15, TA51 and CF2 to that of the TG17 Thr-Gly region sequence reported by Citri *et al.*, (1987). *All five alleles* sequenced as part of this study were identical in composition to the referenced sequence. Underlined sequences represent first and last Thr-Gly pairs.

AGTGGC TCCTCG GGCAAC TTCACC ACCGCC AGTAAC ATACAC ATGAGC AGTGTG ACAAAT

Thr-Gly
ACGAGC ATTGCC GGCACT GGTGGC ACGGGC ACTGGT ACAGGT ACAGGT ACTGGA ACTGGA

ACTGGA ACCGGG ACAGGA ACTGGA ACCGGG ACAGGA ACTGGA ACGGGA ACAGGT

ACAGGC ACAGGC ACAGGC ACTGGA ACAGGC AATGGA ACAAAT TCCGGC ACCGGA
Thr-Gly

ACCGGC ACCGCC AGCTCA TCCAAA GGCGGA AGCGCC GCCATA CCGCCA GTCACG CTGACC

GAATCC CTGCTC AATAAG CACAAC GACGAG ATGGAG AAGTTC ATGCTG

Figure B2: Comparison of the TG20 variant sequence observed in lines WC5, LB8, OT12, TA05 and CA3 to that of the TG20 Thr-Gly region sequence reported by Citri *et al.*, (1987). *All five alleles* sequenced as part of this study were identical in composition to the referenced sequence. Underlined sequences represent first and last Thr-Gly pairs.

AGTGGC TCCTCG GGCAAC TTCACC ACCGCC AGTAAC ATACAC ATGAGC AGTGTG ACAAAT

Thr-Gly
ACGAGC ATTGCC GGCACT GGTGGC ACGGGC ACTGGT ACAGGT ACAGGT ACTGGA ACTGGA
G A T

ACTGGA ⟶ ACCGGG ACAGGA ACTGGA ACCGGG ACAGGA ACTGGA ACGGGA ACAGGT
ACTGGA G A T A C T G A

ACAGGC ACAGGC ACAGGC ACTGGA ACAGGC ⟶ AATGGA ACAAAT TCCGGC ACCGGA
T A G C ACAGGC G
Thr-Gly

ACCGGC ACCGCC AGCTCA TCCAAA GGCGGA AGCGCC GCCATA CCGCCA GTCACG CTGACC
G

GAATCC CTGCTC AATAAG CACAAC GACGAG ATGGAG AAGTTC ATGCTG
T

Figure B3: Comparison of the TG22 variant sequence observed in lines WC2,LC7, OT19, and CD15 to that of the TG20 Thr-Gly region sequence reported by by Citri *et al.*, (1987). Nucleotides below the reference sequence represent those observed in the TG20 sequence. Underlined sequences represent first and last Thr-Gly pairs and arrows indicate sequence insertions.

GGCACT GGTGGC ACGGGC ACTGGT ACAGGT ACAGGT ACAGGT ACTGGA
T A

ACTGGA ACTGGA ACCGGG ACAGGA ACTGGA ACGGGA ACAGGT ACAGGC
G

ACTGGA ACGGGA ACAGGC ACTGGA ACAGGC ACAGGC ACAGGC ACTGGC

ACAGGC AATGGA ACAAAT

Figure B4: Comparison of the TG23 variant sequence (ME2) of Costa *et al.*, (1991) to that of the TG23 variant sequence from the present study (lines WC3, CB3, TA02, OT13 and LC5-shown below fully referenced sequence). Both sequences differ markedly from the sequence reported by Citri *et al.*, (1987), the ME2 variant differing at 21 synonymous sites, but resemble each other very closely.

APPENDIX C

Hypercard Randomization Stack Code

```
on mouseUp
        put card field "nov" into nov
        put empty into card field "z"
        put empty into card file "output"
        put 0 into count

        lock screen

        repeat for nov times
                put empty into card field "z"
                put count + 1 into count
                put count into msg

                put empty into card field "per"
                repeat five times
                        put "b" & return after card field "per"
                end repeat

                put 1 into a

                repeat for five times
                        repeat forever
                          put random (5) into r
                            if line r of card field "per" = "b" then
                                put line r of card field "input"
                                into line r of card field "per"
                            exit repeat
                          end if
                        end repeat

                        put a + 1 into a

                end repeat

                put 0 into sx
                put 1 into a
                  repeat for 5 times
                        put line a of card field "per" into x
                        put sx + x into sx
                        put a + 1 into a
                  end repeat
```

```
put sx/5 into mx

put line 1 of card field "per" into a
put a - mx into zt
put line 2 of card field "per" into a
put a - mx into zc
put line 3 of card field "per" into a
put a - mx into zl
put line 4 of card field "per" into a
put a - mx into zw
put line 5 of card field "per" into a
put a - mx into zo

put zo*zw&return after card field "z"
put zo*zl&return after card field "z"
put zo*zc&return after card field "z"
put zo*zt&return after card field "z"
put zw*zo&return after card field "z"
put zw*zl&return after card field "z"
put zw*zc&return after card field "z"
put zw*zt&return after card field "z"
put zl*zo&return after card field "z"
put zl*zw&return after card field "z"
put zl*zc&return after card field "z"
put zl*zt&return after card field "z"
put zc*zo&return after card field "z"
put zc*zw&return after card field "z"
put zc*zl&return after card field "z"
put zc*zt&return after card field "z"
put zt*zo&return after card field "z"
put zt*zw&return after card field "z"
put zt*zl&return after card field "z"
put zt*zc&return after card field "z"

put zt*zt+zc*zc+zl*zl+zw*zw+zo*zo into szs

put 1 into a
put 0 into sw450
put 0 into sw900
put 0 into sw1250
repeat for 20 times
        put line a of card field "z" into z
        put word 1 of line a of card field "ww" into w450
        put word 2 of line a of card field "ww" into w900
        put word 3 of line a of card field "ww" into w1250

        put sw450 + (w450 * z) into sw450
        put sw900 + (w900 * z) into sw900
```

```
                put sw1250 + (w1250 * z) into sw1250

                put a + 1 into a
            end repeat

            put (5 * sw450) / (8 * szs) after card field "output"
            put "      " after card field "output"
            put (5 * sw900) / (8 * szs) after card field "output"
            put "       " after card field "output"
            put (5 * sw1250) / (4 * szs) after card field "output"
            put "       " after card field "output"


            unlock screen
            lock screen

        end repeat

        unlock screen
        beep

end mouseUp
```

Hypercard Mantel Test Stack Code

```
on mouseUp
        put card field "nov" into nov
        put empty into card field "z"
        put empty into card file "output"
        put 0 into count

        lock screen

        repeat for nov times
                put empty into card field "z"
                put count + 1 into count
                put count into msg

                put empty into card field "per"
                repeat 45 times
                        put "b"&return after card field "per"
                end repeat

                put 1 into a

                repeat for 45 times
                        repeat forever
                          put random (45) into r
                             if line r of card field "per" = "b" then
                                 put line a of card field "x"
                                 into line r of card field "per"
                             exit repeat
                          end if
                        end repeat

                        put a + 1 into a

                end repeat

                        put empty into card field "output"
                        repeat 10 times
                                put "b"&return after card field "output"
                        end repeat

                        put 1 into a
                        repeat for 10 times
                                repeat forever
                                put random (10) into r
                                if line r of card field "output" = "b" then
```

```
                        put line a of card field "input 2"
                        into line r of card field "output"
                        exit repeat
                        end if
                end repeat

        put 1 into a
        put 0 into z
        repeat for 45 times
                put line a of card field "per" into x
                put line a of card field "y" into y
                put z + (x*y) into z
                put a + 1 into a
        end repeat

        put z&return after card field "z"


        unlock screen
        lock screen

    end repeat

    unlock screen
    beep

end mouseUp
```

Chapter 7

BIBLIOGRAPHY

Anderson, P.R. and J.G. Oakeshott. 1984. Parallel geographic patterns of allozyme variation in two sibling *Drosophila* species. *Nature.* **308**: 729-731.

Aquadro, C.F. 1992. Why is the genome variable? Insights from *Drosophila. TIG.* **8**: 84-91.

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

Ashburner, M. 1989. *Drosophila a Laboratory Manual* . 1st ed. Cold Spring Harbor Press, Cold Spring Harbor, New York.

Baker, A.J. 1992. Genetic and morphometric divergence in ancestral European and descendant New Zealand populations of chaffinches (*Fringilla coelebs*). *Evolution.* **46**: 1784-1800.

Barbujani, G. 1987. Autocorrelation of gene frequencies under isolation by distance. *Genetics.* **117**: 777-782.

Barker, J.S.F. and J.C. Mulley. 1976. Isozyme variation in natural populations of *Drosophila buzzatii. Evolution.* **30**: 213-233.

Barker, J.S.F., P.D. East and B.S. Weir. 1986. Temporal and microgeographic variation in allozyme frequencies in a natural population of *Drosophila buzzatii. Genetics.* **112**: 577-611.

Begun, D.J. and C.F. Aquadro. 1991. Molecular population genetics of the distal portion of the X chromosome in *Drosophila*: evidence for genetic hitchhiking of the *yellow-achaete* region. *Genetics.* **129**: 1147-1158.

Begun, D.J. and C.F. Aquadro. 1992. Levels of naturally occuring polymorphism correlate with recombination rates in *D. melanogaster. Nature.* **356**: 519-520.

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

Cavener, D. and M. Clegg. 1981. Multigenic response to ethanol in *Drosophila melanogaster. Evolution.* **35**: 1-10.

Chung, C.T., S. Niemela and R.H. Miller. 1989. One-step preparation of competent *Escherichia coli:* Transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci.USA.* **86**: 2172-2175.

Church, G.M. and W. Gilbert. 1984. Genomic sequencing. *Proc. Natl. Acad. Sci. USA.* **81**: 1991-1995.

Citri, Y., H.V. Colot, A.C. Jacquier, Q. Yu, J.C. Hall, D. Baltimore, M. Rosbash. 1987. A family of unusually spliced biologically active transcripts encoded by a *Drosophila* clock gene. *Nature.* **326**: 42-47.

Claussen, J., D. Keck and W. Hiesey. 1941. Regional differentiation in plant species. *Am. Nat.* **75**: 231-250.

Cliff, A.D. and J.K. Ord. 1981. *Spatial Processes: Models and Applications.* Pion, London.

Colot, H.V., J.C. Hall, M. Rosbash. 1988. Interspecific comparison of the *period* gene of *Drosophila* reveals large blocks of non-conserved coding DNA. *EMBO J.* **7**: 3929-3937.

Costa, R., A.A. Peixoto, J.R. Thackeray, R. Dalgleish and C.P. Kyriacou. 1991. Length polymorphism in the threonine-glycine-encoding repeat region of the *period* gene in *Drosophila. J. Mol. Evol.* **32**: 238-246.

Costa, R., A.A. Peixoto, G. Barbujani and C.P. Kyriacou. 1992. A latitudinal cline in a *Drosophila* clock gene. *Proc. R. Soc. Lond. B.* **250**: 43-49.

Coyne, J.A. and B. Milstead. 1987. Long distance migration of *Drosophila.* 3. Dispersal of *D. melanogaster* alleles from a Maryland orchard. *Am. Nat.* **130**: 70-82.

David, J.R. 1982. Latitudinal variability of *Drosophila melanogaster*: allozyme frequencies divergence between European and Afrotropical populations. *Biochem. Genet.* **20:** 747-761.

DiMichele, L., K.T. Paynter and D.A. Powers. 1991. Evidence of Lactate dehydrogenase-B allozyme effects in the Teleost, *Fundulus heteroclitus. Science.* **253**: 898-900.

Dobzhansky, Th. and B. Spassky. 1953. Genetics of natural populations XXI. Concealed variability in two sympatric species of *Drosophila. Genetics.* **38**: 471-484.

Endler, J.A. 1977. *Geographic Variation, Speciation and Clines.* Princeton University Press, Princeton NJ.

Epperson, B.K. 1993. Spatial and space-time correlations in systems of subpopulations with genetic drift and migration. *Genetics.* **133**: 711-727.

Epperson, B.K. and M.T. Clegg. 1986. Spatial autocorrelation analysis of flower colour polymorphisms within substructured populations of morning glory (*Ipomoea purpurea*). *Am. Nat.* **128**: 840-858.

Erlich, H.A. and N. Arnheim. 1992. Genetic analysis using the Polymerase Chain Reaction. *Ann. Rev. Genet.* **26**: 479-506.

Esteal, S. 1985. The ecological genetics of introduced population sof the giant toad, *Bufo marinus*. III. Geographical patterns of variation. *Evolution.* **39**: 238-246.

Ewens, W.J. 1977. Population genetics theory in relation to the neutralist-selectionist controversy. *Adv. Hum. Genetics.* **8**: 67-134.

Ewer, J., M. Hamblen-Coyle, M. Rosbash and J.C. Hall. 1990. Requirement for *period* gene expression in the adult and not during development for locomoter activity rhythms of imaginal *Drosophila melanogaster*. *J. Neurogenet.* **7**: 31-73.

Feinberg, A.P. and B. Vogelstein. 1983. A technique for radiolabelling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**: 6-13.

Hall, J.C. and C.P. Kyriacou. 1990. Genetics of biological rhythms in *Drosophila*. *Adv. Insect Physiol.* **22**: 221-298.

Harris, H. 1966. Enzyme polymorphism in man. *Proc. R. Soc. Ser. B.* **164**: 298-310.

Hubby, J.L. and R.C. Lewontin. 1966. A molecular aproach to the study of genic heterozygosity in natural populations. I. The number of alleles at different loci in *Drosophila pseudoobscura*. *Genetics.* **54**: 577-594.

Hudson, R.R. 1990. Gene genealogies and the coalescent process. *Oxf. Surv. Evol. Biol.* vol **7**.

Hudson, R.R. and N.L. Kaplan. 1988. The coalescent process in models with selection and recombination. *Genetics.* **120**: 831-840.

Hudson, R.R., M. Kreitman and M. Aguade. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics.* **116**: 153-159.

Huxley, J.S. 1939. Clines: An auxiliary taxonomic principle. *Nature.* **142**: 219-220.

Jackson, F.R., T.A. Bargiello, S.H. Yun and M.W. Young. 1986. Product of *per* shares homology with proteogycans. *Nature.* **320**: 185-188.

Kaplan, N.L., T. Darden and R.R. Hudson. 1988. The coalescent process in models with selection. *Genetics.* **120**: 819-829.

Kaplan, N.L., Hudson, R.R. and C.H. Langley. 1989. The "hitchhiking effect revisited. *Genetics*. **123**: 887-899.

Kettlewell, H.B.D., and R.J. Berry. 1961. The study of a cline. *Amathes glareosa* Esp. and its melanic *f. edda* Stand. (Lep) in Shetland. *Heredity*. **16**: 403-414.

Kimura, M. 1983. *The Neutral Theory of Molecular Evolution* , Cambridge University Press, Cambridge.

Kliman, R.M. and J. Hey. 1993. DNA sequence variation at the *period* locus within and among species of the *Drosophila melenogaster* complex. *Genetics*. **133**: 375-387.

Koehn, R.K., R.I.E. Newell and F. Immermann. 1980. Maintenance of an aminopeptidase allele frequency cline by natural selection. *Proc. Natl. Acad. Sci. USA*. **77**: 5385-5389.

Koehn, R.K., A.J. Zera and J.G. Hall. 1983. Enzyme polymorphism and natural selection. *Evolution of Genes and Proteins* (ed. M. Nei and R.K. Koehn) pp. 115-136. Sinauer, Sunderland, Massachusetts.

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

Kreitman, M. 1987. Molecular population genetics. *Oxf. Surv. Evol. Bol.* vol. **4**.

Kreitman, M. 1991. Detecting selection at the level of DNA. *Evolution at the Molecular Level* (ed. R.K. Selander and A.G. Clark), pp. 204-221, Sinauer, Sunderland, Massachusetts.

Kreitman, M. and M. Aguade. 1986a. Genetic uniformity in two populations of *Drosophila melanogaster* as revealed by filter hybridization of four-nucleotide-recognizing restriction enzyme digests. *Proc. Natl. Acad. Sci. USA*. **83**: 3562-3566.

Kreitman, M. and M. Aguade. 1986b. Excess polymorphism at the *Adh* locus in *Drosophila melanogaster* . *Genetics*. **114**: 93-110.

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

Kyriacou, C.P. 1990. The molecular ethology of the *period* gene in *Drosophila*. *Behav. Genet*. **20**: 191-211.

Lachaise, D., M. Cariou, J.R. David, F. Lemeunier, L. Tsacas and M. Ashburner. 1988. Historical biogeography of the *Drosophila melanogaster* species subgroup. *Evol. Biol*. **22**: 159-225.

Langley, C.H., J. MacDonald, N. Miyashita and M. Aguadé. 1993. Lack of correlation between interspecific divergence and intraspecific polymorphism at the suppresor of forked region in *Drosophila melanogaster* and *Drosophila simulans*. *Proc. Natl. Acad. Sci. USA*. **90**: 1800-1803.

Leigh-Brown, A.J. 1989. Population genetics at the DNA level: a review of the contribution of restriction enzyme studies. *Oxf. Surv. Evol. Biol.* vol **6**.

Lewontin, R.C. 1991. Electrophoresis in the development of evolutionary genetics: Milestone or millstone? *Genetics*. **128**: 657-662.

Lewontin, R.C. and J.L. Hubby. 1966. A molecular approach to the study of genic heterozygosity in natural populations. II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila melanogaster*. *Genetics*. **54**: 595-606.

Long, A.D. 1992. *The Quantitative Genetics of Clinal Variation in Drosophila melanogaster*. M.Sc. thesis, McMaster University.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. *Cancer Res*. **27**: 209-220.

Maynard Smith, J., and J. Haigh. 1974. The hitchhiking effect of a favourable gene. *Genet. Res*. **23**: 23-35.

Menozzi, P., A. Piazza and L.L. Cavalli-Sforza. 1978. Synthetic maps of human gene frequencies in Europeans. *Science*. **201**: 786-792.

Miyashita, N. and C.H. Langley. 1988. Molecular and phenotypic variation of the *white* locus region in *Drosophila melanogaster*. *Genetics*. **120**: 199-212.

Nei, M. 1972. Genetic distance between populations. *Am. Nat*. **106**: 283-292.

Nevo, E. 1978. Genetic variation in natural populations: Pattern and theory. *Theor. Pop. Biol*. **13**: 121-177.

Oakeshott, J.G., J.B. Gibson, P.R. Anderson, W.R. Knibb, D.G. Anderson and G.K. Chambers. 1982. Alcohol dehydrogenase and glycerol-3-phosphate dehydrogenase clines in *Drosophila melanogaster* on different continents. *Evolution*. **36**: 86-96.

Oden, N.L. 1984. Assessing the significance of a spatial correlogram. *Geog. Anal*. **16**: 1-16.

Peixoto, A.A., S. Campesan, R. Costa and C.P. Kyriacou. 1993. Molecular evolution of a repetitive region within the *per* gene of *Drosophila*. *Mol. Biol. Evol*. **10**: 127-139.

Pittendrigh, C.S. 1954. On the temperature independence in a clock system controlling emergence time in *Drosophila*. *Proc. Natl. Acad. Sci. USA*. **40**: 1018-1029.

Powers, D.A., G.S. Greaney and A.R. Place. 1979. Physiological correlation between lactate dehydrogenase genotype and hemoglobin function in killifish. *Nature*. **227**: 240-241.

Prout, T and J.S.F. Barker. 1993. *F* - statistics in *D. buzzatii:* selection, population size and inbreeding. *Genetics*. **134**: 369-375.

Riley, M.A., M.E. Hallas and R.C. Lewontin. 1989. Distinguishing the forces controlling genetic variation at the *Xdh* locus in *Drosophila pseudoobscura* *Genetics*. **123**: 359-369.

Royaltey, H.H., E. Astrachan and R.R. Sokal. 1975. Tests for patterns in geographical variation. *Geog. Anal.* **7**: 369-395.

Saiki, R.K., S. Scharf, F. Faloona, K. Mullis, G. Horn *et al.* 1985. Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science*. **230**: 1350-1354.

Sakai, A.K. and N.L. Oden. Spatial pattern of sex expression in silver maple (*Acer saccharinum* L.): Morista's index and spatial autocorrelation. *Am. Nat.* **122**: 489-508.

Sambrook, J., E.F. Fritsch and T. Maniatis. 1989. *Molecular Cloning a Laboratory Manual.*, 2nd edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

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

Singh, R.S. 1979. Genic heterogeneity within electrophoretic alleles and the pattern of variation among loci in *Drosophila pseudoobscura*. *Genetics*. **93**: 997-1018.

Singh, R.S. and L.R. Rhomberg. 1987a. A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. II. Estimates of heterozygosity and patterns of geographic differentiation. *Genetics*. **117**: 255-271.

Singh, R.S. and L.R. Rhomberg. 1987b. A comprehensive study of genic variation in natural populations of *Drosophila melanogaster*. I. Estimates of gene flow from rare alleles. *Genetics*. **115**: 313-322.

Singh, R.S. and A.D. Long. 1992. Geographic variation in *Drosophila*: From molecules to morphology and back. *TREE*. **7**: 340-345.

Singh, R.S., D.A. Hickey and J. David. 1982. Genetic differentiation between geographically distant populations of *Drosophila melanogaster*. *Genetics*. **101**: 235-256.

Singh, R.S., R.C. Lewontin and A.A. Felton. 1976. Genetic heterogeneity within electrophoretic alleles of xanthine dehydrogenase in *Drosophila pseudoobscura*. *Genetics*. **84**: 609-629.

Slatkin, M., and T. Maruyama. 1975. Genetic drift in a cline. *Genetics*. **81**: 209-222.

Sokal, R.R. 1979. Ecological patterns inferred from spatial correlograms. *Contemporary Quantitative Ecology and Related Econometrics* (ed. G.P. Patil and M.L. Rosenzweig), pp. 167-196. Internatinal Cooperative Publishing House, Maryland.

Sokal, R.R. and N.L. Oden. 1978. Spatial autocorrelation in biology. *Biol. J. Linn. Soc*. **10**: 199-228.

Sokal, R.R. and P. Menozzi. 1982. Spatial autocorrelation of HLA frequencies in Europe support demic diffusion of early farmers. *Am. Nat*. **119**: 1-17.

Sokal, R.R. and F.J. Rohlf. 1981. *Biometry*. Freeman Publishing Company. San Francisco, California.

Sokal, R.R. and D.E. Wartenberg. 1983. A test of spatial autocorrelation analysis using an isolation by distance model. *Genetics*. **105**: 219-237.

Sokal, R.R., J. Bird and B. Riska. 1980. Geographic variation in *Pemphigus populicaulis* (Insecta: Aphididae) in Eastern North America. *Biol. J. Linn. Soc*. **14**: 163-200.

Sokal, R.R., N.L. Oden, and J.S.F. Barker. 1987. Spatial structure in *Drosophila buzzatii* populations: simple and directional spatial autocorrelation . *Am. Nat*. **129**: 122-142.

Sokal, R.R., R.M. Harding and N.L. Oden. 1989. Spatial patterns of human gene frequencies in Europe. *Am. J. Phys. Anthrop*. **80**: 267-294.

Southern, E.M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. *J. Mol. Biol*. **987**: 503-517.

Strobeck, C. 1983. Expected linkage disequilibrium for a neutral locus linked to a chromosomal arrangement. *Genetics*. **103**: 545-555.

Thackeray, J.R., C.P. Kyriacou. 1990. Molecular evolution in the *Drosophila yakuba period* locus. *J.Mol. Evol*. **30**: 389-401

Van Delden, W. 1982. The alcohol dehydrogenase polymorphism in *Drosophila melanogaster*: selection at an enzyme locus. *Evol. Biol.* **15**: 187-222.

Wheeler, D.A., C.P. Kyriacou, M.L. Greenacre, Q. Yu, J.E. Rutila, M. Rosbash and J.C. Hall. 1991. Molecular transfer of a species specific courtship behaviour from *Drosophila simulans* to *Drosophila melanogaster*. *Science*. **251**: 1082-1085.

Wright, S. 1951. The genetical structure of populations. *Ann. Eugenics*. **15**: 323-354.

Wright, S. 1978. *Evolution and Genetics of Populations*. IV. Variability within and among natural populations. University of Chicago Press, Chicago, Illinois.

Yu, Q., H.V. Colot, C.P. Kyriacou, J.C. Hall and M. Rosbash. 1987. Behaviour modification by in vitro mutagenesis of a variable region within the *period* gene in *Drosophila*. *Nature*. **326**: 765-769.