AN ANALYSIS OF THE POPULATION GENETIC STRUCTURE AND SPECIES HISTORY OF *DROSOPHILA MELANOGASTER* AND *DROSOPHILA SIMULANS* USING RESTRICTION FRAGMENT LENGTH POLYMORPHISMS OF MITOCHONDRIAL DNA

By

© LAWRENCE RICHARD HALE, B.Sc. (Hons.)

A thesis
Submitted to the School of Graduate Studies
in Partial Fulfilment of the Requirements
for the Degree Doctor of Philosophy

McMaster University
August, 1989

MITOCHONDRIAL DNA VARIATION IN NATURAL POPULATIONS OF *DROSOPHILA*

DOCTOR OF PHILOSOPHY (1989)
(Biology)

McMASTER UNIVERSITY
Hamilton, Ontario

TITLE: An analysis of population genetic structure and species history of Drosophila melanogaster and Drosophila simulans using restriction fragment length polymorphisms of mitochondrial DNA

AUTHOR: Lawrence Richard Hale, B.Sc. (Hons.)
(Simon Fraser University)

SUPERVISOR: Professor Rama S. Singh

NUMBER OF PAGES: xiv , 216

ABSTRACT

Animal mitochondrial DNA (mtDNA) has several features that give it great utility in the study of geographic structure of natural populations. Its small size and covalently closed circular conformation make it easy to purify. Strict maternal inheritance and homoplasmy makes the effective copy number of mtDNA as little as 1/4 that of nuclear loci; this renders populational complements of mtDNA less susceptible to the homogenizing effects of gene flow due to migration, and more susceptible to founder effects due to fluctuations in effective population size. The absence of recombination allows for comparatively simple reconstruction of genealogies of mtDNA variants. A comparatively high rate of base substitution assures that most species will have enough mtDNA variants to make population genetic inferences meaningful.

Finally, sequence variants of mtDNA harboured in natural populations are presumed to be selectively neutral (i.e. has no effect on organismal fitness). This means that the distribution of mtDNA variants in populations will be due entirely to the stochastic and deterministic effects of species history, and not to natural selection. As a result, mtDNA can be used to infer species history directly and, after allowing for differences in modes of transmission, to infer the action of natural selection on other genetically determined factors like allozymes. The research of this dissertation was to survey world-wide natural populations of Drosophila melanogaster and its sibling species D. simulans for restriction fragment length polymorphisms of mtDNA. These cosmopolitan species have been widely studied for variation of allozymes and many other genetically determined factors (e.g. chromosomal inversions, morphometric characters).

A total of 144 isofemale lines of D. melanogaster were

analyzed from 18 geographic populations. Considerable size variation was observed in this sample. Most size variation occurs in the major non-coding region (A+T-rich region), yielding a total size range of 18.2 kbp to 19.9 kbp. The occurrence of several size variants among all haplotypes indicates that the rate of size mutation is quite high. Further, the frequency distribution of size variants suggests that there is selection against larger sized mtDNA molecules; a replication advantage to smaller sized molecules is a likely explanation. Finally, there is evidence for 'small-scale' size variation (i.e. a total range of 20 bp) in the coding region of mtDNA. However, it could not be determined if the observed mobility variation of the fragments in question was actually due to addition/deletion of DNA or due to conformational effects of particular base substitutions.

Using 10 restriction enzymes, 23 restriction haplotypes were observed in D. melanogaster mtDNA. The phylogeographic distribution of these haplotypes allowed populations to be roughly divided into three longitudinal regions. Euro-African populations showed high intra- and inter-populational diversity and were inferred to be the oldest. Far East populations showed low intra-populational diversity but high inter-populational diversity; these populations similarly have a lengthy, but more complex, history. Western Hemisphere populations have low intra- and inter-populational diversity, and were inferred to be the youngest. A colonization history of this species is proposed. The species history suggested by mtDNA alone is quite similar to one proposed from the collective results of studies on several genetically-influenced traits in D. melanogaster (David and Capy 1988). mtDNA analysis is therefore shown to be a very efficient means by which to study species history. The very different histories of these regions reinforces the

notion that the parallel latitudinal clines of allozyme alleles observed in this species (Singh et al 1982) are due to natural selection and are not historical.

A total of 79 isofemale lines of D. simulans from 14 geographic populations (13 continental populations, and the Seychelles Islands) were surveyed. The 10 restriction enzymes revealed a very discontinuous distribution of variation. The continental populations harboured only four haplotypes, while the Seychelles population harboured two haplotypes that were very different from the continental ones. The discontinuous distribution suggests that D. simulans evolved as a series of allopatric ancestral populations, as has been previously suggested. The relative lack of diversity among the continental mtDNAs, coupled with the observation that none of the four continental haplotypes are unique to a single population, strongly suggests that continental populations of D. simulans result from a recent world-wide expansion from one of the ancestral populations, possibly concurrent with the expansion into the Western Hemisphere by D. melanogaster. A recent expansion would explain the lower degree of allozyme population structure seen in D. simulans than in D. melanogaster, and is more parsimonious than the alternative explanations of a narrower niche-width or the adoption of a general purpose genotype.

ACKNOWLEDGEMENTS

I wish to express my special gratitude to Dr. Rama Singh for giving me the freedom to move in directions new to his laboratory, and for his consistent encouragement and support. Particular thanks also go to Dr. Richard Morton for never letting me get too comfortable.

I have been privileged to have a fine complement of fellow Drosophila apprentices: Shanta Thomas, Ling-Wen Zeng, Tony Long, Mike Coulthart, Madhu Choudhary, Bill McMillan, Jane Pursey, John Prevec, Nicholas Okoampah, and Dan Houpt. They were my chief sounding-board for speculations and ideas.

I have benefitted scientifically from many other people in the department. Special mention must go to the graduate students and technicians in the laboratories of Drs. Mak, Prevec and Graham, for helping me with the realities of molecular biology.

I wish to thank Abe Haddad, Debbie Long, Scott Reid, and Rodney Rhem for being special friends during my time here. I will miss you all terribly. The collective insanity of the 'ladies' in the departmental office (and glass-washing facility) never failed to cheer me up when things got a little tough. Thanks gang.

Above all, I wish to thank my dear wife, Linda, for understanding what graduate work requires of a spouse, and loving me anyway. Finally, my thanks to Alan for making this past year a special joy and keeping me cognizant of what life is really all about.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PAGE

1 INTRODUCTION

1.1 Historical overview

The common ancestry of all life through evolutionary change has been corroborated to the point of a scientific law. So central to modern biology is the concept of evolution that, as Dobzhansky put it:

> "Nothing in biology makes sense except in the light of evolution" (Dobzhansky, 1973).

Yet evolution remains a source of energetic scientific dispute. What is at issue, however, is not whether evolution has occurred, but rather how it has occurred. That is, what are the biological mechanisms by which the continuity of life through heredity and evolution has been manifest. Lewontin (1974) has distinguished these two issues into "the fact of evolution" and "evolutionary theory". Evolutionary theory seeks to bring into focus the interactions within and between the various hierarchical levels of biological organization, and identify which biotic and non-biotic components have been important in shaping each level.

One area of evolutionary science, 'population genetics', seeks to characterize genetic variation in

populations and species and explore the forces that shape the distribution of this variation into the various nodal classes, i.e. demes, populations and species. In the course of introducing the research I have undertaken, I will briefly outline how evolutionary theory has developed into its current state, particularly with respect to the dynamics of genetic variation within and between populations of organisms.

1.1.1 Evolutionary change and its mechanisms

Evolution was hardly a novel concept at the time that Charles Darwin's "Origin of Species" (1859) was published. Theories that the diversity of life was not unchanging has been an integral part of the entire history of human thinking about the world (a topic exhaustively reviewed by Mayr, 1982), and were implicit in the ideas of the great Greek philosophers (Osborn 1894).

Contemporary ideas of evolutionary change germinated with renewed interest in detailed study of the natural world during the Renaissance. It was from this that important developments in classifying organisms arose. Two of the most important are a consistent hierarchical system of taxonomic designation (Linnaeus), and the demonstration of internal structure as an important guide by which to make some of these groupings. However, all of these latter advancements were made in the absence of any notion that types of organisms change.

Ideas of change of living forms re-appeared in natural science in the 18th and 19th centuries. In the 18th century,

the weight of geological and biological observations led the French naturalist C. de Buffon to conclude that living organisms had undergone much alteration. However, he had no consistent mechanism by which this change could come about. In the 19th century, another French naturalist, J. de Lamarck proposed that species were continually changing to fit the challenges of their environments, and that the direction of this change was inherently toward increasing complexity. The mechanism for this change was the influence of the environment, perhaps due in part to the close association (adaptation) between the organisms and their environment. Thus arose the theory of evolution by the inheritance of acquired characters, the concept for which he is most famous (Jordanova 1984). Lamarck's theory was well-constructed and very influential. It represents the first formal evolutionary theory of the modern age.

It was Charles Darwin, however, that gave evolutionary theory its current foundations by providing a mechanism. Unlike Lamarck, who inferred that new variations in a species were directed by the demands of the environment, Darwin suggested that a pool of variation arose independently of environmental influences. Organisms would then leave fewer or more offspring according to the interactions of the environment and the variations they possessed, and therefore would be 'naturally selected'.

A fundamental component of this concept is the heritability of favourable variations. That this was the case

was made clear to Darwin through his interest in pigeon-breeding and the phenotypic changes that can occur through the action of artificial selection. Yet the exact mechanism of this heritability was unknown to Darwin. Blending inheritance, which was the most commonly accepted genetical theory at the time, would have resulted in the rapid dilution of adaptively significant variation. It was not until the rediscovery of Mendel's work on the particulate model of inheritance by Correns, deVries and von Tschermak in 1900 that it was possible to incorporate a genetical framework into Darwin's natural selection.

That incorporation was not without its difficulty, however. The heritable variation that could be seen to behave according to the framework suggested by Mendel was decidedly discontinuous, and this led to the belief of early Mendelian-era geneticists that evolutionarily important variants were infrequent and incorporated large phenotypic differences. However, phenotypic variability seen in nature was not of this sort, but continuous, subtle and ubiquitous. Thus, the nature of variation and its role in evolution became the point of contention between the early geneticists and the biometricians in England, and later on between geneticists and naturalists in North America. It required further discoveries by such workers as Johansen, Nilson-Ehle, East and others (Mayr 1982) about the genetical control over the phenotype (eg. epistasis, pleiotropy, polygenic inheritance) to ultimately produce a consistent model under which continuous

and discontinuous phenotypic variation could co-exist under the same Mendelian framework.

1.1.2 Population genetics and the abundance of genetic variability

The new emphasis on the population as a pool of discretely inherited yet interdependent genetic variation with enormous evolutionary potential led to a formal mathematical description of the dynamics of this variation within and between populations of organisms, and the implications for evolutionary change. At the forefront of this development of 'population genetics' were R.A. Fisher, J.B.S. Haldane, and S. Wright (Provine 1973). Their formulations resulted in a description of evolution as one rooted in changes of the frequencies of allelic forms of a gene.

With a workable mathematical description of population genetics in place, there eventually came about an integration of this model into the mathematical theories of systematics, paleontology, and of natural selection. Huxley (1942) referred to this fusion of models of evolutionary change at widely divergent levels of organization as "the modern synthesis" (a retrospective by the principal contributors can be found in Mayr and Provine, 1980). There are three basic ideas that result from this fusion. First, the ultimate source of phenotypic variability upon which natural selection acts is gene mutation and recombination of the different alleles. Second, natural selection acts to change the frequencies of

these allelic forms within populations, leading to gradual evolution of populations. It is significant that other influences such as gene flow, random genetic drift and non-random mating were ascribed a rather minor role. Finally, the processes at work to change allele frequencies at the population level are the same ones responsible for such higher level phenomena as speciation and the divergence between higher taxonomic groups. In other words, macro-evolution was seen to be an extension of micro-evolutionary advance.

However, what was still sorely missing at this point were any hard data concerning the actual amount of genetic variation that was present within and between populations. Are genes generally homozygous or heterozygous? Do populations show a lot of differences, or few? Without this information it was not possible to distinguish potential mechanisms of evolutionary change from actual mechanisms of such change, even at the population level, since the evolutionary dynamics of a population will to some extent depend on the amount of genetic variability present in it (Fisher 1930).

Early attempts to quantify the amount of variability in the genome were necessarily directed at genetically determined characters that, for one reason or another, could provide only very imperfect estimates. Morphometric traits, for example, could be determined by a few genes or many, any or all of which could be heterozygous; environmental effects also play important roles in many phenotypic traits and are hard to control. A much better system, and the one around

which the field of evolutionary genetics was effectively established (Lewontin et al 1981), is the arrangement of genes on chromosomes. In Drosophila, the banding patterns found on the polytene chromosomes of larval salivary glands let Dobzhansky survey populations for discrete chromosomal inversion and duplication/deletion polymorphisms. Still, there was no way of knowing the degree of correlation between chromosomal and genic variation. As a result, hypotheses about the level of genetic variation that should be present in natural populations was wholly dependent on the set of assumptions as to the interactions between genes, the phenotype and the environment that were held.

In 1955, Dobzhansky characterized then-current thinking on this subject into two schools (Dobzhansky 1955). Based on Fisher's (1930) mathematical theory of natural selection and Muller's (1927, 1950) view of genes and the nature of mutations, the classical school envisioned populations as maintaining the 'best' allelic variant for any given gene by the action of 'purifying' selection. New mutations are generally held to be deleterious and are usually eliminated soon after being generated. Individuals would therefore be homozygous for most genes, and any gene would have one dominant allele with a few transient low frequency variants at any given time. Rare beneficial mutations would eventually displace the previously 'best' allele after many generation of natural selection. Meaningful evolution would therefore proceed only by the substitution of new beneficial

alleles at a number of loci in a population in a slow and gradual fashion.

The balance school differs sharply from the classical school in that it predicts that populations will harbour a great deal of genetic variation at any given time. The basic idea behind this model was first suggested by Sewall Wright (1931), who felt that strong gene interactions, frequency dependent selection, and non-selective forces (eg. drift and gene flow) had substantial effects on the direction of gene frequency change, and thereby of evolution. Wright suggested a subdivided population structure would lie at the heart of the evolutionary mechanism, with more successful populations influencing others by gene flow. Dobzhansky also believed that the amount of genetic variation would be substantial, but that it is maintained largely by heterosis. In this view, the high level of variability is part of the 'adaptive norm' of the population, rather than being a result of the interactions between populations as suggested by Wright.

The advent of protein electrophoresis made possible for the first time gene-by-gene precise estimates of genic variation (Harris 1966, Hubby and Lewontin 1966). It gave immediate support to one of the primary premises of the balance hypothesis, in that there is considerable variation in natural populations. Subsequent analyses of a great number of other species (reviewed in Nevo 1984) confirmed this.

1.1.3 Neutral theory and the effects of history

Both the balance and classical schools ascribe a major role to natural selection in adaptation and evolution. There is, however, considerable controversy about the extent to which the large amount of genetic variation is under the influence of natural selection and therefore pertinent to adaptation and evolution of the population. Kimura (1968) has argued that the amount of genetic variation observed in populations is simply too great to be ascribed to natural selection alone, balancing or otherwise. This led him, along with King and Jukes (1969) to propose that the majority of the observed genetic variation could just as easily be a result of the combined effects of the irrelevance of genetic variants to organismal fitness, and random genetic drift which would come about with fluctuating population size.

Most mutants would be harmful and would be eliminated by natural selection. Neutral mutations would be ultimately fixed or lost by random effects after a period of transient polymorphism. At intervals roughly equal to the inverse of the neutral mutation rate, a new variant would substitute for the previously dominant one. A consequence of this latter feature of neutral theory is that substitution will occur in a fashion roughly linear with time. The specific rates would, of course, differ with varying functional constraints placed on the macromolecule. Whether or not this prediction has been borne out is a matter of considerable controversy. Many believe that the apparent constancy of the rate of molecular evolution

(Zuckerkandl and Pauling 1965, Wilson et al 1977) is merely a result of averaging over a long period of time (Lewontin 1974, Fitch and Langley 1974), and that the observed variances of the rate are greater than that allowed under the neutral model (Gillespie 1986).

The neutral theory introduced a framework around which a mathematically rigourous model of the dynamics of molecular evolutionary change could be constructed. Yet many feel that the contention that most polymorphism in natural populations is being maintained by stochastic factors alone is wrong, and that the neutral theory is merely an excellent and essential 'null hypothesis' through which the types and strength of selective forces can be analyzed. The debate between the balance and classical schools which centered on the amount of genetic variation has now been shifted to a debate between 'neutralists vs. selectionists' which centers on the role of selection in evolution and is at the center of virtually all contemporary work. The debate, however, is not about which idea is absolutely right and which is absolutely wrong. Neutralists do not deny a role for natural selection, and selectionists do not deny that there is physiologically irrelevant genetic variation. Rather, it is a debate about which mechanism is playing a greater role in the overall level of variability that we see within and between populations and species.

Coming to some sort of conclusion as to the relative importance of the neutral or selectionist explanations is a

daunting prospect. To do so would require experimental demonstration of the absence of selection pressures on alleles of a locus under a wide array of possible environmental variables. A handful of enzymes and their alleles have been analyzed *in vitro* and *in vivo* for changing responses to controlled changes in environmental variables (reviewed in Koehn et al 1983; also White et al 1988, Place and Powers 1984), but these cannot be taken as a general result for all genetic variation. Realism restricts us to an 'educated guess' based on the predictions of the neutral theory about the abundance and distribution of allelic variability in species. Tests based on distribution of allozyme variation are hampered by the lack of information regarding the genealogy between variants. New models that utilize the relationship between inter-specific divergence and intra-specific polymorphism at the nucleotide level (Hudson, Kreitman and Aguade 1987) are promising, but are yet restricted to analyses of single loci.

A significant difference between the neutralist and selectionist explanations of genetic variability has to do with the effect of species and population history on observed patterns. Historical effects are those forces aside from mutation and natural selection that can have an affect on genetic variability and its distribution, such as colonization, recurrent migration, and effective population size.

Under the neutral theory, the abundance and distribution of neutral variation will be more closely tied

to species history than will variation under natural selection. This is because natural selection can counteract some of the effects of history. For instance, adaptive selection of populations to differing environments will maintain genetic differences between populations in the face of gene flow, which tends to reduce population differentiation (Slatkin 1987). The departure of variants of particular loci from historical influences will vary with the type and strength of selection pressures. In addition, neutral variants may depart from historical influences due to linkage to selected variants at other loci (Franklin and Lewontin 1970).

The susceptibility of neutral variation to historical influences offers an indirect way to assess the relative importance of natural selection on the genetic structure of geographical populations. If the history of a species can be independently ascertained, then the departure of genetic variants from that history can be assessed. One way by which the species history can be studied is to use a marker for which it can be assumed that natural selection does not affect the distribution of variants. Therefore, the population structure indicated by this 'neutral marker' can be used as a standard against which the other elements can be tested for influence by natural selection.

Clines of allozyme alleles are one of the strongest examples of selection and yet they can be explained by both selective and historical factors (reviewed in Rhomberg and Singh 1989). On the historical side, such clines are the

result of migration between two founding populations that differ strongly in allele frequency, such that equilibrium has not been reached. Alternatively, such clines can be simply due to a gradient in some environmental parameter, which is tracked by the locus as a result of selection. A neutral marker could distinguish between these possibilities in that a cline by migration would be expected with the neutral marker as well.

An important question, however, is what can be considered a neutral marker. Tests of nucleotide polymorphism in nuclear genes can do this, but the techniques with which large-scale sequencing of genes can be done in numbers needed for population genetic studies are still very new. There is, however, another genetic element that is considered, on the basis of theory and observation, to be a neutral marker. That element, in animals at least, is mitochondrial DNA (mtDNA). In the next section I will discuss mtDNA in terms of its application to evolutionary science.

1.2 Organelle DNA

It has been known for close to forty years that many key metabolic genes are inherited extrachromosomally. The first clear indication of this was the finding of Ephrussi (1949) that the factors determining the slow growing 'petite' phenotype of the yeast Saccharomyces cerevisiae were cytoplasmically inherited. With the discovery by Watson and Crick (1953) that DNA is the 'material of heredity', it was

soon inferred that the agent of eukaryotic cytoplasmic inheritance was the DNA found in organelles (ie. mitochondria and chloroplasts) (Sager & Ishida 1963, Chun et al. 1963).

For the next decade, organelle DNA was a curiosity in the field of 'biological chemistry', with most work focusing on physical properties of the molecule, including its mode of transcription (Aloni and Attardi 1971, Murphy et al 1975), base composition (Lagerkvist 1978), and conformational structure (Fauron and Wolstenholme 1976, Shah and Langley 1977). Uses of organelle DNA in evolutionary studies was largely restricted to using these physical properties to shed light on the origins of the different organelles themselves. Yet some work on interspecific divergence was accomplished using heteroduplex reannealing of denatured DNA (Wolstenholme et al 1979, Dawid 1972). It was not until the technological advances of molecular biology that detailed information about the genetic content of organelle DNA could be gathered.

The features that unite all classes of organelle DNA are actually rather few in number (Brown 1985, Palmer 1985a). All are found within membrane-bound organelles in eukaryotic cells, and trace their origins to eubacteria that became endosymbiotic with primitive eukaryotic cells (Gray and Doolittle, 1982). All are inherited in a non-Mendelian fashion, are comparatively small relative to the nuclear genome (although there is an enormous difference in size ranges between different types of organelle DNA: 14 kbp to 34 kbp in animal mtDNA, 340 kbp to 2500 kbp in plant mtDNA), and

all harbour rRNA and tRNA genes that are used for transcription and translation of organelle encoded genes. There are strong differences in just about every other category, including gene content, gene arrangement, gene structure, genetic code, and mode of transmission from parent to offspring. A description of these characters in chloroplast DNA (cpDNA) and non-animal mtDNA has been reviewed in several places (Palmer 1985a,b, Sederoff 1984, Rochaix 1985). In the following section, I will briefly describe some of the features of animal mtDNA.

1.2.1 Features of animal mtDNA

Animal mtDNA is a covalently closed circular molecule, that has a normal size range of 15-20 kbp. Encoded within mtDNA are 22 tRNAs genes, 2 rRNA genes, and genes for 13 polypeptides that are involved in the electron transport system or ATP synthesis (Merten and Pardue 1981). All of the electron transport polypeptides are subunits of polymeric enzymes of which the other subunits are encoded in the nuclear genome. The mtDNA molecule is replicated from a single initiation point in a highly asymmetric fashion (Clayton 1982). Genes are transcribed as a single cistron, again from a single initiation point (Clayton 1984). mtDNA is found in multiple copies within each eukaryotic cell, with a few molecules within each mitochondrion. As far as can be deduced, animal mtDNA is inherited maternally (Gyllensten et al 1985, Lansman et al 1983b), and there is no inter-molecular

recombination (Moritz and Brown 1986).

Perhaps the most remarkable feature of animal mtDNA is the extreme structural efficiency of the molecule. The molecule consists of two regions: a coding region, where all coding elements are found, and a usually smaller non-coding region, which contains the site of initiation of replication but nothing else. Within the coding region there is virtually no inter-genic DNA, and genes do not contain introns. Genes are separated by tRNA genes, which are thought to also act as signals for transcript processing. Animal mtDNA has been referred to as the epitome of genetic economy (Attardi 1985).

Another remarkable feature of animal mtDNA is the evolutionary conservation of the molecule across taxa. This is manifest in several different respects. First, all animal mtDNAs are found in the covalently closed circular form. A striking exception is in two Hydra species, where the mtDNA is found as two linear molecules (Warrior and Gall 1985). Second, the gene content of the coding region is remarkably constant. Mitochondrial genomes contain far fewer genes than their eubacterial progenitors. In addition to simple loss of genes, there is strong evidence for gene transfer from the organelle genome to the nuclear genome (Palmer 1985a). It is unclear why organelle gene content is so consistent. In contrast, the non-coding region is quite variable in sequence content (Fauron and Wolstenholme 1980a,b). Animal mtDNA has a total size range of 14-34 kbp (Moritz et al 1987, Snyder et al 1987). Third, the order of genes within the coding region

is quite well-conserved. It was this observation that originally led to the suggestion that recombination was lacking in mtDNA (Borst 1972). The reason is probably due to the lack of inter-genic sequences in the mtDNA molecule, making intra-molecule rearrangement rather improbable. (Inter-molecular recombination could still occur but, because of uniparental inheritance and homoplasmy, it would be of no physiological or phylogenetic consequence). Nevertheless, there are a number of cases where gene order has changed (Moritz et al 1987). Usually, these are detected in inter-phylum comparisons, but intra-genus differences have been reported. One of these differences is within the genus Drosophila, as a single inversion has occurred in D. yakuba mtDNA, differentiating it from the mtDNA of other Drosophila species (Clary et al 1982).

The conservative nature of animal mtDNA is also manifest at another level, that of the sequence of the mitochondrial gene products. The amino acid sequences of mitochondrially encoded polypeptides are very well-conserved, reflecting a rather low rate of evolution (Brown 1983). This stands in strong contrast to the generally higher rate of nucleotide substitution in mtDNA over nuclear genes (Brown et al 1979). This information will be of considerable importance when the presumed neutrality of animal mtDNA sequence variants is considered.

1.2.2 Animal mtDNA and species phylogeography

The utility of organelle DNA to studies of organismal evolution was recognized in the mid-1970's, using animal mtDNA (Brown and Vinograd 1974). Certain key characteristics of animal mtDNA made it useful for such work (reviewed in Avise et al 1987). First, its covalently closed circular conformation makes it relatively easy to separate from other DNA. This makes the study of mtDNA analogous to being able to separate one set of tightly linked genes from each organism under consideration. Second, virtually all of the mtDNA copies within an individual have the same sequence (homoplasmic), since the oocyte complement are generally identical. Although some somatic mutation will occur during the growth of the organism, the frequency of any mutation will almost assuredly be small enough to preclude detection in restriction analysis.

Third, the lack of recombination results in the mtDNA being clonally inherited from mother to offspring, accumulating mutations with passing generations. The order of particular mutational events can therefore be traced, and mtDNA sequence variants can be grouped according to the sharing of particular mutational states. Inter-molecular recombination between mtDNA molecules within a single organelle may occur, but because of the homoplasmy in the oocyte, such recombinations would have no phylogenetic effect.

These properties were used by Wesley Brown and his colleagues to infer the phylogenetic relationships of primate species by assessing sequence change through the use of

restriction enzymes (Brown and Vinograd 1974, Brown et al 1979, 1982). One of the important discoveries from this work was that the rate of base substitution in animal mtDNA is often quite high, up to 10 times that of single copy nuclear DNA (scnDNA). This results in mtDNA sequence analysis having a much higher resolving power than other biomolecules. However, this also results in resolving power being lost for groups much above the species level due to repeated substitutions at the same sites. Despite this higher resolving power, Wilson's group was unable to resolve the human/chimp/gorilla trichotomy. The higher rate in mtDNA was thought to be due to an inefficient proof-reading mechanism in the gamma DNA polymerase, which replicates mtDNA (Clayton 1982). However, the rate of mtDNA substitution relative to scnDNA has since been shown to be highly variable among higher taxa (Vawter & Brown 1986), and the rate in Drosophila mtDNA has been shown to be not significantly higher than nuclear DNA (Caccone et al, 1988). The difference may be varying degrees of rate slowdown in scnDNA rather than an elevated rate in mtDNA (Vawter and Brown 1986).

Uniparental inheritance gives animal mtDNA a special utility in the study of intra-specific hybridization, particularly with respect to geographic zones of contact between closely related species. When such species are able to produce fertile hybrids, the fertile offspring are usually female (as is the case with Drosophila). Such hybrids will have mtDNA from the maternal species only, while the nuclear

genes will be mixed. If the hybrids backcross to the paternal species, the next generation will still have the mtDNA of the original maternal species, but the nuclear genes will be 75% from the paternal species. Further crosses to the paternal species will yield the apparent result of introgression of the mtDNA of one species into the other. This feature has been used to analyze characteristics of hybrid zones, including the size of the zone and the direction of hybridization. A particularly well-documented example is the long, narrow zone of contact in Europe between two sub-species of mice, Mus musculus domesticus which is found in western and southern Europe, and M. m. musculus which is found in eastern and northern Europe. Allopatric populations of the two sub-species have distinct mtDNAs. At the western end of the hybrid zone in Denmark, domesticus mtDNA has been found 750 km into the musculus range in mice shown to be musculus by diagnostic allozyme loci (Ferris et al 1983, Vanlerberghe et al 1988b). There was no introgression of musculus mtDNA into domesticus. At the eastern end in Bulgaria, the direction of introgression of mtDNA is also one-way, but is from musculus into domesticus (Vanlerberghe et al 1988a). Interestingly, there is introgression of nuclear alleles from domesticus quite deep into musculus territory. This studies indicate that the nature of hybrid contact zone can be quite heterogeneous. Mitochondrial introgression has also been shown for other species systems in Drosophila (Powell 1983), crickets (Harrison et al 1987), frogs (Spolsky and Uzzel 1984, 1986)

and deer (Carr et al 1986) among others.

One other feature of animal mtDNA is of key importance to this study. Sequence variants (or restriction variants) are widely considered to be selectively neutral. That is, they have no effect on organismal fitness. As is the case with other genes, the empirical evidence for the effects of selection on genetic variants is lacking, and therefore support for the notion is extremely cautious (Avise et al 1987). The evidence for and against selective neutrality will be taken up in detail in the discussion.

If intra-specific mtDNA variants can be taken to be selectively neutral, then the abundance and distribution of variants can be considered predominantly to be due to species history alone. It is therefore theoretically possible to use mtDNA to distinguish the historical component in order to assess the selection component in allozyme variation distribution.

1.3 The comparative population structure of closely related species: D. melanogaster and D. simulans

One of the basic tenets of the theory of evolution by natural selection is that natural selection utilizes the genetic variation available to it, but does not direct what variants are generated. Therefore, the response of a population to a selective challenge will, to a large extent, depend on the pool of genetic variability that it specifically has retained. As a result, the response of different species,

or even of different populations within a species, to the same environmental challenge may be quite different.

One way of assessing the relative importance of different genetic 'strategies' in adaptation is to compare population structures of different species. Usually such comparisons are made between species that are closely related, such that the range of physiological responses by the organism to the environment are as similar as possible (eg. Ayala and Powell 1972, Gonzalez et al 1982). It is also desirable that the geographic and climatic range of the species be similar, so that the selective pressures themselves would be similar. All of this serves to minimize the effect of external variables when making such assessments.

One such species pair that has been under intense scrutiny is the sibling species pair Drosophila melanogaster and D. simulans, perhaps the best known species of the melanogaster subgroup. They are two of a handful of species that have a cosmopolitan distribution (Bock 1980). As such, they have had to adapt to a wide array of environmental conditions, both tropical and temperate (Parsons 1975, 1983). How have these two species managed to adapt to such diverse environments? More specifically, how do the genetic mechanisms by which they have each accomplished this differ?

These questions have been actively pursued at several levels including comparative ecology, behavior, and morphology, as well as genetics (reviewed in Lemeunier et al 1986). The general finding of most of these studies has been

that there is less genetic variability in D. simulans than in D. melanogaster, both with respect to total genetic heterozygosity, and differentiation between populations. However, results from more detailed analyses at the molecular level are somewhat different. Choudhary and Singh (1987) observed that D. simulans is polymorphic at fewer allozyme loci than D. melanogaster, but is equally heterozygous at those loci. Coulthart and Singh (1988a,b) observed that some classes of genes, in particular those encoding polypeptides found in the glands of the male reproductive tract (which may play a significant role in the establishment of reproductive isolation, and therefore of speciation) are somewhat more polymorphic in D. simulans than in D. melanogaster. Finally, Aquadro et al. (1988) found 4-6 times more restriction polymorphism at the rosy locus in D. simulans than in D. melanogaster from a single population in the United States.

A number of hypotheses have been put forward to explain the differences in population genetic structure between these two species. Choudhary and Singh (1987) list the most likely options as either the adoption of a generalist genetic strategy to adaptation by D. simulans, as opposed to a specialist strategy by D. melanogaster, or to a much more recent expansion of D. simulans from its ancestral stock. Under a generalist strategy, the lack of population structure in the species is due to the active selection of a populational genetic composition that is adaptive under many different environmental conditions. In contrast, the reduced

geographic differentiation by means of recent colonization simply means that D. simulans is much further from a species-wide ecological equilibrium.

The goal of the present research is to help distinguish between these two possibilities by utilizing the neutral marker, mtDNA, to assess the comparative histories of the two species. Preliminary studies of D. melanogaster and D. simulans mtDNA indicated that the total variability of D. simulans mtDNA is significantly lower than in D. melanogaster (Shah and Langley 1979, Baba-Aissa et al 1984). However, there are no rigorous surveys of inter-populational distribution of mtDNA variation in D. melanogaster (Shah and Langley 1979, Reilley and Thomas 1980), and studies with D. simulans have been hampered by the lack of even a single high frequency polymorphism.

The experimental approach taken was to analyze many restriction sites in each mtDNA sampled in order to maximize the amount of observed polymorphism. This was particularly important in the case of D. simulans, for which very little restriction polymorphism had been identified. Specifically, I used restriction enzymes that cut DNA molecules often. Although Lansman et al. (1981) argued against this approach on the grounds of uninterpretable restriction patterns when the number of fragments was too high, unique attributes of Drosophila made it allowable. First, the bias toward A and T nucleotides reduced the total number of cuts and, second, the apparent degree of nucleotide diversity is low compared to

other higher animals. Approximately 100 restriction sites were sampled in each species per sample. Preliminary reports of this work in D. melanogaster have already appeared in print (Hale and Singh 1986, 1987).

2

MATERIALS AND METHODS

2.1 *Drosophila* Stocks

The strains of *Drosophila* used are listed in Table 2.1, with information on their origins and sources from which they were obtained.

2.2 Preparation and Maintenance of *Drosophila* Culture

All the strains were maintained as isofemale lines (i.e. established from a single, inseminated, wild caught female) under ambient atmospheric humidity, with a diurnal photic cycle (12 hr light, 12 hr dark) at 20°C. Culture stocks were maintained in groups of 25 x 95 mm glass vials on standard cornmeal medium (Table 2.2). A density of approximately 25-30 individuals per vial was usually maintained, with discrete generations (flies were changed every 14-21 days).

2.3 Line Sampling

mtDNA was purified from approximately 1 g of adult flies from each isofemale line used. Sampling cultures were started by taking adult flies from stock culture vials older than 21 days (so as not to interfere with stock culture maintenance), and placing them in a culture bottle with

Table 2.1

Iso-female stocks of *Drosophila* used in this study

Species/Population/ Line	Place	Date	Source
D. melanogaster			
France (12 lines)	Villeurbanne, France	1978	J.R. David CNRS, France
England (8 lines)	Covent Gardens, U.K.	1980	M. Kidwell U. Arizona
West Africa (9 lines)	Benin	1978	J.R. David
Central Africa (5 lines)	Brazzaville, Congo	1978	J.R. David
India (9 lines)	Varanasi, India	1983	B.N. Singh Baranas Hindu U., India
Korea (11 lines)	Seoul, Korea	1979	R. Paik Seoul Nat. U., Korea
Japan (10 lines)	Jume, Japan	1980	M. Kidwell
Taiwan (8 lines)	Taipei, R.O.C.	1977	F.-J. Lin Taiwan Nat. U., Taiwan
Vietnam (6 lines)	Ho-Chi-Minh City, Vietnam	1978	J.R. David
Australia (6 lines)	Fairfield, Australia	1980	P.A. Parsons LaTrobe U., Australia
British Columbia (6 lines)	Pt. Coquitlam, B.C., Canada	1983	A. Beckenbach S. Fraser U. Vancouver, BC
Ottawa (12 lines)	Ottawa, Ont. Canada	1978	G. Carmody Carleton U.

Hamilton (5 lines)	Hamilton, Ont. Canada	1977	R.A. Morton McMaster U.
Massachusetts (6 lines)	Amherst, Mass. U.S.A.	1978	D. Hickey U. of Ottawa
Florida (7 lines)	Miami, Fla., U.S.A.	1983	M. Kidwell
Texas (11 lines)	Brownsville, Texas, U.S.A.	1978	D. Hickey
California (6 lines)	Lakeside, Calif., U.S.A.	1982	M. Kidwell
Argentina (5 lines)	LaPlata, Argentina	1980	S. Coscaron LaPlata, Arg.
D. simulans			
France (12 lines)	Porquerolles, France	1983	J.R. David
Tunisia (10 lines)	Nasrallah, Tunisia	1983	J.R. David
Congo (14 lines)	Brazzaville, Congo	1983	J.R. David
South Africa (12 lines)	Cape Town, S.A.	1983	J.R. David
Australia			
(4 lines)	Melbourne	1983	J.R. David
(4 lines)	Sydney	1983	J.R. David
United States			
(1 line)	Amherst, Mass.	1984	M. Kidwell
(1 line)	Weynmouth, RI	1984	M. Kidwell
(1 line)	Beltsville, Md.	1985	J. Coyne U. of Chicago
(4 lines)	Raleigh, N.C.	1985	C.F. Aquadro Cornell U.
(1 line)	Miami, Fla.	1984	M. Kidwell
(5 lines)	Santa Ynez, Calif.	1988	A. Hoffman LaTrobe U.

Table 2.2

Composition of Cornmeal Culture Medium

Yellow cornmeal	60 g
Dried brewer's yeast	30 g
Agar	10 g
Malt powder	15 g
Sugar	15 g
Corn syrup	10 ml
Water	1 l
Tegosept solution: (10g p-hydroxybenzoic acid methyl ester + 100 ml 95% ethanol)	24 ml

cornmeal medium. Bottles were left for 3 days or until first instar larvae were visible. These adults were then transferred to a second bottle. A rolled kimwipe was added to each bottle to assist efficient pupation. Adult flies were collected from pair of bottles and frozen at -70°C until the bottles were exhausted. Flies were then kept at -70°C until use, or up to one year.

2.4 mtDNA preparation

2.4.1 Solutions

The sources of chemicals used in this study (sections 2.4, 2.5 and 2.6) are listed in Table 2.3 along with their abbreviations. Solution recipes are described in Table 2.4.

2.4.2 Procedures

The mtDNA isolation technique is essentially the differential centrifugation protocol of Lansman et al. (1981), with some modifications (W. Brown, pers. comm.).

Frozen flies were homogenized in 14 ml of cold homogenization buffer, with 1 ml of DDCA solution. The DDCA solution and the cold temperature are to inhibit nuclease activity throughout the preparation. All procedures, up until mitochondrial lysis, were carried out under cold conditions. An initial homogenization was done in a Dounce tissue homogenizer with a motor driven teflon-coated pestle, to assure that all flies were ground. The homogenate was then transferred to a homogenizer with a tight fitting glass pestle

Table 2.3

Sources of Chemicals

Chemical Name	Abbreviation	Source
Acrylamide (electrophoresis purity)	-	1
Agarose (type V: low EEO)	-	2
Ammonium persulphate	AP	1
N,N'-methylene bisacrylamide	Bis	1
Boric acid	-	3
Diethyl-dithio-carbamic acid (sodium salt)	DDCA	2
Cesium chloride	CsCl	4
Ethanol	EtOH	-
Ethidium bromide	EtdBr	2
Ethylene-diamine-tetraacetic acid (disodium salt)	EDTA	5
2-propanol	-	6
Sephadex (50-150 m)	-	2
Sodium chloride	NaCl	5
Sodium dodecyl sulphate	SDS	5
Sucrose	-	5
N,N,N,N'-tetramethyl ethylene diamine	TEMED	2
Tris(hydroxymethyl)aminomethane (electrophoresis purity)	Tris	1

1) BioRad, Richmond, Calif.
2) Sigma, St. Louis, Mo.
3) Caledon Labs, Georgetown, Ont.
4) Terochem Labs, Edmonton, Alta.
5) BDH Chemicals, Toronto, Ont.
6) Fischer Scientific, Fairlawn, N.J.

Table 2.4

Recipes of solutions

Homogenization buffer	
1.0 M Tris (pH 7.5)	100 ml
0.1 M EDTA (pH 7.5)	100 ml
Sucrose	85.6 g
H_2O to 1 litre	
STE	
1.0 M Tris (pH 7.5)	100 ml
0.1 M EDTA (pH 7.5)	100 ml
NaCl	1.9 g
H_2O	to 1 litre
TBE (10x stock)	
Tris	121 g
Boric Acid	61 g
0.1 M EDTA (pH 7.5)	100 ml
H_2O	to 1 litre
TE	
0.1 M Tris (pH 7.5)	10 ml
0.01 M EDTA (pH 7.5)	10 ml
H_2O	to 100 ml
EtdBr solution	
20 mg Etd/Br in 100 ml STE	
SDS solution	
10 mg SDS in 100 ml STE	
NaCl-loaded isopropanol	
5 M NaCl in H_2O; 200 ml added to 800 ml of 2-propanol. Shake and let stand overnight.	

to assure complete cell disruption. About seven strokes with the glass pestle were applied.

The homogenate was added to a 40 ml polystyrene centrifuge tube, and homogenization buffer was added to a total volume of about 30 ml. The homogenate was centrifuged in a Sorvall SS-34 rotor at 3,000 rpm for 5 min at 4°C. The supernatant was then transferred to a clean tube, and a second spin was done. These slow speed spins are meant to pellet larger body components (i.e. cell walls, and unbroken cells and nuclei), with the cytoplasmic constituents being left in the supernatant. The supernatant was then transferred to a clean tube and spun at 13,000 rpm for 20 min at 4°C. The supernatant was discarded, leaving a pellet that is enriched for cytoplasmic constituents (i.e. mitochondria). The pellet was then resuspended in 3.7 ml of STE. SDS solution (150 µl) was then added to lyse the mitochondria.

The volume of the lysate was adjusted to 4.0 ml. To this was added 4.2 g of dry CsCl. This solution then repeatedly inverted until all of the CsCl was dissolved. This solution was then added to a 13 x 51 mm polyallomer heat-sealable ultracentrifugation tube containing 200 µl of EtdBr solution. The tubes were topped up to the base of the neck with a CsCl in STE solution. The sealed tubes were then centrifuged in a Beckman Vti65.2 rotor at 55,000 rpm at 17°C for about 16 hours.

As mtDNA is a covalently closed circular supercoiled molecule, its density is greater than that of nuclear DNA,

and therefore bands at a different point on the isopycnic CsCl gradient. The exact point is between 5 and 10 mm below the prominent nuclear DNA band. Since the yield of mtDNA by this procedure is typically quite small, the mtDNA band was rarely seen. Therefore, all of the gradient between 5 and 10 mm below the nuclear band was collected. Using the side-puncture method, about 0.5 ml of the gradient was collected.

Ethidium bromide was eluted from the mtDNA samples by washing three times with an equal volume of NaCl-saturated isopropanol. mtDNA was then precipitated away from CsCl by adding two volumes of glass-distilled water, transferring to a 13 x 51 mm polyallomer tube, and filling with 95% EtOH. This was allowed to precipitate at -20°C for at least 3 hours (more typically overnight), after which the tubes were spun at 20,000 rpm for 30 min at 5°C in a Beckman SW50.1 rotor. The supernatant was poured off and the tubes were allowed to dry, inverted, at room temperature for about 2 hours. mtDNA was resuspended in 60 μl of TE and transferred to a sterile microfuge tube for storage at -20°C.

2.5 Restriction and Labelling of mtDNA

2.5.1 Restriction enzyme digestions

A total of 10 restriction enzymes were used in this study. Table 2.5 lists them with their recognition sequence and supplier. The 'r-value' of an enzyme refers to the number of bases in its recognition sequence. Enzymes with more than one recognition sequence (such as AvaII) may have non-integer

Table 2.5

Restriction enzymes used

Enzyme	Recognition sequence	r-value[a]	supplier[b]
For polymorphism survey:			
AvaII	G/G(A/T)CC	14/3	1
DdeI	C/TNAG	4	1
EcoRI	G/AATTC	6	2
HaeIII	GG/CC	4	2
HindIII	A/AGCTT	6	1
HinfI	G/ANTC	4	1
HpaII	C/CGG	4	2
MboI	/GATC	4	3
TaqI	T/CGA	4	1
XbaI	T/CTAGA	6	1
For survey of D. simulans mating incompatability only:			
ClaI	AT/CGAT	6	2
HpaI	GTT/AAC	6	1

a) Length of recognition sequence, according to Nei (1987)

b) 1) Boehringer Manheim Corp., Dorval, P.Q.
2) Bethesda Research Laboratories, Burlington, Ont.
3) Pharmacia Inc., Uppsala, Sweden.

r-values. All enzymes chosen were ones leaving either a 5' overhang at the cleavage site or a blunt end, to facilitate the labelling reaction. Digestion conditions were as recommended by the supplier, using the supplier's buffer when it was available, except for the following differences. First, because the yield of mtDNA was very low (approx. 1 μg per isolation), restriction digests typically had only approx. 0.1 μg of DNA. Also, 3-4 units of enzyme were used to assure complete digestion. Digestions were allowed to run for 2 hours, at which point either the labelling reaction was carried out, or the digested DNA was stored at -20°C until such time as the labelling reaction could be carried out.

2.5.2 Radioactive labelling of restricted mtDNA

Labelling of restricted mtDNA was done by the DNA polymerase I large fragment (Klenow enzyme) 'fill in' method. This enzyme has two pertinent activities: an efficient 5'-3' polymerase activity, and a less efficient 3'-5' exonuclease activity. When the restriction enzyme leaves a 5' overhang, Klenow labels the DNA by adding complementary bases to unpaired bases along the overhanging strand. An alpha ^{32}P-deoxyribonucleoside triphosphate (α^{32}P-dNTP) corresponding to the complement of the first unpaired base proximal to the 3' terminating stand was typically used. However, unlabelled dNTP's can be used if the available labelled dNTP is further distal along the overhang strand. When the DNA is blunt ended, the 3'-5' exonuclease activity takes out the first available

base on the 3′ terminating strand, upon which the polymerase activity fills it in with a labelled dNTP if one is available.

For mtDNA digested with 5′ overhang enzymes, labelling was done in the restriction buffer with 0.5 units of Klenow and 2 μCi of labelled dNTP (plus an equimolar amount of cold nucleotide, if needed) for 20 min at room temperature. The labelling reaction mixture added was kept to a minimum to maximize reaction efficiency. For DNA digested with blunt end enzymes, the conditions were the same except 1 unit of Klenow was used and the reaction was left for 1 hour. Both changes are to compensate for the lower efficiency of the exonuclease activity.

Background radioactivity of autoradiograms was shown to be reduced if unincorporated nucleotides were removed after the labelling reaction. The following method accomplished this. To each sample of labelled DNA, 75 μl of TE was added. The diluted DNA was then filtered through mini-columns of G25-150 Sephadex equilibrated in TE. The filtered DNA was transferred to a 500 μl microfuge tube which was then filled with 95% EtOH. The DNA was allowed to precipitate at -20°C for at least 2 hours, and then spun in a microfuge tube in the cold room for 20 min. The EtOH was poured off and the tubes were allowed to dry, inverted, at room temperature of at least 4 hr. The restricted, labelled mtDNA was then resuspended in 10 μl of TE.

2.6 Gel Electrophoresis

2.6.1 Agarose Gel Electrophoresis

2.6.1.1 Electrophoretic apparatus

Horizontal agarose gels were run in a plexiglas box (33 cm long, 22 cm wide, 9 cm high) with the central part (22 cm long, 22 cm wide) raised by 5 cm. Anodal and cathodal wires ran the width of the box to either side of the raised central part. Glass plates of 19 x 20 cm supported the gel. The pocket formers made pockets 2 x 7 mm.

2.6.1.2 Preparation of Gels

Gel concentrations ranged from 0.5% to 1.0%, depending on the enzymes used. Typically, gels of 0.5% were used with 6-cutter digested mtDNA, as most fragments were larger than 2.0 kbp. Gels of 1.0% were used with 4-cutter digested mtDNA, as many fragments were in the range of 2.0 to 0.5 kbp, the lower end of resolution of these gels.

The gel solution was prepared by adding an appropriate amount of high melting temperature agarose and boiling, with stirring, on a hot plate. The solution was allowed to boil until all of the agarose was dissolved. With the glass plate lying flat in the gel box and the pocket former in place, the somewhat cooled gel solution was poured directly onto the plate and allowed to spread to the periphery. Surface tension at the edge of the plate prevents spillover, underscoring the importance of correct pouring temperature. Any air bubbles were removed immediately with a pasteur pipet. This results

in a thin gel (approx. 4 mm thick) that allows sharp resolution of fragments.

After the gel was set, the box was filled with 1.5 litres of TBE buffer, the pocket former was removed, and the plate was secured.

2.6.1.3 Electrophoretic procedure

To the labelled mtDNA restriction fragments was added 1 μl of 10x loading dye (30% glycerol, 2.5 mg/ml bromophenol blue). Half of this (5.5 μl) was added to its gel pocket using a P20 Pipetman. Typically, each gel had 18 pockets. The size standard for agarose gels was HindIII-cut lambda-phage DNA fragments. Gels were run at about 75 V (25 mA) for about 7 hours, or until the bromophenol blue tracking dye was about 1 cm from the end of the gel.

2.6.2 Polyacrylamide gel electrophoresis

Vertical acrylamide gels were run in a plexiglas box (30 cm long, 14 cm wide, 18 cm high). The glass plate/gel sandwich was supported by a plexiglas stand with an anode tank at the top. Below the anode well on the stand is the cathode wire, which comes into contact with the tank buffer.

The glass plates are 20 cm wide by 23 cm long. One of each pair has a notch out of the top 17 cm by 3 cm, to allow gel contact with the anode buffer. The plates were spaced by 1.7 cm thick plexiglas strips.

2.6.2.2 Preparation of gels

Most gels were run according to the recipe of Maniatis et al (1982). Gels were generally 8.5% in concentration. 4.93 g of acrylamide and 0.17 g of bis-acrylamide were added to 60 ml of 1xTBE buffer. After the acrylamide was dissolved, the solution was filtered into a 125 ml buchner flask and placed under vacuum for at least 1/2 hour to degas the solution.

Denaturing gels were used in the later stages of this study. The gel recipe is 2.28 g acrylamide, 0.12 g bis-acrylamide, 28.6 g urea in 60 ml of 1xTBE. Otherwise, the same preparation procedures were followed.

Spacer strips were placed along the periphery of glass plate pairs (except for the top) and clamped together with metal buffalo clamps. The inside and outside edges were sealed by dripping a boiling 1% agar in water solution along the edges. This was allowed to cool for a few minutes before the gel solution was added.

Once the solution had been degassed, 250 μl of a 10% ammonium persulphate solution and 95 μl of TEMED was added to start polymerization. After swirling briefly, the solution was pipetted into the space between the glass plates to about 5 mm from the bottom of the notch on the notched plate. The pocket-former was inserted, being careful not to introduce air bubbles, and the gel was allowed to set for 6 hours to overnight.

After polymerization was complete, the buffalo clamps were removed, the bottom spacer strip was removed, and the gel

sandwich was attached to the vertical support with plastic clamps. The top anode tank was filled with 600 ml of TBE buffer, and buffer was added to the bottom tank so that it would immerse the cathode wire.

2.6.2.3 Electrophoretic procedure

5.5 µl of the labelled mtDNA fragments (with loading buffer) was added to each pocket with a P20 Pipetman. HpaII-cut pBR322 DNA is used as a size standard on acrylamide gels. Gels were run at approximately 250 V (20 mA) for about 4 hours, or until the bromophenol blue tracking dye was about 1 cm from the bottom of the gel.

2.6.3 Autoradiography of gels

Gels were vacuum-dried onto Whatman 3MM filter paper (drying apparatus from Eltech Designs), and wrapped with a single layer of plastic wrap. This was placed in a film cassette with a 20 x 30 cm piece of Kodak XAR-5 X-ray film. For agarose gels, a cassette with an intensifying screen was used. The dried gels were allowed to expose the film for approximately 48 hours, at which point the gel was removed and the film was developed in a Kodak automated film developing system.

2.7 Data handling

2.7.1 Display of restriction data

Variant restriction patterns are designated in the

text by appending an upper case letter to the enzyme name. For instance, Ava-A and Ava-B refer to the A & B patterns for enzyme AvaII. In this research, the differing cleavage patterns between conspecific haplotypes were sufficiently simple that precise mutational relationships could be established. Inference of the total number of restriction sites, both variant and invariant, in part takes advantage of the circularity of the mtDNA molecule. For circular DNA molecules there is a one-to-one correspondence of the number of restriction sites and number of fragments generated by a particular enzyme. Therefore, using one particular restriction variant as a reference point, restriction sites are either added or designated polymorphic as each variant is compared to it. Polymorphic sites were then assigned a lower-case letter designation, eg. Ava-a, Ava-b, etc.. Eventually the number of monomorphic and variable sites is identified over all haplotypes. Haplotypes can then be described by the presence/absence of all inferred variable restriction sites. Note that restriction mapping is not necessary, although map availability can assist in interpreting the mutational relationships between haplotypes.

2.7.2 Genetic distance measures

Measures of sequence divergence from the restriction site data are based on measures of 'genetic distance' between haplotypes (Nei 1987). There are several different methods by which to estimate genetic distances between variants and

populations from restriction site data (Ewens et al 1981, Engels 1981, Hudson 1982, Kaplan 1983, Nei 1987). These methods differ somewhat in the assumptions underlying such parameters as rates of substitution and frequency of each nucleotide, linkage between sites, and the effect of differing r-values in the enzyme complement. Comparisons between all these methods reveal that when the distance value is not large (ie. < 0.01), all of the methods yield very similar results. The methods of Nei (Nei and Li 1979, Nei and Tajima 1981, 1983, Nei 1982, 1987) are the most widely used in studies of animal mtDNA, and have therefore been adopted here.

The nucleotide distance measure (d_{ij}) estimates the number of nucleotide substitutions per nucleotide site between a given pair of haplotypes. Because the values of d_{ij} prove to be small, considerations of such factors as unequal rate of substitution between the four different nucleotides, and unequal frequencies of the four nucleotides in the DNA molecule, both of which have been observed in Drosophila, do not have a significant effect on the value of d_{ij}.

Estimates of d_{ij} can be calculated by either of two types of data: the proportion of shared restriction fragments, or by the presence/absence of specific restriction sites. While the latter yields a more reliable result, restriction site data often cannot be accurately calculated as differences in restriction patterns may be complex and precise mutational steps cannot be established between variants. Shared fragments have the difficulty that fragments of similar size may not

always be due to the same cleavage sites, and that fragments may differ by insertion/deletion of nucleotides and not be restriction site changes.

Calculating d_{ij} from restriction site data uses a maximum likelihood method that incorporates separate estimates from enzymes with different r-values. The formula (equation 5.50 in Nei, 1987)

$$\hat{p} = \hat{p}_n * \frac{\sum_n r_n (\hat{m}_n - m_{xyn}) / [\{1 - (1 - \hat{p}_n)^{rn}\} \{2 - (1 - \hat{p}_n)^{rn}\}]}{\sum_n r_n \hat{m}_n / [2 - (1 - \hat{p}_n)^{rn}]}$$

incorporates the proportion of shared to unshared sites for each r-value (r_n) and iterates until a result stabilizes. $\hat{m}_n$ equals the average of the number of sites for that r-value, and m_{xyn} equals the number of shared sites between the two haplotypes for that r-value. The formula iterates by successively placing the calculated $\hat{p}$ value in place of $\hat{p}_n$ and re-calculating. Usually about five iterations gives a stable result. Summation in each iteration is over the r-values. A trial value of $\hat{p}_n$ must first be calculated using

$$\hat{p}_n = 1 - (m_{xyn} / m_n)^{1/r}$$

Once a value of p is reached, it is converted to d_{ij} by

$$d_{ij} = - 3/4 \ \ln (1 - 4/3 \ \hat{p})$$

(Jukes and Cantor 1969). This equation converts the value for the proportion of nucleotide differences to a proportion of nucleotide substitutions, but is usually only important when p is large. The standard deviation of each d_{ij} is calculated by first calculating a variance term for each r-value (equation 5.44 in Nei 1987)

$$V(\hat{d}_n) = \frac{9\,(1-\hat{p})^2\,(2-(1-\hat{p})^r)\,(1-(1-\hat{p})^4)}{2\,r^2\,m\,(3-4\hat{p})^2\,(1-\hat{p})^r}$$

and then applying these to get the total variance (equation 5.51 in Nei 1987)

$$V(d_{ij}) = 1 \,/\, \Sigma_i\,[1 - V(\hat{d}_i)]$$

The square root of $V(d_{ij})$ yields the standard deviation.

Calculating d_{ij} by the shared fragment method (Nei and Li 1979, Nei 1987) is conceptually similar to that used for restriction sites in that initial estimates are iterated until the result stabilizes (equations 5.53-5.55 in Nei, 1987). However, the formula does not incorporate the estimates from each set of enzymes. d_{ij} is calculated for each r-value and a simple weighted average is taken to give the final value of d_{ij}. The proportion of shared restriction fragments for each r-value is

$$F = 2\,m_{xy} \,/\, (m_x + m_y)$$

where m_{xy} is the number of shared fragments, m_x is the total number of fragments in haplotype x, and m_y is the total number of fragments in haplotype y. $F^{1/4}$ is used as a trial value of G in

$$\hat{G} = [\ \hat{F}\ (3 - 2\hat{G_1})\ ]^{1/4}$$

Calculated values of $\hat{G}$ are successively reintroduced as G_1, to iterate until a value stabilizes. $\hat{d}$ is calculated for this r-value by

$$\hat{d} = -\ (2\ /\ r)\ \ln\ \hat{G}$$

Once $\hat{d}$ is calculated for each r-value, an average is taken to yield the d_{ij} estimate.

Pairwise values of d_{ij} between haplotypes can then be used to calculate intra-populational diversity (equation (1) of Nei, 1982)

$$d_s = \Sigma_{ij}\ x_i\ x_j\ d_{ij}$$

where x_i equals the proportion of haplotype i in the population. Regional diversity (d_t) is calculated in a similar fashion, where x_i equals the total <u>weighted</u> proportion of haplotype i in the region (averaged over populations). Nucleotide distances between populations are calculated by

(equation 10.20 of Nei 1987)

$$d_{xy} = \sum_{ij} x_i \, y_j \, d_{ij}$$

where x_i equals the proportion of haplotype i in population x, and y_j equals the proportion of haplotype j in population y.

2.2.7.3 Construction of phylogenetic trees

As pointed out by Nei (1987), methods of reconstructing phylogenies and gene genealogies fall under two broad classes. Distance matrix methods utilize values of genetic distance (eg. d values) to find a minimum total distance topology, usually by a clustering algorithm (eg. UPGMA; Sneath and Sokal, 1973), which depends on a pairwise similarity averaging technique. Maximum parsimony methods generally utilize discrete two-state data (eg. the presence/absence of restriction sites) to find a topology that infers the fewest mutational changes. Parsimony methods are generally preferred when discrete restriction site data are available (Felsenstein, 1983), since distance matrix methods assume a constant substitution rate in all parts of the tree (although the maximum parsimony method can also fail if substitution rates differ greatly (Felsenstein 1978). The debate over the choice of the two types of methods is far from resolved (Sober 1983)).

For this research, a Wagner parsimony tree (Kluge and

Farris 1969) for the D. melanogaster mtDNA haplotypes was constructed using the MIX program of the PHYLIP package (v. 3.1), kindly provided by Joe Felsenstein. The assumptions of Wagner parsimony are that character state changes are independent of the state of any other character, character states can change in either direction (0 - 1 is as likely as 1 - 0), and that the ancestral states are not known (therefore the network is not rooted). The MIX program adds species one at a time and finds the 'best' tree after each addition. As such, the program is somewhat dependent on the input order of species. Local rearrangements are carried out after each addition of a species, and this helps alleviate the effects of input order. Several different runs of MIX, using different input orders, were done to reduce this even further. The data input was the matrix of presence/absence of restriction sites among all haplotypes observed.

3

RESULTS

3.1 Restriction variation of Drosophila mtDNA: technical notes

3.1.1 Enzyme selection

An important preliminary step in this research was establishing which enzymes would be most useful in providing the desired information with respect to intra-specific variation. Two properties are desirable in the restriction patterns that are generated. First, the patterns generated should be sufficiently polymorphic over the species to yield enough diagnostic information to make the analyses as unambiguous as possible. This property is balanced against the second, which demands that the polymorphism be simple enough to draw precise mutational relationships.

The ultimate choice of enzymes will depend on the level of sequence diversity in the species. If the level of diversity is comparatively high, then enzymes which cut at a few sites will be best, while if the level is low, then enzymes which cut often should be chosen. The expected number of cut sites is a function of the size of the recognition sequence (r-value) and the A+T:G+C composition of the sequence, the latter being a factor when the A+T:G+C ratio of the DNA being examined is significantly different from unity.

In order to determine which enzyme would be appropriate for this research, a series of eight enzymes with a variety of r-values and A+T:G+C composition in the recognition sequence were tested against a single isofemale line from all 18 D. melanogaster populations, and the six D. simulans populations available at the start of this project. The results are shown in Table 3.1. In D. melanogaster, the enzymes with r = 6 were all monomorphic. With the exception of HpaII, all enzymes with r < 6 revealed polymorphism. The enzymes with the highest expected number of restriction sites were the most polymorphic in this sample, and yet the patterns were simple enough to draw clear mutational relationships. Therefore, in addition to the four enzymes that revealed some variation, two more enzymes with r = 4 and equal A+T:G+C ratio were introduced to the enzyme regime. The four enzymes that were monomorphic in this initial survey were assumed to be monomorphic throughout the entire sample, and were not surveyed further.

In D. simulans, no enzymes revealed polymorphism, a finding that was not particularly unexpected (Baba-Aissa et al 1984). In order to increase the chances of finding polymorphism, all enzymes were retained and the two extra r = 4 enzymes were added.

3.1.2 Fragment mobility and analysis

The sizes of observed fragments for all variants of all enzymes for D. melanogaster and D. simulans mtDNA are

Table 3.1

Restriction variants observed in a preliminary survey of one iso-female line from each population of *D. melanogaster* and *D. simulans*

Population	EcoRI	HindIII	HpaII	XbaI	AvaII	HaeIII	TaqI	MboI
D. melanogaster								
Korea	A	A	A	A	A	B	A	B
Japan	A	A	A	A	A	B	B	A
Taiwan	A	A	A	A	A	A	A	A
Vietnam	A	A	A	A	A	A	E	A
Australia	A	A	A	A	E	B	A	A
India	A	A	A	A	A	A	E	A
Central Africa	A	A	A	A	A	B	B	A
West Africa	A	A	A	A	A	A	B	A
France	A	A	A	A	C	A	A	D
England	A	A	A	A	A	A	A	A
Ottawa	A	A	A	A	A	B	B	A
Hamilton	A	A	A	A	A	B	B	A
Massachusetts	A	A	A	A	A	B	B	A
Florida	A	A	A	A	A	B	B	A
Texas	A	A	A	A	A	B	B	A
British Columbia	A	A	A	A	A	B	A	B
California	A	A	A	A	A	B	B	A
Argentina	A	A	A	A	B	B	B	A

Table 3.1 (continued)

Population	EcoRI	HindIII	HpaII	XbaI	AvaII	HaeIII	TaqI	MboI
D. simulans								
France	A	A	A	A	A	A	A	A
Tunisia	A	A	A	A	A	A	A	A
Congo	A	A	A	A	A	A	A	A
Cape Town	A	A	A	A	A	A	A	A
Australia								
Melbourne	A	A	A	A	A	A	A	A
Sydney	A	A	A	A	A	A	A	A

listed in Appendix I and II, respectively. The number of fragments visualized by 1% agarose gels alone (Fig. 3.1) of MboI, TaqI, HinfI and DdeI were significantly less than the total expected. Therefore, 8.5% polyacrylamide gels (Fig. 3.2) were employed to resolve fragments less than 500 bp; the maximal resolution was about 25 bp. Using this two-gel system, virtually all fragments were visualized.

There is, however, evidence that many of the fragments visualized on the 8.5% acrylamide gels did not appear in their correct position. Analyzing the nucleotide sequence of a 13,215 bp portion of D. melanogaster mtDNA (deBruijn 1983, Clary and Wolstenholme 1985) (Fig. 3.3) using the University of Wisconsin Genetics Computer Group analysis package (Devereux et al 1984) revealed that the fragments between 500 bp (the upper limit of analysis) and 200 bp in size were not running on the 8.5% acrylamide gels according to the position predicted by the published sequence. That this was a property of the mtDNA, and not of the gel, was suggested by the observation that size standard DNAs, HpaII-cut pBR322 and HindIII-cut lambda, were in their expected positions on the gels. The principle difference between the size standard DNAs and Drosophila mtDNA has to do with the A+T:G+C ratio. Specifically, the marker DNAs have a ratio close to unity, while Drosophila mtDNA is 75-80% A+T in the coding region. Short A or T tracts (which are not uncommon in Drosophila mtDNA) can introduce secondary structure into the restriction fragments that may alter electrophoretic mobility. To

Figure 3.1

Agarose gel (1%) of MboI digests of D. melanogaster mtDNA. HindIII-cut lambda-DNA is the size standard; fragment sizes are denoted in the margin. Lanes 1,3,10: Mbo-D; Lanes 2,7,8: Mbo-A; Lane 6:Mbo-K.

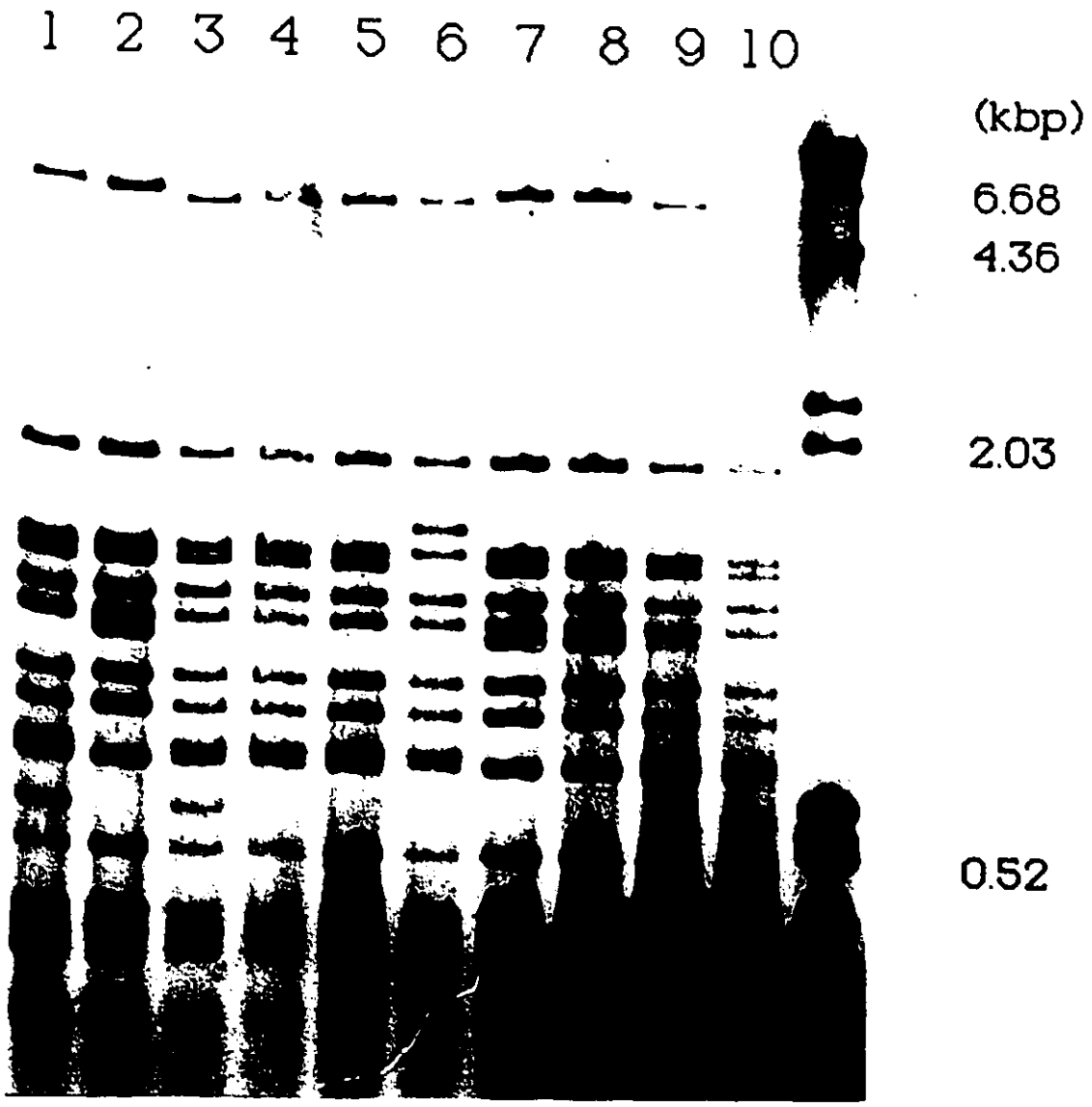

1 2 3 4 5 6 7 8 9 10
(kbp)
6.68
4.36
2.03
0.52

Figure 3.2

Non-denaturing acrylamide gel (8.5%) of MboI digests of D. melanogaster mtDNA. HpaII-cut pBR322-DNA is the size standard; the sizes of a few of these fragments are denoted in the margin. Lane 6 (Mbo-K) is the only pattern in this gel to show restriction polymorphism at this level of resolution.

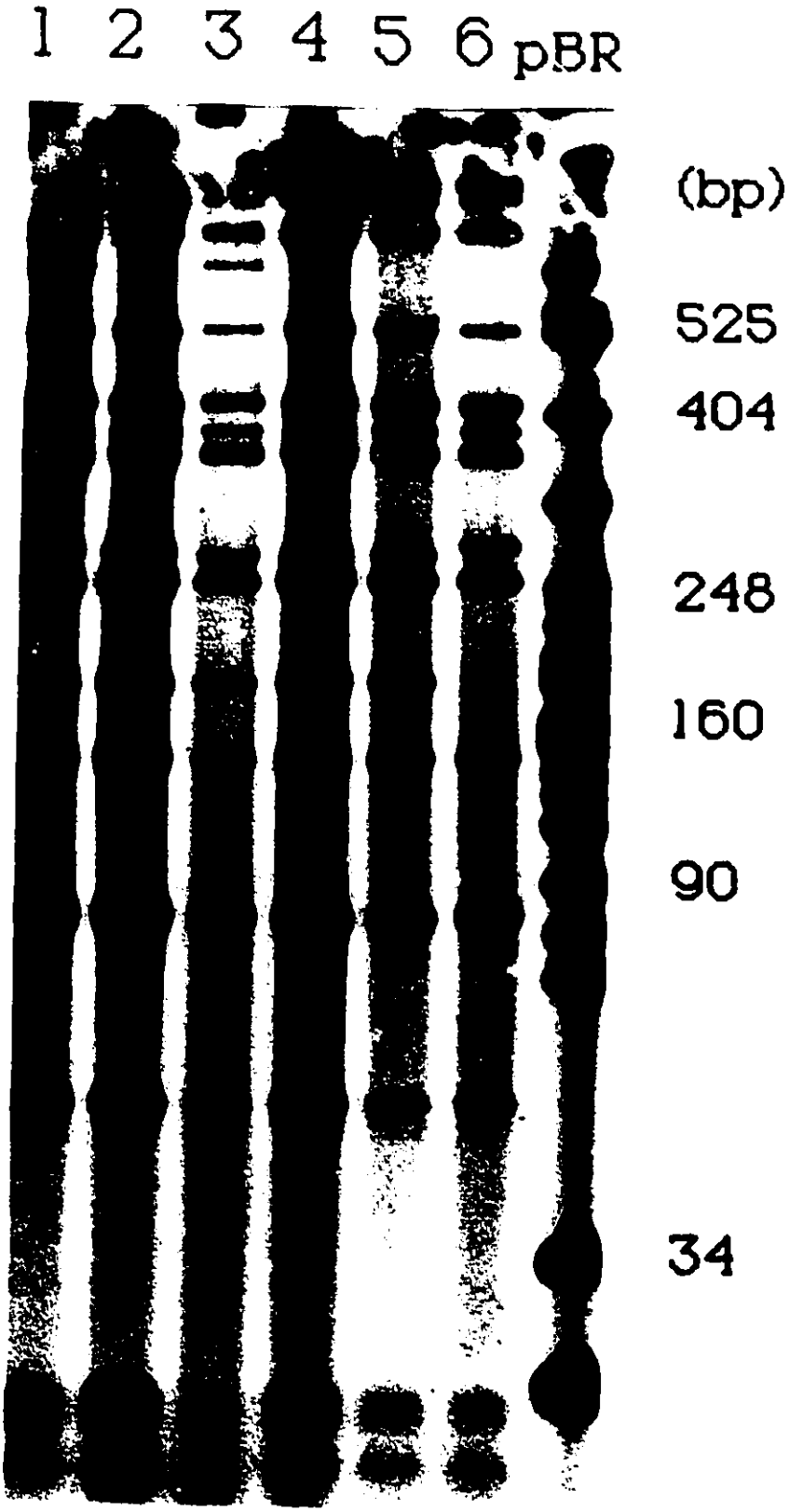
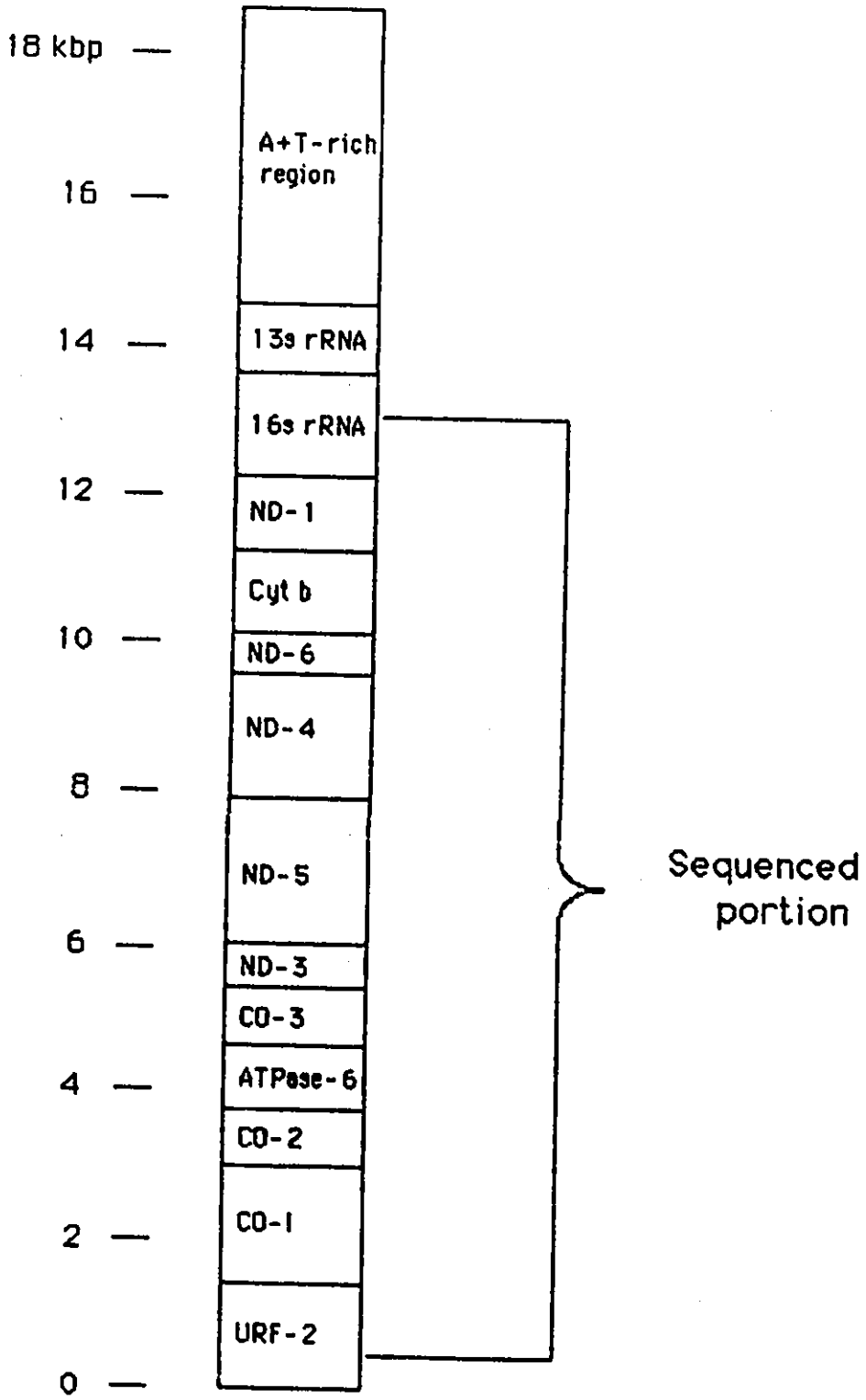

1 2 3 4 5 6 pBR
(bp)
525
404
248
160
90
34

Figure 3.3

Simplified genetic map of D. melanogaster mtDNA showing the bounds of the portion sequenced by deBruijn (1983) and Garesse (1988). Gene designations are: URF-2, unidentified reading frame 2; CO-1,2,3, cytochrome c oxidase subunits I,II,and III; ND-1,3,4,5,6, NADH reductase complex subunits; Cyt-b, cytochrome b. mtDNA is a circular molecule; the A+T-rich region and the URF-2 gene are contiguous.


18 kbp
16
14
12
10
8
6
4
2
0
A+T-rich region
13s rRNA
16s rRNA
ND-1
Cyt b
ND-6
ND-4
ND-5
ND-3
CO-3
ATPase-6
CO-2
CO-1
URF-2
Sequenced portion

determine whether such conformational effects are operative in these gels, I utilized 4.0% denaturing acrylamide gels for a small sample of lines from each species for each of the r = 4 enzymes in question. The fragments on these gels ran much closer to that predicted by the published sequence.

3.1.3 Frequency of restriction cleavage

The number of restriction sites observed for most enzymes were in keeping with the number expected (Lansman et al 1981) (Table 3.2). However, two non-A+T-biased r = 4 enzymes, HinfI and DdeI, have significantly fewer total restriction sites than expected ($p < 0.05$, t-test). These two enzymes differ from the others in that their recognition sequences have a non-specific base that interrupts the otherwise palindromic sequence. This base does not influence the calculation of expected sites. Analysis of the published sequence confirms that the mtDNA molecule does have fewer HinfI and DdeI sites. The depression in the number of sites may be due to a deleterious conformational effect on the mtDNA by interrupting the palindrome, but any possible reasons for this are not obvious.

3.2 Size variation of *Drosophila* mtDNA

As has been indicated earlier, intra-specific variation of restriction patterns is generated not simply by nucleotide substitution and the resultant gain/loss of restriction sites. Insertion/deletion of sequences of variable

Table 3.2

Expected and observed numbers of restriction sites for each enzyme in D. melanogaster and D. simulans

Enzyme	r value	expected	non-variable[a]	variable[a]	average[a,b]
AvaII	14/3	2.7	2/6	4/0	5/6
HaeIII	4	3.6	3/3	3/0	3.2/3
DdeI	4	32.1	10/11	2/0	11/11
EcoRI	6	4.5	4/8	0/0	4/8
HindIII	6	4.5	4/4	0/0	4/4
HinfI	4	32.1	14/14	4/3	16.2/15.5
HpaII	4	3.6	4/3	0/0	4/3
MboI	4	32.1	25/23	5/0	27.5/23
TaqI	4	32.1	31/25	2/0	32/25
XbaI	6	4.5	4/4	0/0	4/4

a) - The left side of each slash indicates the number of sites in D. melanogaster, while the right side of each slash indicates the number of sites in D.simulans.

b) - The average of the total number of sites (both variable and non-variable) occuring in each variant of the enzyme in each species.

length have been documented for mtDNA of many different organisms (Powers et al 1986, Bermingham et al 1986, Wallis 1987, Wolstenholme and Dawid 1968).

Polymorphism of this sort falls into two classes, both of which were observed in this research. The most frequent class concerns sequence insertion/deletion in the non-coding region that contains the start site for transcription. In non-insects and non-Drosophilid insects, this is the D-loop region, so named because of its secondary structure. In Drosophila, this region is strongly A+T-rich (95%; Klukas and Dawid 1976). The scale of size variation in the non-coding region can be anywhere from a few base-pairs to a few thousand base-pairs, and will be referred to here as 'large-scale' size variation.

The other class of possible size variation is in the coding region (Aquadro and Greenberg 1983). This class is rarer because of the extremely compact organization of animal mtDNA, with spaces between genes punctuated by the presence of tRNA genes. Also, mtDNA genes do not contain introns. Clearly therefore, there would be little chance for insertion or deletion outside the A+T-rich region that would not disrupt one of the gene products encoded by mtDNA. Drosophila mtDNA, however, does appear to have somewhat more inter-genic sequences than that of most other animals (Brown 1985), but it is still very little. The scale of coding-region size variation is only a few base-pairs, and will be referred to as 'small-scale' size variation.

3.2.1 A+T-rich region size polymorphism

There are three features that distinguish length polymorphism from restriction site polymorphism. First, there is no gain/loss of small restriction fragments (ie. only one fragment is altered between any two variants). Second, the length polymorphism will usually be apparent in digests with all restriction enzymes. One fragment in each digest will shift to the same extent. Third, the affected fragment from each enzyme will overlap on the restriction map. All of these criteria were met in observing large-scale size variation in this research. Fig. 3.4 depicts a simplified restriction map showing the fragment in which the large-scale variation was evident. In each case, the affected fragment contained the A+T-rich region.

The easiest enzyme with which to analyze the large-scale polymorphism in D. melanogaster is HindIII. There are HindIII sites that closely flank the ends of the A+T-rich region, leaving about 700 bp between each HindIII site and the edge of the A+T-rich region. The A+T rich region containing HindIII fragment is the second largest HindIII fragment; the largest and third-largest fragments lie close enough on agarose gels to act as a very good standard by which to observe size variation in the affected fragment. HindIII digests of size variant mtDNA are shown in Fig. 3.5.

In the agarose gels used, a resolution of about 0.1 kbp was possible in the region where the size variant fragment occurred. A total of 13 size classes were observed, spanning

Figure 3.4

Restriction map of *D. melanogaster* mtDNA as described by previous studies (Shah and Langley 1979, Fauron and Wolstenholme 1980, Reilley and Thomas 1980) for four restriction enzymes: E=EcoRI, H=HaeIII, N=HindIII, P=HpaII. The fragments that show the large scale size variation are shown below: a) HpaII, b) HaeIII, c) EcoRI, d) HindIIII

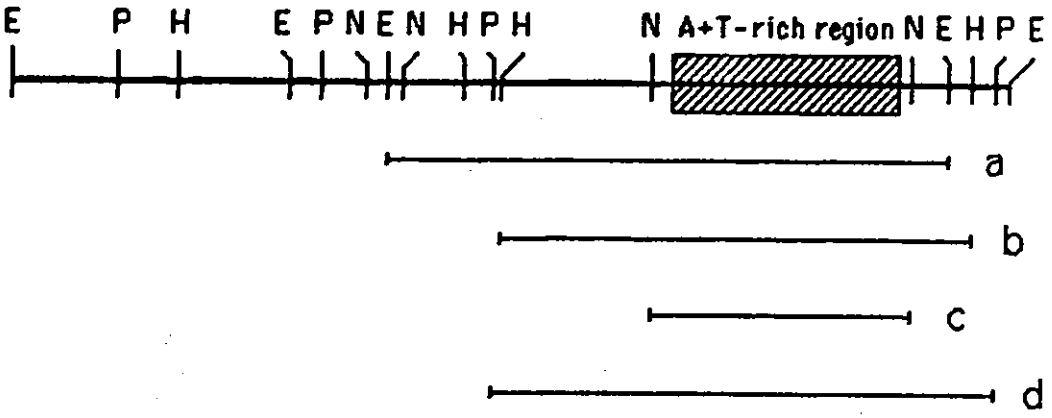
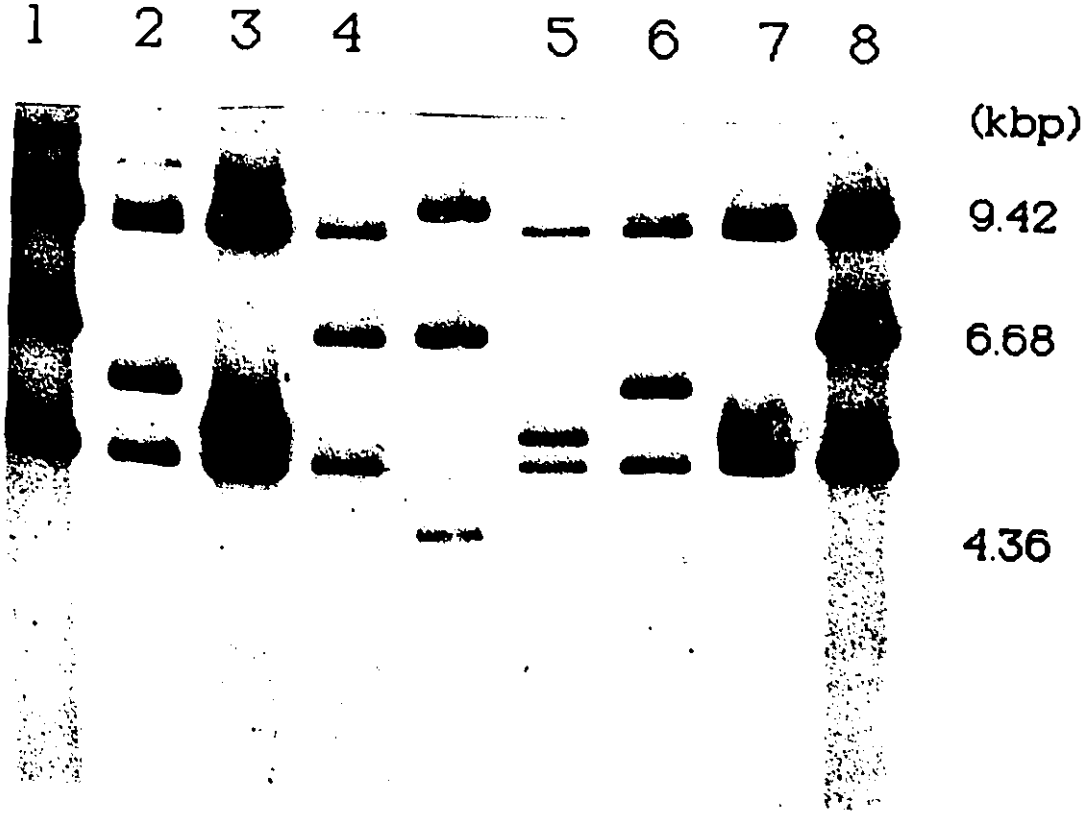

E P H E P N E N H P H N A+T-rich region N E H P E
a
b
c
d

Figure 3.5

HindIII digests of D. melanogaster mtDNA showing size variation in the fragment containing the A+T-rich region (Fig. 3.12). HindIII-cut lambda-DNA is the size standard; fragment sizes are denoted in the margin. Lanes 3, 5 & 7, 18.6 kbp size class; lanes 2 & 6, 19.1 kbp size class; lanes 1, 4 & 8, 19.9 kbp size class. The gel is 0.6% agarose.


1
2
3
4
5
6
7
8
(kbp)
9.42
6.68
4.36

a range of 1.7 kbp.

Table 3.3 shows the distribution of size variation among the observed restriction haplotypes (Table 3.4) for those lines to which a size class(es) could be precisely assigned. A striking observation is that all haplotypes which occur more than once in the entire sample vary in size. In addition, those haplotypes that were observed more often demonstrated more size classes, For example, haplotype #7 was the most frequently observed haplotype and revealed more size classes (nine) than any other haplotype. Haplotypes #1 and #2, which were each the second most frequent haplotype, had six and five size classes, respectively, among 13 lines each.

An important consequence of this observation has to do with the rate of restriction site gain/loss relative to sequence insertion/deletion. If each restriction site haplotype can be assumed to have a unique evolutionary origin, then these data would support the notion that the size variants in each haplotype had multiple origins and are therefore being generated quite regularly. Also, since the number of size classes are restricted to the limit of resolution of the gel, these size classes may be (and probably are) heterogeneous, and therefore the number of different size classes may have been sharply underestimated. One of the unfortunate consequences of this apparent high rate of size mutation is that the size variation is of little use in assessing phylogeny and population structure. As can be seen in Table 3.5, every population (except Japan) harbours a great

Table 3.3

Number of isofemale lines (homoplasmic/heteroplasmic) showing A+T-rich region size variation in different restriction haplotypes in *D. melanogaster*

	Size class (kbp)													
Haplotype #	18.2	18.6	18.7	18.8	18.9	19.0	19.1	19.2	19.3	19.4	19.5	19.6	19.9	n
1		2/2	0/2		1/0	1/1	3/2	0/1	2/0					13
2	1/0	9/2		0/1		1/0	0/1							13
3		1/0												1
4		0/1				3/0	2/1							6
5		1/1		2/0			0/1			1/0				5
6											1/0			1
7		29/6	12/2	1/3	1/2	2/2	2/2		0/1	1/0	1/0		1/0	59
8		1/0					0/1	3/0		0/1				5
9		0/1					0/1						1/0	2
10		1/0					2/0							3
11			1/0											1
12		3/2	1/1	1/0			0/1				1/0			8
13											0/1		3/1	4
14				1/0										1
15		1/0												1
16									1/0					1
17		1/1		0/1								1/0		3
18		1/0												1
19				1/0										1
20	1/0													1
21														0
22		2/0	1/1	1/0	0/1									5
23							1/0							1

Table 3.4

Restriction variants observed for the six polymorphic enzymes in each haplotype in *D. melanogaster*

Haplotype #	AvaII	HaeIII	TaqI	MboI	HinfI	DdeI
1	A	A	A	A	B	B
2	A	A	A	A	F	B
3	A	A	E	A	A	C
4	A	A	E	A	A	B
5	A	A	E	E	B	B
6	A	B	A	A	A	B
7	A	B	B	A	A	C
8	B	B	B	A	A	C
9	C	A	A	D	C	A
10	C	A	A	D	C	B
11	E	B	A	A	A	B
12	E	B	A	B	C	B
13	E	B	A	B	D	B
14	E	B	A	K	C	B
15	E	B	B	B	A	C
16	A	A	A	A	A	C
17	A	A	A	C	B	B
18	A	A	E	E	B	C
19	A	B	A	B	A	B
20	A	D	A	A	B	C
21	D	C	A	C	B	B
22	A	A	B	A	A	C
23	A	B	A	B	B	A

Table 3.5

Number of isofemalelines (homoplasmic/heteroplasmic) showing A+T-rich region size variation in geographic populations in *D. melanogaster*

	Size class (kbp)													
Population	18.2	18.6	18.7	18.8	18.9	19.0	19.1	19.2	19.3	19.4	19.5	19.6	19.9	n
FRA		2/2	0/1	1/0	0/1		4/3	0/1					1/0	12
KOR		4/2					0/2				0/1		3/1	10
OTT		1/0	5/0	2/1	0/1	1/0	1/0			1/0				12
AUS		2/1	1/0	2/1										6
MAS		1/1	1/1			1/0				1/0			1/0	6
TAI	1/0	5/1		0/1		1/0								8
VIE		2/1	2/0				0/1							5
BC		1/1	1/0			0/1	1/0				1/0			5
CAF		1/1	1/0	1/0	0/1	1/0								5
JAP		10/0												10
ARG		1/0					0/1	3/0		0/1				5
CAL		3/0	1/0		1/0		0/1		0/1					6
FLO		4/0	2/1	0/1										7
ENG		2/1	0/1	1/0	1/0		1/0		1/0					7
IND		0/1				3/0	2/1							6
HAM		3/1	1/0			0/1								5
TEX		6/2	0/1			0/1	1/0				2/0			11
WAF	1/0	4/1	0/1	0/1			0/1		2/0			1/0		10

deal of size polymorphism, even when monomorphic for restriction site haplotypes.

Fig. 3.6 shows the frequency distribution of the size classes over the entire sample. The most important point is the sharp skewness of the distribution, as the shape departs significantly from that of a normal distribution ($p < 0.01$, t-test; Sokal and Rohlf 1981, pp. 114-116 and 174-175). The 18.6 kbp size class is by far the most frequent, accounting for over 40% of the genomes. Above this value, there is a sharp elongated drop-off in frequencies, except for smaller modes at 19.1, 19.5 and 19.9 kbp. There are only two lines below this value, with mtDNA sizes of 18.2 kbp.

Very few restriction sites are expected to occur within the A+T-rich region. Also, this region has proven extremely resistant to cloning (Garesse 1988). Consequently, finer analysis of the mechanisms of how size variation is generated is not available. However, work on size variation within the melanogaster species group suggests that a tandemly repeating unit may be largely responsible (Solignac and Monnerot 1986, Baba-Aissa et al 1988).

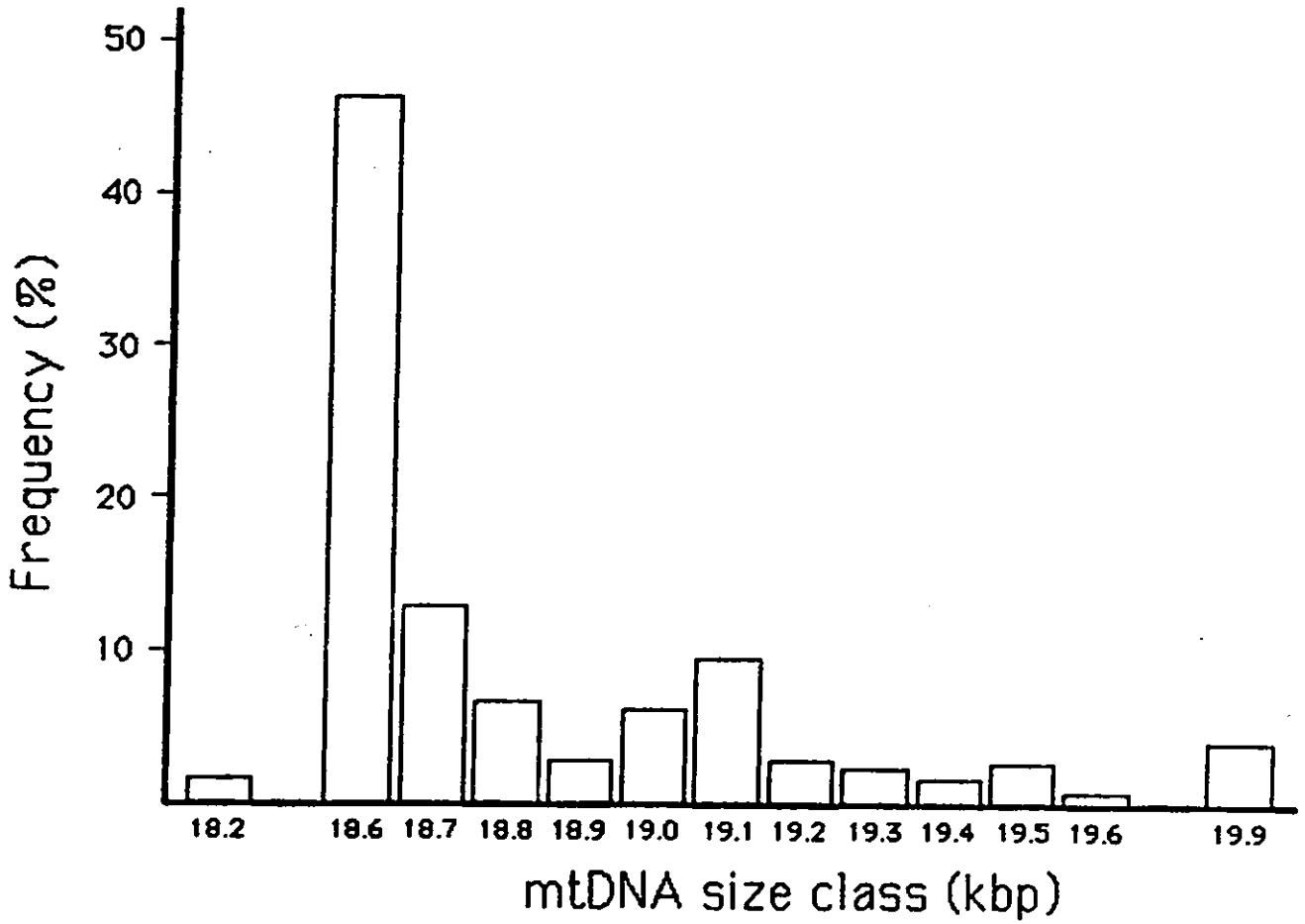
Virtually no A+T-rich region size variation was seen in the sample of D. simulans. Only a single line from France showed a somewhat larger A+T-rich region (Fig. 3.15); all of the rest fell into the 18.6 kbp size class.

3.2.2 Heteroplasmy of size variant mtDNAs

In a significant number of D. melanogaster lines (24

Figure 3.6

Frequency distribution of A+T-rich region size variants in D. melanogaster. In calculating frequencies, the variants in heteroplasmic lines were given half the weight of those in homoplasmic lines (Table 3.3).

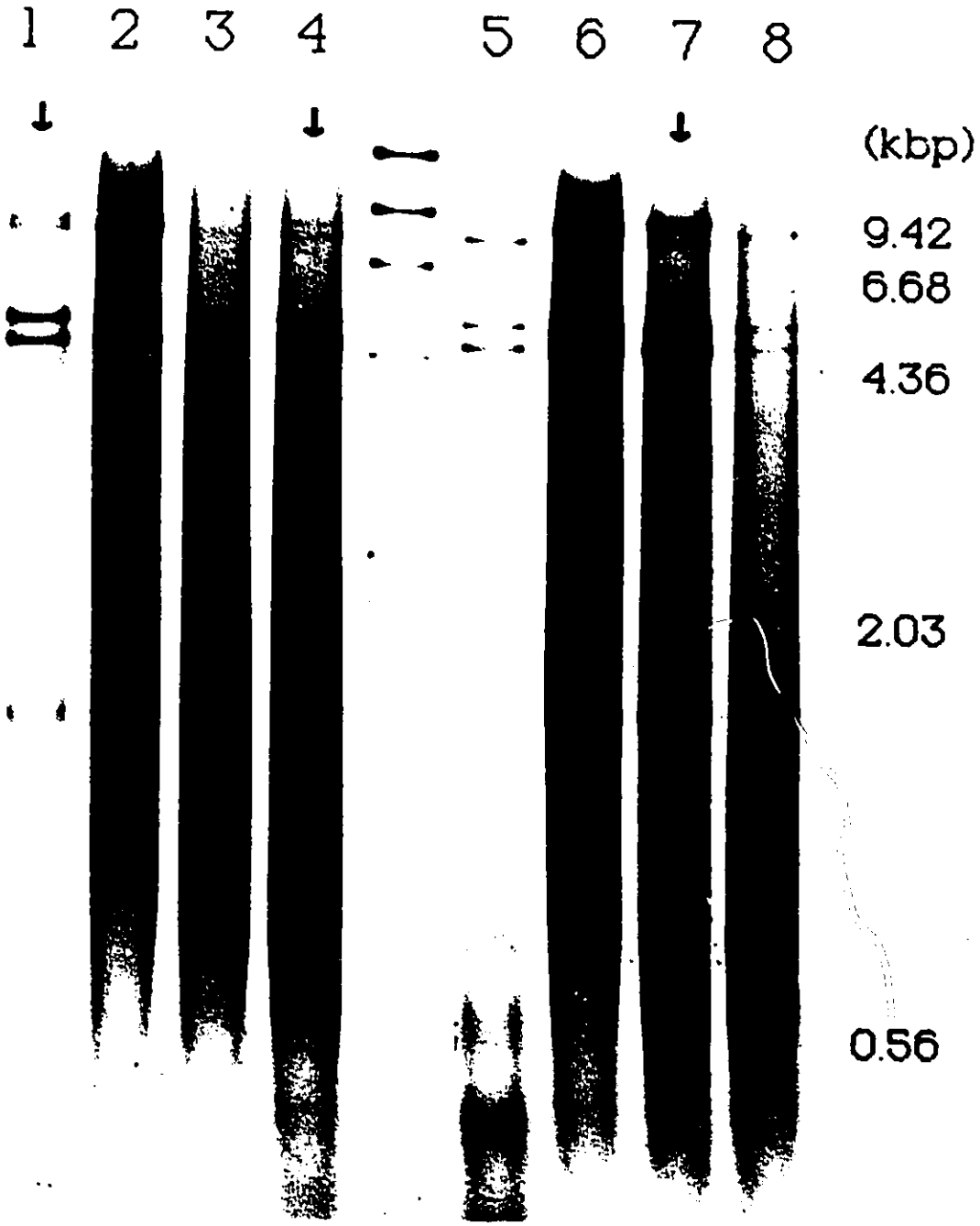

50
40
30
20
10
Frequency (%)
18.2
18.6
18.7
18.8
18.9
19.0
19.1
19.2
19.3
19.4
19.5
19.6
19.9
mtDNA size class (kbp)

of 144 = 16%), the mtDNA complement appeared to be heteroplasmic for two different size classes. This was inferred when the fragment containing the A+T-rich region was found, for all enzyme digestions, to exist as a doublet with each found substoichiometric to other fragments in the lane (Fig. 3.7). Incubation of mtDNA under conditions of five-fold excess restriction enzyme confirmed that incomplete digestion was not the reason.

The appearance of heteroplasmy can also be produced by paternal leakage of mtDNA, or by contamination of the stocks by flies harboring mtDNA of a different size class. Paternal leakage was not tested directly here since a properly conducted experiment would have required inordinate effort. The weight of evidence against paternal transmission in other animal systems would argue strongly against it in this case (Lansman et al 1983b, Gyllensten et al 1985). Since the lines in this study had been maintained in the laboratory for as long as five years before they were studied, contamination of lines seemed a more pressing concern. To test this, a sample of four putatively heteroplasmic lines were sub-cultured from single females for three generations to see if different size classes segregated. In every case, the sub-cultured flies were heteroplasmic with the same complement as the original line. This would seem to discount the possibility that the lines are simply contaminated, but are rather truly heteroplasmic. Also, no changes in the relative stoichiometry of the two mtDNA size classes occurred, suggesting that over the three

Figure 3.7

ClaI digests of *D. melanogaster* mtDNA demonstrating three cases of heteroplasmy for A+T-rich region size variants (lanes with arrows). HindIII-cut lambda-DNA is the size standard; fragment sizes are denoted in the margin. The gel is 0.6% agarose.


1
2
3
4
5
6
7
8
(kbp)
9.42
6.68
4.36
2.03
0.56

generations, the heteroplasmy was stable within experimental observation.

Of the 24 size heteroplasmic lines, 16 included the dominant 18.6 kbp class. These lines were heteroplasmic with size classes up to 19.1 kbp, which may be due to the gain of an extra 470 bp repeat unit. Lines heteroplasmic with the 19.1 kbp class also have mtDNA sizes both above and below 19.1 kbp. In one case, a line is heteroplasmic for two rather large mtDNAs (19.5 and 19.9 kbp), which may correspond to two and three copies, respectively, of the 470 bp repeat unit over the 18.6 kbp class.

There were no size heteroplasmic D. simulans lines observed in the present study.

3.2.3 Length polymorphism in coding region

Length polymorphism outside the major non-coding region in animal mtDNA is comparatively rare relative to that inside it. Insertion/deletions in regions where there is virtually no non-coding DNA is very likely to introduce frameshifts into a gene. Such mutants would ultimately be eliminated by organismal selection if it increased in frequency in the germ-cell lineage. What variation has been reported has been of very small size, usually on the order of 10 bp or less (Aquadro and Greenberg 1983). Such size polymorphism can only be detected, short of direct sequencing, by analysis of restriction fragments less than 500 bp in length, where the resolution of the gels are sensitive enough

to detect small mobility shifts. In this work, there were four separate cases of fragments that appeared to reveal such small-scale size variation.

The fragment most regularly demonstrating mobility shifts was an MboI fragment which on non-denaturing gels show an apparent size of 0.26 kbp. This fragment showed altered mobilities spanning the equivalent of 20 bp (Fig. 3.8a). All populations and haplotypes showed altered mobility in this fragment. Even the lines from Japan, which were monomorphic both for site haplotypes and for A+T large scale size variation, was highly polymorphic for this fragment. The maximum resolution in acrylamide gels at this size is approximately 3 bp; six different size classes could be discerned. The frequency distribution for the affected MboI fragment is shown in Fig. 3.9. Although the skew coefficient is significantly different than expected under a normal distribution ($p < 0.05$, t-test), it is not as strong as that seen for the A+T-rich region size distribution.

Since mtDNA fragments often have incorrect mobility in these non-denaturing gels, it is difficult to specify where on the restriction map this mobility polymorphism is occurring. (For purposes of clarity, references to fragments observed on the gels will be referred to as a 'fragment' and expressed in kilo-basepairs (eg. 0.26 kbp fragment), while references to fragments hypothesized from the sequence analysis will be referred to as a 'sequence' and expressed in terms of basepairs (eg. 260 bp sequence)). The sequence

Figure 3.8

Non-denaturing acrylamide gels (8.5%) showing examples of apparent size variation in small fragments. The sizes of all mtDNA fragments in these regions are denoted in the margin. The positions of those fragments showing small-scale variation are indicated in each lane by an arrow. A) Variation in the 0.26 kbp MboI fragment. Four size classes are shown. B) Variation in the 0.34 kbp TaqI fragments. A total of three different fragments show variation.

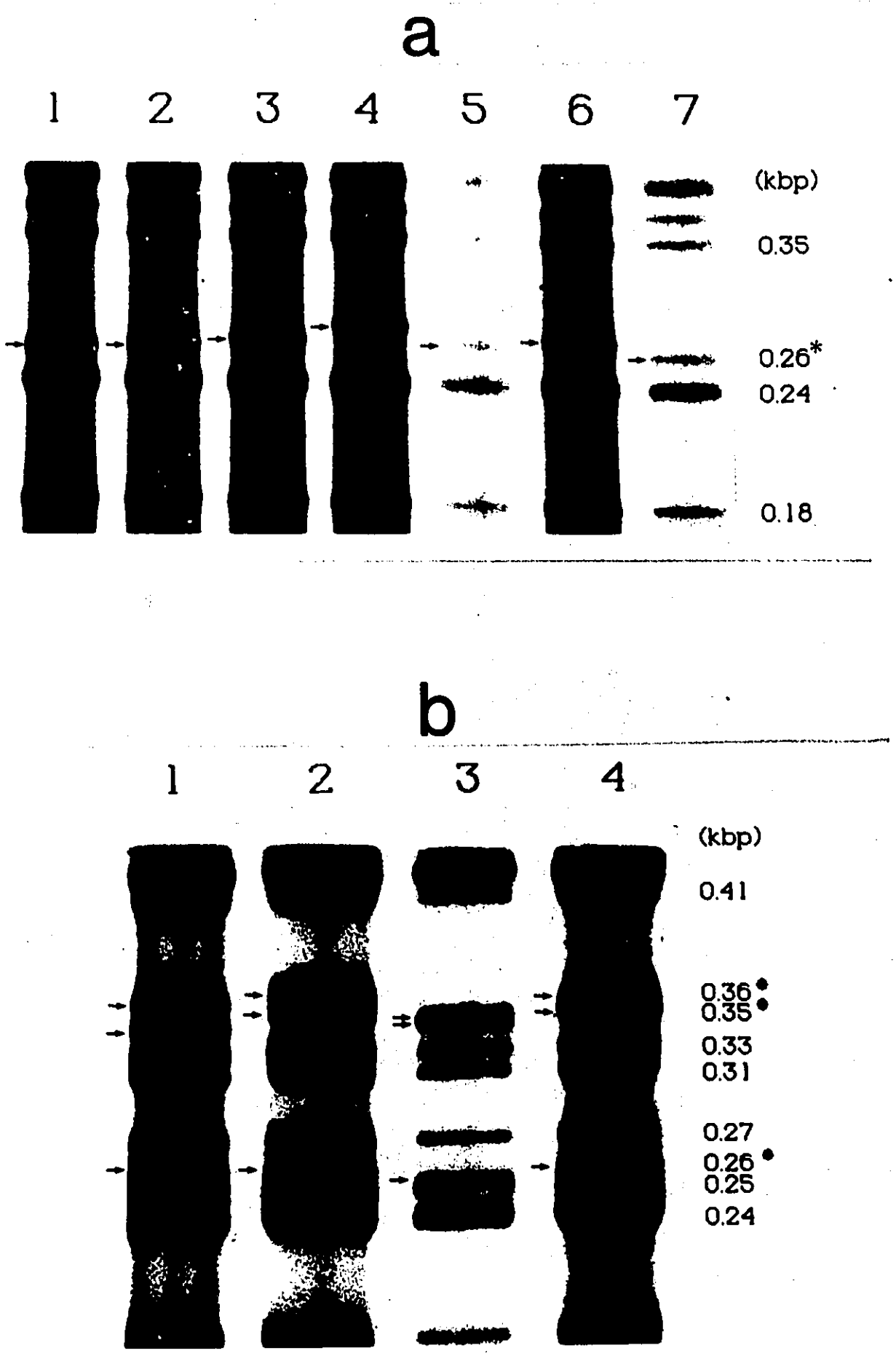
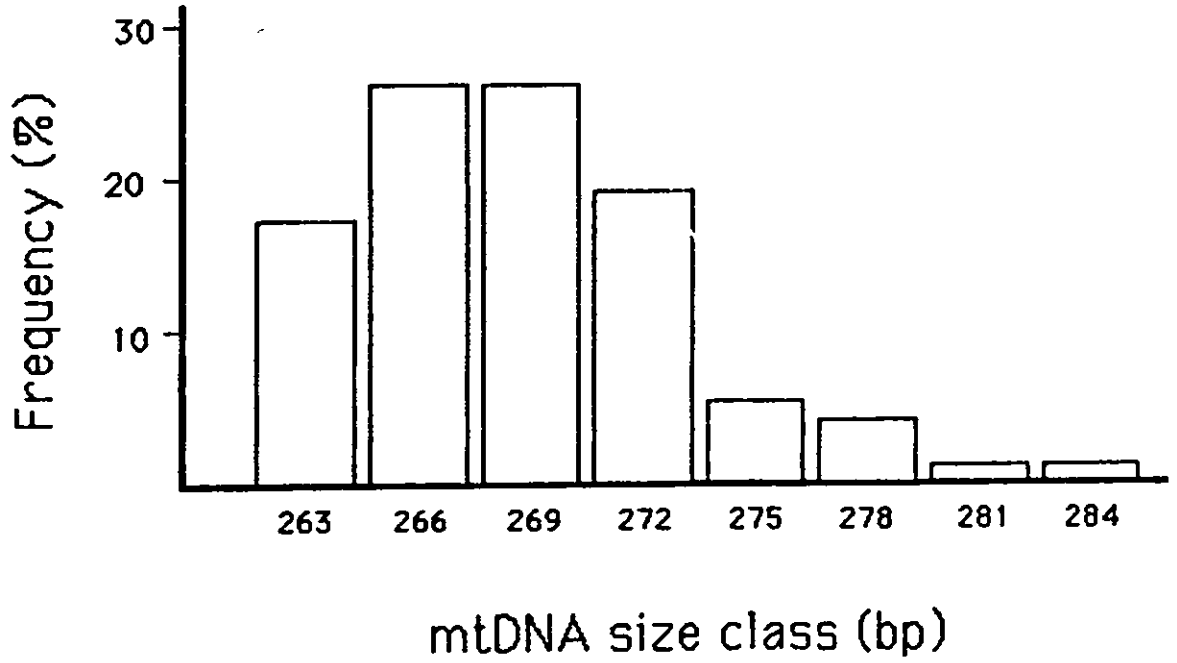

a
1
2
3
4
5
6
7
(kbp)
0.35
0.26*
0.24
0.18
b
1
2
3
4
(kbp)
0.41
0.36
0.35
0.33
0.31
0.27
0.26
0.25
0.24

Figure 3.9

Frequency distribution of apparent size variation in the 0.26 kbp MboI fragment. Size classes are based on the 'apparent' size of the fragment within the level of resolution, and is not intended to suggest the precision of the measurements.


30
20
10
Frequency (%)
263
266
269
272
275
278
281
284
mtDNA size class (bp)

analysis reveals three sequences that could potentially be visualized as an 0.26 kbp fragment (Table 3.6). Non-denaturing gels were unsuccessful at resolving this. Of the three sequences, only the 243 bp sequence contains any intergenic DNA into which nucleotide pairs could be repeatedly added/deleted without disrupting a gene product. The tRNA-ala gene lies wholly within this sequence, and there is a total of 35 bp of intergenic DNA at the termini of the gene (Fig. 3.10). The other two sequences lie within genes, and therefore have no intergenic DNA. It is tempting, therefore, to interpret the altered mobility of the 0.26 kbp MboI fragment as being the result of insertion/deletion into intergenic DNA within the 243 bp sequence.

As indicated earlier, one of the distinguishing features of size variation over site variation is that size variation will be revealed regardless of the restriction enzyme that is used. Therefore, there should be evidence for this small-scale polymorphism in the 0.26 kbp fragment in the digests of other enzymes, providing that it occurs in a fragment that is small enough to allow resolution on gels. The 243 bp MboI sequence is fully overlapped by a 330 bp TaqI sequence (Fig. 3.10). This size should allow resolution of the scale of polymorphism observed with the 0.26 kbp fragment. Therefore if the mobility polymorphism in the 0.26 kbp fragment is due to addition/deletion of DNA in the 243 bp MboI sequence, then a TaqI fragment should show concordance. However this is not the case. There are three TaqI fragments

Table 3.6

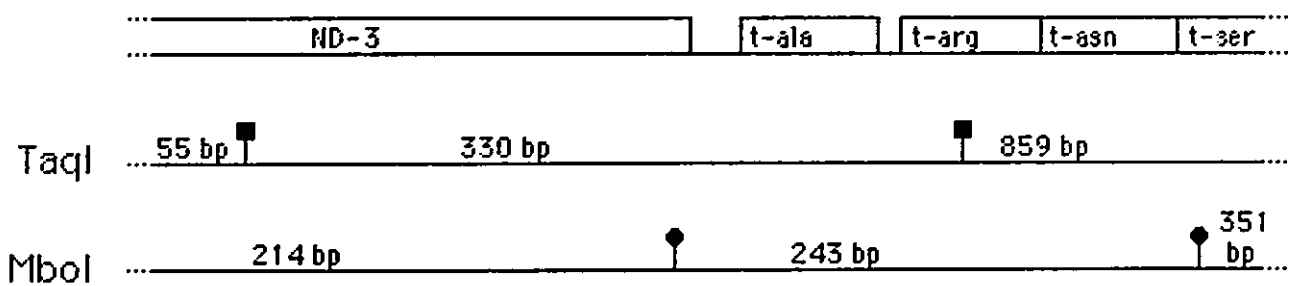
Restriction fragments of D. melanogaster mtDNA displaying mobility polymorphism in 8.5% acrylamide gels and surrounding fragments, and approximately equally sized sequences from sequence analysis

Enzyme	Observed (kbp)	From sequence analysis (bp)
MboI *	0.26	243
	0.245	234
	0.240	214
TaqI *	0.36	330
*	0.35	318
	0.33	311
	0.31	267
	0.27	261
*	0.26	243
	0.25	233
	0.238	231
	0.238	227

* - Fragment showing small-scale mobility polymorphism

Figure 3.10

Detailed restriction and genetic map showing the postion of the 243 bp MboI sequence and the 330 bp TaqI sequence relative to the genes in that region of the D. melanogaster mtDNA molecule.


ND-3
t-ala
t-arg
t-asn
t-ser
TaqI
55 bp
330 bp
859 bp
MboI
214 bp
243 bp
351 bp

that show some mobility polymorphism (Fig. 8b). But in each case the mobility variants of these fragments are very low frequency. Only eight lines show a low frequency mobility variant for any of the three TaqI fragments. Of these, four are observed in France, with one each in Korea, Ottawa, Japan and West Africa.

It would seem, therefore, that if the mobility polymorphism observed in the 0.26 kbp fragment were due to addition/deletion of DNA, it would have to occur in one of the other two similarly-sized MboI sequences with intergenic DNA to readily accommodate it. Alternatively, the mobility polymorphism could be due to conformational effects on the 0.26 kbp MboI fragment by base substitutions that result in slightly altered mobility of the fragment. Any overlapping TaqI fragment would have different conformational properties and sensitivities to base substitution, and therefore a concordance would not be expected.

The sizes of the TaqI fragments to show mobility variation are two at 0.34 kbp, and one at 0.26 kbp. There are approximately equally sized TaqI sequences that have some intergenic DNA: 330 bp (35 bp of intergenic DNA), 311 bp (29 bp), and 277 bp (20 bp).

3.3 Sequence divergence and evolution of mtDNA in *D. melanogaster*

Among the 150 lines of *D. melanogaster* studied, 23 composite restriction types (haplotypes) were identified.

These are listed in Table 3.4, along with their numerical designation.

A total of 104 restriction sites were observed with the six enzymes; if the four monomorphic enzymes are included, then the number of total restriction sites is 120. Of these, 20 sites were polymorphic. The distribution of monomorphic and polymorphic sites per enzyme is shown in Table 3.2.

3.3.1 Distribution of variant restriction sites

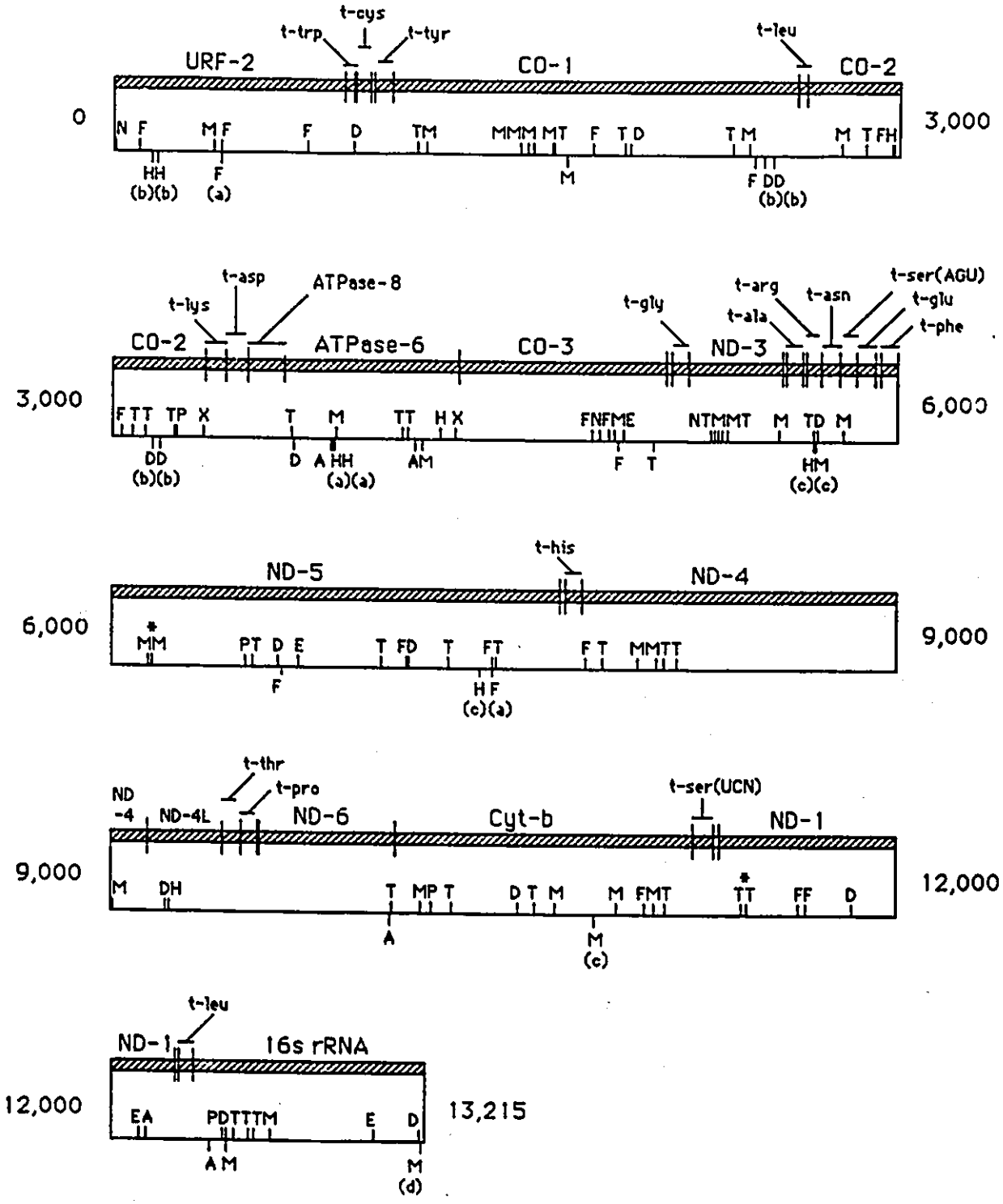
The published sequence allowed construction of a comprehensive restriction map that includes most of the monomorphic and variable restriction sites (Fig. 3.11).

On the basis of fragments observed down to the level of resolution of the acrylamide gels (20 bp) there were 100 monomorphic sites. The sequence analysis predicts the presence of two fragments below this resolution limit. The second site responsible for generating these unseen fragments have been included in the map (and denoted by an asterisk), but were omitted from the calculations of genetic distance since they were not observed directly and it could not be certain that they were not polymorphic.

Of the 102 monomorphic sites, 94 can be mapped within the 13,215 bp sequenced portion. In only one case is there uncertainty as to the position of a monomorphic site. Of the other eight sites determined to lie outside the sequenced portion, six necessarily lie to the 16s rRNA side, one necessarily lies to the URF-2 side, and one may lie to either

Figure 3.11

Restriction map for the sequenced portion of D. melanogaster mtDNA (Fig. 3.3). The 13,215 bp sequence is broken into four consecutive 3,000 bp sections with a remainder section. For each section, the genetic map is shown above the restriction map. Genes designations in the form 't-ala' refer to tRNA genes with the mitochondrial anticodon for the respective amino acid (in this case alanine). Other genes are as designated in Fig. 3.3. On the restriction map, monomorphic restriction sites are marked above the line, while polymorphic sites are marked below the line. Letters designate the respective restriction enzyme: A=AvaII, D=DdeI, E=EcoRI, F=HinfI, H=HaeIII, M=MboI, N=HindIII, P=HpaII, T=TaqI, X=XbaI. When the position of a variable site is uncertain, all possible positions within the sequenced portion are shown along with the small-case letter designation of the site (as in Table 3.6). Asterisks above two pairs of monomorphic sites indicates that only one was detected by electrophoresis with the other inferred by the sequence (see text).


t-cys
t-trp
t-tyr
t-leu
URF-2
CO-1
CO-2
0
3,000
t-asp
ATPase-8
t-lys
t-gly
t-arg
t-asn
t-ser(AGU)
t-glu
t-ala
t-phe
CO-2
ATPase-6
CO-3
ND-3
3,000
6,000
t-his
ND-5
ND-4
6,000
9,000
t-thr
t-pro
t-ser(UCN)
ND -4
ND-4L
ND-6
Cyt-b
ND-1
9,000
12,000
t-leu
ND-1
16s rRNA
12,000
13,215

side.

Mapping the polymorphic sites took some more careful analysis. Inferring a particular site to correspond to a particular mutational event was a seemingly simple matter of identifying the site which most closely produced the observed fragment sizes when present or absent. There was some ambiguity, particularly in cases when the causative sites were not present in the source molecule for the published sequence. Finding these sites required invoking an option in the GCG analysis program that allowed for a 1 base mismatch when searching for restriction sites. In this way, potential restriction sites are identified.

In total, eight of the 20 variable sites could not be unambiguously assigned to a particular location within the sequenced portion. One polymorphic site, Taq-b, was determined to definitely lie outside the sequenced portion to either side. One site, Mbo-d, was determined to lie either in the sequenced portion of the 16s rRNA gene, or to the URF-2 side of the sequenced portion. Two sites were determined to occupy one of two close-by sites within the same gene. Three sites were determined to occupy one of two possible sites in different genes. One site, Dde-b, could only be localized to four possible sites that lie in two different genes.

No sites were determined to have occurred within the A+T-rich region; only 2.5 sites were expected to occur within the region (based on the approximate minimum size of 4 kbp) over the complement of 10 enzymes.

There are too few sites to assess inter-genic comparisons of variability. However, one gene (ATPase-6) seemed to have a disproportionately high number of variable sites (5) relative to other genes. Inter-specific sequence divergence between D. melanogaster and D. yakuba, however, reveals that ATPase-6 (7.4% nucleotide diversity) is not more variable than surrounding genes (eg. CO-III, 7.4%; CO-I, 7.1%; ND-3, 7.9%) (Clary and Wolstenholme 1985, Garesse 1988).

3.3.2 Phylogenetic relationships between haplotypes

The data used to find the most parsimonious tree of D. melanogaster haplotypes are a 20 x 23 matrix that specifies the presence or absence of the 20 variable sites among the 23 observed haplotypes (Table 3.7).

Twenty different runs of the MIX program were done with different random number seeds (for haplotype input). Eight trees requiring the apparent minimum of 30 mutational steps were described. Two trees which differ only by the position of haplotype #6 (and thereby by the extra loss of one site over another) were produced far more often than the others (15 of the 20 runs produced only these two trees). Unfortunately, this is not evidence that these trees are 'better' than the others.

Two methods were employed to distinguish statistically between the various trees. The two trees appearing most frequently are depicted (as a single tree showing the alternative positions of haplotype #6) in Figs. 3.12 and 3.13

Table 3.7

Presence (1) / absence (0) matrix for variable restriction sites in D. melanogaster restriction haplotypes

Haplotype	AvaII abcd	HaeIII abc	TaqI ab	MboI abcde	HinfI abcd	DdeI ab
AAAABB-1	1111	100	10	01001	1100	10
AAAAFB-2	1111	100	10	01001	1000	10
AAEAAC-3	1111	100	00	01001	0100	11
AAEABB-4	1111	100	00	01001	1100	10
AAEEBB-5	1111	100	00	01101	1100	10
ABAAAB-6	1111	000	10	01001	0100	10
ABBAAC-7	1111	000	11	01001	0100	11
BBBAAC-8	0111	000	11	01001	0100	11
CAADCA-9	1100	100	10	11011	1110	00
CAADCB-10	1100	100	10	11011	1110	10
EBAAAC-11	1101	000	10	01001	0100	11
EBABCB-12	1101	000	10	11001	1110	10
EBABDB-13	1101	000	10	11001	1111	10
EBAKCB-14	1101	000	10	11000	1110	10
EBBBAC-15	1101	000	11	11001	0100	11
AAAAAC-16	1111	100	10	01001	0100	11
AAACBB-17	1111	100	10	00001	1100	10
AAEEBC-18	1111	100	00	01101	1100	11
ABABAB-19	1111	000	10	11001	0100	10
ADAABC-20	1111	101	10	01001	1100	11
DCACBB-21	1011	110	10	00001	0100	11
ABABBA-22	1111	100	11	01001	0100	11
ABABBA-23	1111	000	10	11001	1100	00

Figure 3.12

Wagner parsimony network for mtDNA restriction haplotypes of D. melanogaster giving detailed information on the particular restriction site changes. Haplotype #6 is in parentheses to denote that there are two equally likely positions, both of which are shown. The tree is unrooted and branch lengths have no meaning. The inferred restriction site changes between haplotypes are denoted by a single large case letter designated the enzyme (A=AvaII, D=DdeI, F=HinfI, H=HaeIII, M=MboI, T=TaqI), a lower-case letter designating the specific site (as in Table 3.6), and an arrow solely to indicate which direction the site is gained.

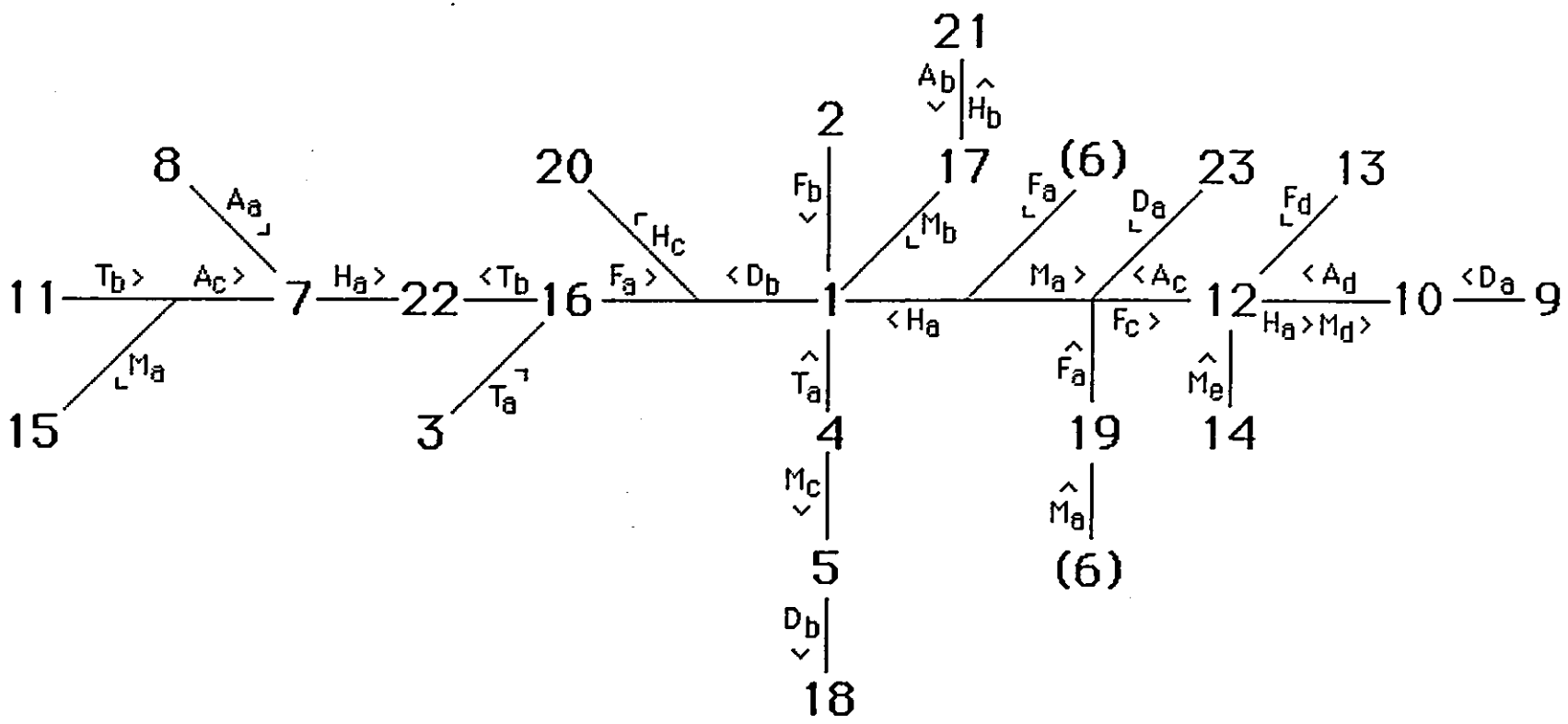
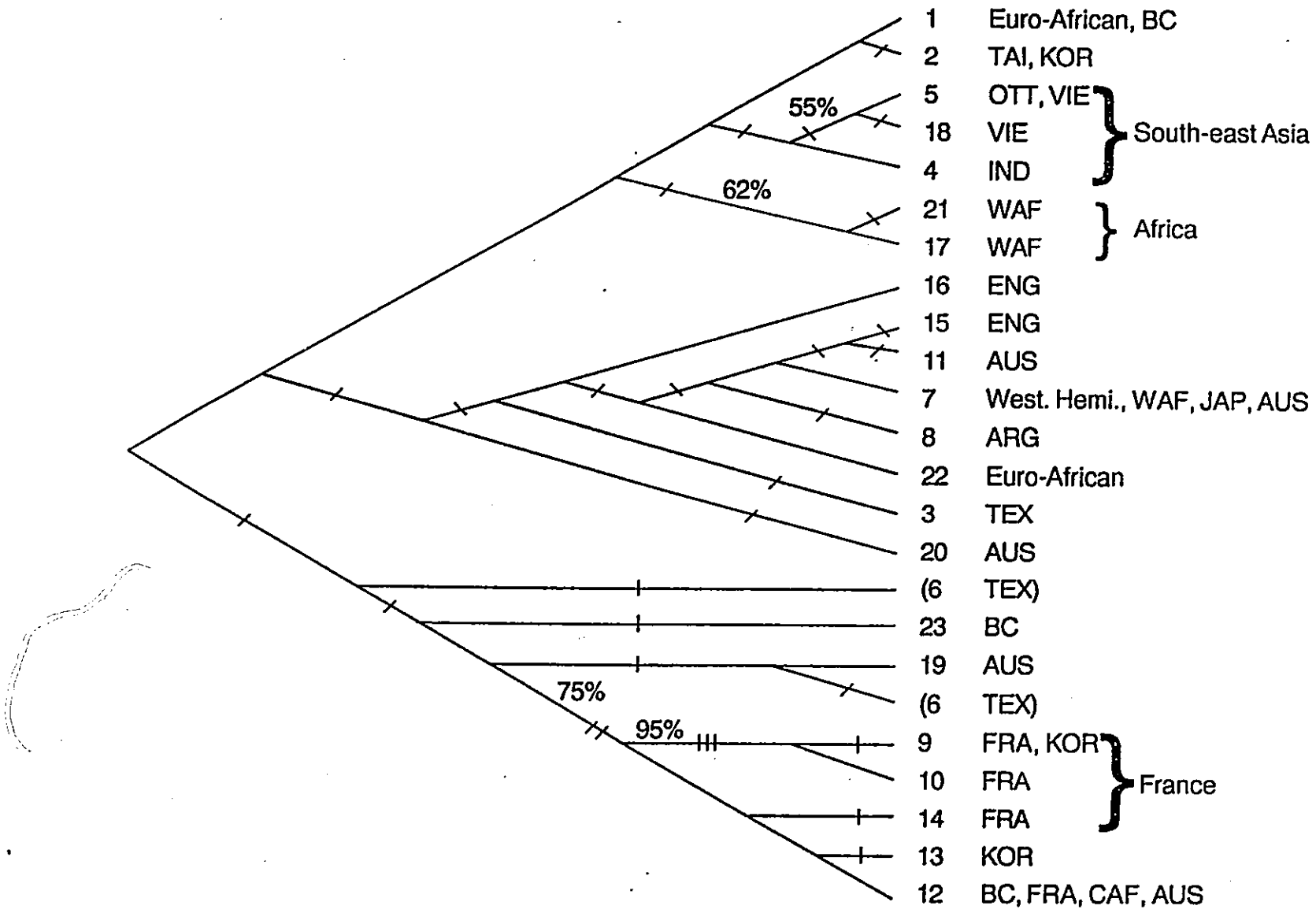

21
Ab
Hb
2
8
20
17
(6)
23
13
Aa
Fb
Fa
Da
Fd
Hc
Mb
11
Tb>
Ac>
7
Ha>
22
<Tb
16
Fa>
<Db
1
Ma>
<Ac
12
<Ad
10
<Da
9
Ma
<Ha
Fc>
Ha>Md>
15
3
Ta
Ta
Fa
Me
4
19
14
Mc
Ma
5
(6)
Db
18

Figure 3.13

Alternate depiction of Wagner parsimony networks giving detailed information as to the occurrence of each D. melanogaster mtDNA haplotype in geographic populations. Haplotype #6 is in parentheses to denote that there are two equally likely positions, both of which are shown. The population designations are as defined in the text. A single hatch along branches indicates one restriction site mutational event (see Fig. 3.11 for details on site changes); no hatch indicates a branch of zero length. Numbers along certain branches indicate the bootstrap confidence value for that monophyletic group; only values of > 50% are shown.


55%
62%
75%
95%
1 Euro-African, BC
2 TAI, KOR
5 OTT, VIE
18 VIE
4 IND
South-east Asia
21 WAF
17 WAF
Africa
16 ENG
15 ENG
11 AUS
7 West. Hemi., WAF, JAP, AUS
8 ARG
22 Euro-African
3 TEX
20 AUS
(6 TEX)
23 BC
19 AUS
(6 TEX)
9 FRA, KOR
10 FRA
14 FRA
France
13 KOR
12 BC, FRA, CAF, AUS

solely (for now, anyway) to illustrate these tests. The depiction in Fig. 3.12 gives detailed information about the particular restriction site change(s) between haplotypes. Figure 3.13 gives detailed information about the phylogeographic distribution of haplotypes. It must be emphasized that the topology is the same in each case.

The first test method places confidence limits on monophyletic branches of the tree (Felsenstein 1985). The BOOTM program uses a bootstrapping method to resample data and estimate the proportion of times a particular monophyletic grouping was found. Several runs of this program demonstrated that none of the branches which were variable among the eight trees were found at even 50% frequency (Fig. 3.13).

The second one is based on the finding by Templeton (1983) that parallel losses of a restriction site (i.e. when there are repeated changes of a site, both involve loss of the site from and ancestral 'present' state) are far more likely than parallel gains of a site. Therefore, the tree(s) that infer fewer parallel gains will be a 'better' tree. Unfortunately, knowing the evolutionary direction (gain/loss) of a site change requires that the tree be 'rooted', the determination of which can be quite arbitrary.

In the present case, ten of the mutational events will represent a repeated change at a particular sites, as the genealogies require 30 events among 20 variable sites. For the trees in Fig. 3.12, double events occur at six sites (Ava-c, Taq-b, Taq-b, Dde-a, Dde-b and, depending on which of the two

trees is selected, one of Mbo-a or Hinf-a) and triple events at two others (Hae-a, and the other of Mbo-a or Hinf-a). Haplotype #1 was chosen as the root because of its internal position in the genealogy, and for its high frequency in geographic locations that have been suggested as the place of origin of the species. With this root in place, seven of the parallel changes are losses (at sites Ava-c, Taq-a, Dde-a, Hae-a, and Hinf-a), and two are gains (at sites Dde-b and Mbo-a). One site (Taq-b) undergoes a gain and a subsequent loss, which is intermediate in likelihood (Templeton 1983). However, this count is exactly the same for all of the other six trees produced by MIX. Therefore, the trees cannot be statistically distinguished on the basis of fewer parallel gains.

The result is that no tree(s) among the eight can be adopted as a 'best' tree. The reason behind this is certainly the high degree of parallelism of site changes. Fortunately, this does not affect phylogeographic inferences to any significant degree. The one important common feature of all eight trees is that there is not a great deal of phylogeographic clustering. These clusterings which can be identified are associated with branches yielding a confidence estimate of greater than 50% by the bootstrap analysis. The major difference between the depicted trees and the other six concerns the position of the branch extending to the immediate left of haplotype #1 in Fig. 3.12; in the other six, the branch extends from the network near the trifurcation which leads to haplotypes #12, 19 & 23. The major effect of this

change is that haplotype #1 is no longer internal in the phylogeny. Having #1 in an internal position is intuitively attractive since, as indicated earlier, it can then be comfortably thought of as an 'ancestral' type. For this reason, and for the fact that these two trees are generated by the computer most often, they are adopted for working purposes in the remainder of the present research.

3.3.3 Distances among haplotypes

In addition to generating a phylogeny, the presence/absence matrix is used to generate measures of genetic distance between haplotypes (d_{ij}). These values are shown in Table 3.8 together with the minimum number of site differences between each pair (these are not corrected for parallel gain/loss of sites as indicated by the phylogeny). The genetic differences range from 0.108×10^{-2} for two haplotypes that differ by the presence/absence of a single r=4 site, to 1.23×10^{-2} for haplotypes at opposite ends of the phylogeny and which do not share restriction patterns for any of the enzymes. The average pairwise value is 0.053×10^{-2}, while the standard error terms for these estimates range from 0.112×10^{-2} to 0.391×10^{-2}. In the case of the lower values of d_{ij}, the relatively large standard error reflects the fact that only a single polymorphic site is determining the value of d_{ij}.

Table 3.8

Nucleotide distances between restriction haplotypes (d_{ij} x 10^{-2}) of *D. melanogaster*[a]

Haplo-type	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
1	-	.108	.327	.108	.216	.219	.436	.549	.658	.544	.440	.436	.544	.549	.657	.217	.109	.325	.327	.215	.327	.325	.327
2	1-0	-	.440	.219	.327	.330	.549	.663	.773	.658	.554	.549	.657	.663	.773	.327	.219	.436	.440	.325	.440	.436	.440
3	3-0	4-0	-	.219	.327	.330	.327	.440	.999	.882	.330	.773	.881	.889	.549	.109	.440	.217	.440	.325	.663	.217	.663
4	1-0	2-0	2-0	-	.109	.330	.549	.663	.773	.658	.554	.549	.657	.663	.773	.327	.219	.217	.440	.325	.440	.436	.440
5	2-0	3-0	3-0	1-0	-	.440	.657	.773	.882	.766	.663	.657	.766	.773	.881	.436	.327	.108	.549	.432	.549	.544	.549
6	2-0	3-0	3-0	3-0	4-0	-	.219	.330	.890	.773	.221	.440	.549	.554	.440	.219	.330	.549	.109	.436	.554	.327	.330
7	4-0	5-0	3-0	5-0	6-0	2-0	-	.109	1.11	.990	.219	.657	.766	.773	.217	.217	.549	.544	.327	.432	.773	.108	.549
8	4-1	5-1	3-1	5-1	6-1	2-1	0-1	-	1.22	1.10	.330	.773	.882	.890	.326	.327	.663	.657	.440	.544	.890	.217	.663
9	4-2	5-2	7-2	5-2	6-2	6-2	8-2	8-3	-	.108	.891	.437	.545	.550	.883	.882	.773	.990	.773	.874	1.00	.990	.549
10	3-2	4-2	6-2	4-2	5-2	5-2	7-2	7-3	1-0	-	.774	.325	.433	.427	.767	.766	.658	.874	.658	.759	.882	.874	.658
11	3-1	4-1	2-1	4-1	5-1	1-1	1-1	1-2	7-1	6-1	-	.440	.549	.554	.219	.219	.554	.549	.330	.436	.780	.327	.554
12	3-1	4-1	6-1	4-1	5-1	3-1	5-1	5-2	3-1	2-1	4-0	-	.108	.109	.436	.657	.549	.766	.327	.652	.773	.766	.327
13	4-1	5-1	7-1	5-1	6-1	4-1	6-1	6-2	4-1	3-1	5-0	1-0	-	.217	.545	.766	.657	.874	.436	.759	.882	.874	.436
14	4-1	5-1	7-1	5-1	6-1	4-1	6-1	6-2	4-1	3-1	5-0	1-0	2-0	-	.549	.773	.663	.881	.440	.766	.890	.881	.440
15	5-1	6-1	4-1	6-1	7-1	3-1	1-1	1-2	7-1	6-1	2-0	4-0	5-0	5-0	-	.436	.773	.766	.327	.652	.999	.325	.549
16	2-0	3-0	1-0	3-0	4-0	2-0	2-0	2-1	6-2	5-2	1-1	5-1	6-1	6-1	3-1	-	.327	.325	.327	.215	.549	.108	.549
17	1-0	2-0	4-0	2-0	3-0	3-0	5-0	5-1	5-2	4-2	4-1	4-1	5-1	5-1	6-1	3-0	-	.436	.440	.325	.219	.436	.440
18	3-0	4-0	2-0	2-0	1-0	5-0	5-0	5-1	7-2	6-2	4-1	6-1	7-1	7-1	6-1	3-0	4-0	-	.657	.322	.657	.432	.657
19	3-0	4-0	4-0	4-0	5-0	1-0	3-0	3-1	5-2	4-2	2-1	2-1	3-1	3-1	2-1	3-0	4-0	6-0	-	.544	.663	.436	.219
20	2-0	3-0	3-0	3-0	4-0	4-0	4-0	4-1	6-2	5-2	3-1	5-1	6-1	6-1	5-1	2-0	3-0	3-0	5-0	-	.544	.322	.544
21	2-1	3-1	5-1	3-1	4-1	4-1	6-1	6-2	6-3	5-3	5-2	5-2	6-2	6-2	7-2	4-1	1-1	5-1	5-1	4-1	-	.657	.663
22	3-0	4-0	2-0	4-0	5-0	3-0	1-0	1-1	7-2	6-2	2-1	6-1	7-1	7-1	2-1	1-0	4-0	4-0	4-0	3-0	5-1	-	.657
23	3-0	4-0	6-0	4-0	5-0	3-0	5-0	5-1	3-2	4-2	4-1	2-1	3-1	3-1	4-1	5-0	4-0	6-0	2-0	5-0	5-1	6-0	-

a) - Above diagonal: genetic distance (d_{ij} x 10^{-2});
Below diagonal: minimum number of site differences [(enzymes with r=4)-(AvaII)]

3.3.4 Distribution of mtDNA variants in D. melanogaster

A fundamental goal of this research was to assess historical relationships among D. melanogaster populations. This requires careful analysis of the distribution of variants among populations (Table 3.9), the genealogical relationships between haplotypes (Fig. 3.12), and of the genetic distance between populations (Table 3.10).

On the basis of intra- and inter-populational diversity of mtDNA observed in this study, D. melanogaster populations can be roughly divided into three geographical regions. One of these regions is the Western Hemisphere. The other two are both in the eastern hemisphere. One encompasses all populations in Europe and Africa, and is hereafter designated as Euro-African. The other encompasses the populations found in India and points east, and is hereafter designated as Far East. I will consider each region separately.

3.3.4.1 Euro-African populations

Euro-African populations of D. melanogaster are easily the most diverse. An average of 4.5 variant haplotypes were found in each sample, compared to 2.0 in the Far East and 1.8 in the Western Hemisphere. With the exception of Central Africa, these populations have at least two high frequency variants (i.e. observed in more than one line). This results in an average intra-populational diversity measure (d_s) for these populations of 0.106×10^{-2}; this is significantly larger

Table 3.9

Distribution of D. melanogaster restriction haplotypes by population.

Haplo-type	Western Hemisphere								Euro-African				Far East					
	OTT	HAM	MAS	FLO	BC	TEX	CAL	ARG	ENG	FRA	WAF	CAF	IND	KOR	JAP	TAI	VIE	AUS
1					1				3	3	5	1						
2														6		8		
3						1												
4													9					
5	1																5	
6						1												
7	11	5	6	7	3	9	6					1			10			2
8								5										
9										1				1				
10										3								
11																		1
12					1					4		1						2
13														4				
14										1								
15									2									
16									1									
17											3							
18																	1	
19																		1
20											1							
21											1							
22									2		1	2						
23					1													
n	12	5	6	7	6	11	6	5	8	12	11	5	9	11	10	8	6	6

Table 3.10

Average mtDNA nucleotide distances between populations (d_{xy} x 10^{-2}) of D.melanogaster.

	OTT	HAM	MAS	FLO	BC	TEX	CAL	ARG	ENG	FRA	WAF	CAF	IND	KOR	JAP	TAI	VIE	AUS
OTT	-	.055	.055	.055	.298	.096	.055	.164	.290	.772	.455	.284	.512	.667	.055	.530	.587	.338
HAM		-	0	0	.274	.050	0	.109	.272	.732	.467	.262	.549	.678	0	.549	.638	.310
MAS			-	0	.274	.050	0	.109	.272	.732	.467	.262	.549	.678	0	.549	.638	.310
FLO				-	.274	.050	0	.109	.272	.732	.467	.262	.549	.678	0	.549	.638	.310
BC					-	.290	.274	.385	.354	.542	.416	.353	.457	.530	.274	.457	.565	.353
TEX						-	.050	.159	.279	.711	.445	.276	.499	.656	.050	.529	.592	.324
CAL							-	.109	.272	.732	.467	.262	.549	.678	0	.549	.638	.310
ARG								-	.383	.849	.580	.373	.663	.794	.109	.663	.753	.422
ENG									-	.551	.335	.305	.384	.522	.272	.384	.487	.386
FRA										-	.492	.554	.494	.429	.732	.494	.621	.494
WAF											-	.367	.218	.427	.467	.218	.338	.489
CAF												-	.415	.533	.262	.415	.520	.382
IND													-	.429	.549	.219	.127	.531
KOR														-	.678	.309	.555	.527
JAP															-	.549	.638	.310
TAI																-	.345	.531
VIE																	-	.640
AUS																		-

than for either of the other two regions ($p < 0.05$ for both comparisons: Mann-Whitney U test). At least one of the high frequency variants in each population is unique to that population. This results in a total regional diversity measure (d_t) of 0.196×10^{-2}. Further, although one of the shared haplotypes, #1, is found at high frequency in 3 of the 4 populations, the populations are strongly differentiated, with G_{st} ($(d_t-d_s)/d_t$) of 0.459. The average inter-populational nucleotide diversity (d_{xy}) is 0.043×10^{-2}.

There are varying degrees of phylogenetic clustering of the haplotype complements in Euro-African populations (Fig. 3.13). The French haplotypes (# 9,10,12,14) form a (nearly) monophyletic group. The bootstrap confidence limit for this group is approximately 70%. The haplotypes from England (# 15,16) are not monophyletic, but all derive from just one of the five lineages branching from haplotype #1. Of the haplotypes from West Africa, two (#17 and #22) together comprise one of the lineages branching from #1. The bootstrap confidence limit on this branch is also greater than 50%. The other two West African haplotypes derive from a separate branch.

The greater diversity of Euro-African populations suggests a higher effective population size, either as a result of a longer history in this region, or as a result of greater stability in numbers compared to other regions. The relatively large number of unique haplotypes argues that this region has a longer history. It is noteworthy that Africa is

widely held to be where D. melanogaster originated, and the European colonization is said to have occurred in prehistoric times, 10-14 thousand years ago (Lachaise et al 1987). However, the mtDNA analysis does not distinguish between the European or African populations as being older than the other.

3.3.4.2 Far East populations

Far East populations are individually less diverse than Euro-African populations. Most populations have one dominant haplotype. This results in a d_s measure of 0.056×10^{-2}. However, the dominant haplotype differs among populations, so that over the region many different haplotypes reach high frequency. The d_t value for the region is 0.203×10^{-2}, which is close to that seen in the Euro-African region. As a result, the G_{st} value for this region is 0.724, which is noticeably higher than that calculated for the Euro-African region. On this basis, Far East populations would seem to be more differentiated, although individually less diverse. However, the average distance (d_{xy}) between Far East populations (0.046×10^{-2}) is not significantly greater than that seen in Euro-African populations ($p > 0.05$).

The large d_t in the face of a small d_s can be explained in two ways. First, the high degree of differentiation could be due to colonization of the region from a common founding stock, followed by independent stochastic effects in each population. The original colonization of the Far East region would have occurred

sufficiently far back in time to allow new variants to become established, such that the colonization would leave little or no residual effect on population structure. The low degree of intra-population diversity could be a result of smaller population sizes in this region that would facilitate drift, or to the past occurrence of a region-wide bottleneck that would allow otherwise polymorphic and relatively unstructured populations to fix different variants.

Alternatively, these populations could have been colonized by individual founding groups at different times and in a sequential manner. Under this hypothesis, routes of colonization can be inferred on the basis of mutational relationships between haplotypes. For instance, there is a clear relationship between the populations from India and Vietnam. The three haplotypes found here constitute a monophyletic branch off haplotype #1, although the bootstrap confidence limit placing #4 at that position is somewhat low (40%). It may be that one population was colonized from a distant source, differentiated to high frequency for haplotype(s) lacking the Taq-a site, and then acted as a source for the other population. As these haplotypes apparently derive from #1, a Euro-African population was likely the original source.

There is also a clear relationship between the populations from Korea and Taiwan. The two dominant haplotypes found in these populations are one mutational step removed from haplotypes found at high frequency in Euro-African

populations, particularly in France. A singleton line in Korea is of a haplotype (#9) found in France as well. These data seem to indicate that a founder group for these populations would have had to be polymorphic and of Euro-African origin, possibly Europe. The population from Taiwan is fixed for haplotype #2, suggesting that this population was colonized from Korea after the #2 haplotype was differentiated.

The Australian population does not have any dominant haplotypes, although this could be simply an effect of sample size. Its two highest frequency haplotypes (2 of 6 lines each) are not unique to either Australia or to the Far east region as well. The two unique haplotypes (# 11,19) derive from the other two, at distant points along the phylogeny. This suggests that the Australian population has also a Euro-African origin and possibly has received immigrants from the Western Hemisphere, but has not yet differentiated to the same degree as other Far East populations.

The Japanese population presents a special problem in that it contained only haplotype (#7), which is predominantly a 'Western Hemisphere' type. It is tempting to speculate that this population may have been colonized from North America. Nevertheless, a Euro-African origin in this case can not be ruled out.

3.3.4.3 Western Hemisphere populations

The Western Hemisphere populations are the least diverse of all the regions. All have either a single

haplotype, or a dominant one with 1-3 singletons. As a result, the regional d_s value ($0.033x10^{-2}$) is quite low in comparison to that from the Old World ($0.076x10^{-2}$). The dominant haplotype is the same in 7 of 8 populations (#7) and so the d_t value ($0.059x10^{-2}$) for the region is quite low as well. This also contrasts quite sharply to the situation for either Euro-African or Far East populations (ie. Old World), which is a result of many different haplotypes being found at high frequency.

The low degree of inter-populational differentiation of Western Hemisphere populations is best shown by the Nei genetic distance (d_{xy}) between populations (Table 3.10). The average distance between Western Hemisphere populations is $0.118x10^{-2}$, which is significantly different from corresponding values from the Euro-African and Far East regions ($0.434x10^{-2}$ and $0.462x10^{-2}$, respectively).

The comparative lack of diversity within Western Hemisphere populations suggests low effective population sizes. The lack of inter-populational diversity indicates that, unlike Far East populations, there has been no great degree of diversity in the region's recent history. One explanation that is consistent with these observations is that Western Hemisphere populations of D. melanogaster originated from a common stock within which there were founder effects that led to the predominance of haplotype #7. The recent colonization of America (in the middle of the nineteenth century; David and Tsacas 1981) would explain the general lack

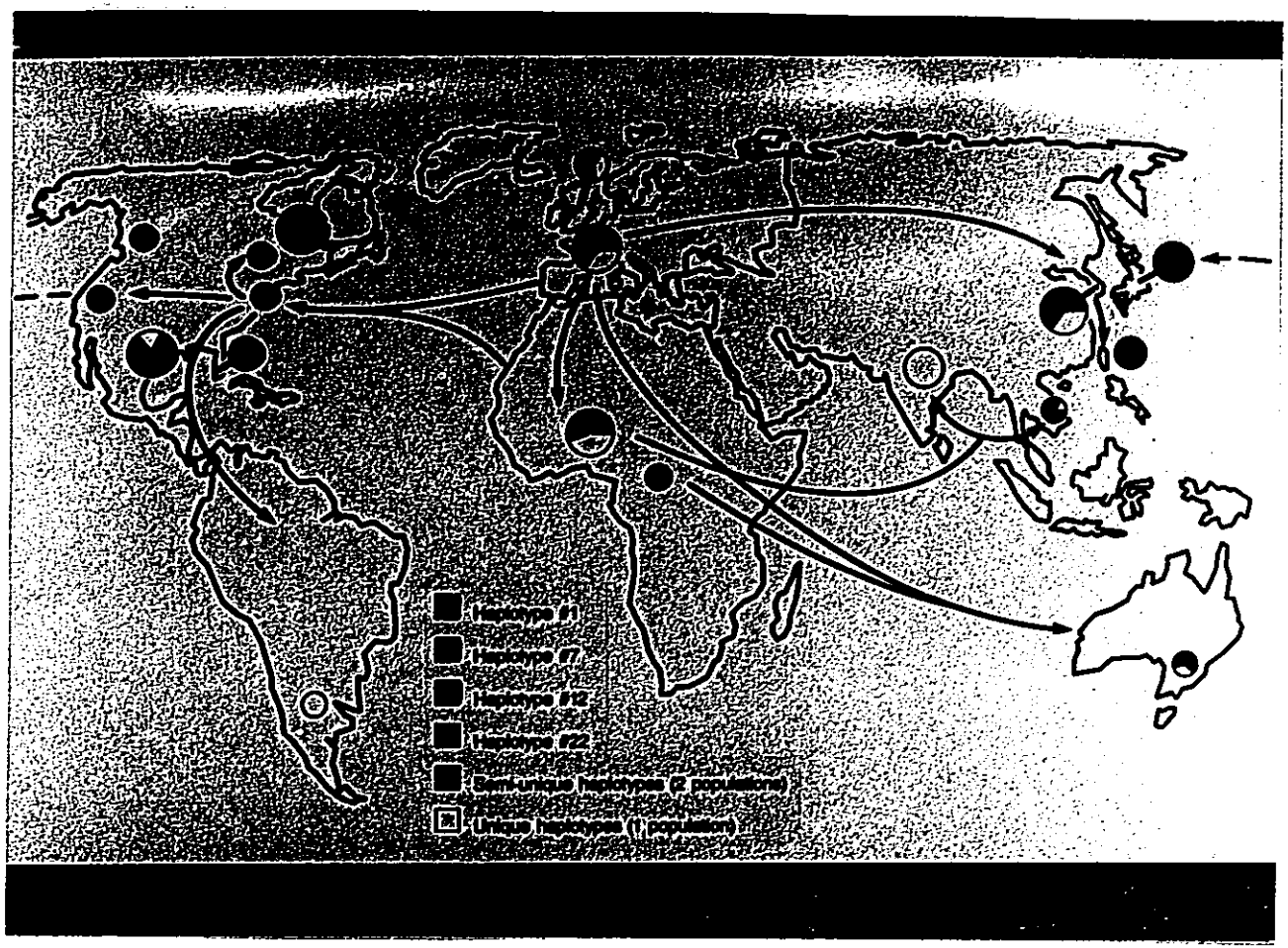
of new mtDNA variants in this region. It is also possible that the lack of variability in Western Hemisphere D. melanogaster is due to strong selection for haplotype #7 in North American populations. While selection of any sort is difficult to disprove (neutrality of mtDNA variants will be taken up further in the discussion), it is important to note that no other genetic element shows the effects of what would have to be intense purifying selection.

The lack of inter-populational diversity between Western Hemisphere populations would suggest that the expansion across the Western Hemisphere would have been reasonably recent. Yet the populations are beginning to differentiate. There may be significant differentiation between North and South America as Argentina already contains a fixed new variant. Also, all of the singleton variants in other Western Hemisphere populations are not derived from haplotype #7. This would indicate that despite its isolation recurrent introduction of flies into the Western Hemisphere is quite common.

The descriptions of D. melanogaster species history described in this section (3.3.4) are based solely on the mtDNA data. They are summarized in Fig. 3.14, which shows the inferred colonization history of D. melanogaster as predicted by mtDNA restriction site variation.

Figure 3.14

Possible colonization routes of D. melanogaster as inferred from mtDNA restriction haplotypes. Pie diagrams denote the haplotype composition of the population according to the insert. The pies correlate approximately with the size of the sample from each population.


Haplotype #1
Haplotype #7
Haplotype #12
Haplotype #22
Semi-unique haplotypes (2 populations)
Unique haplotypes (1 population)

3.4 Sequence divergence and evolution of mtDNA in *D. simulans*

3.4.1 Distribution of restriction variants in *D. simulans* populations

Prior work on restriction variation in mtDNA from natural populations of *D. simulans* have revealed a very discontinuous distribution of variants. Baba-Aissa et al (1984, 1988) and Solignac et al (1986a) found three distinct mtDNA types. One type, designated sI, was restricted to the Seychelles Islands, which is an archipelago in the Indian Ocean, and a few south Pacific islands. Another, designated sIII, was found to be restricted to the island of Madagascar, also in the Indian Ocean. The third type, sII, was found in all continental populations. The three types are strongly diverged from one another, with a range of pairwise nucleotide distance measures of 1.49×10^{-2} - 3.29×10^{-2} (Solignac et al 1986b). By contrast, the maximum distance measures for *D. melanogaster* haplotypes is 1.22×10^{-2}, with the vast majority under 1% (Table 3.8). In this study, all of the *D. simulans* sampled from continental localities were of type sII, and all of the seven lines from the Seychelles Islands were of type sI.

The distinctive aspect of the distribution of *D. simulans* mtDNA polymorphism is that, despite the comparatively large mutational distance between major types (ie. sI, sII, and sIII), there is virtually no polymorphism within each type. All of the variation is inter-type, while virtually none

is intra-type (Baba-Aissa et al 1988). The same pattern was also observed in the present research. Of the seven sI mtDNAs, all were identical over the nine enzymes, except for one line where a single DdeI site had been lost. The total number of restriction sites examined in sI mtDNA was 94.

In the 68 sII mtDNAs surveyed, only one enzyme, HinfI revealed any polymorphism. There were four variant HinfI patterns (Fig. 3.15) which are easily linked by gain/loss of three restriction sites. These three HinfI sites are the only polymorphisms in the sII lines surveyed. The total number of restriction sites in sII mtDNAs was 104.

There are, therefore, four sII haplotypes based on the four HinfI variant patterns. The evolutionary relationship between them is simple (Fig. 3.16) with each separated from adjacent haplotypes by a single restriction site gain/loss. The average distance measure between haplotypes is exceedingly small (0.0019). By comparison, the unweighted average distance between the 23 D. melanogaster haplotypes is 0.0053.

The distribution of the four HinfI defined haplotypes into populations is shown in Table 3.11. As can be seen, two types, 'A' and 'B', are much more frequent than the other two, 'C' and 'D'. Along the Euro-African transect, the 'A' variant predominates in the more northern two populations (France and Tunisia), while the 'B' variant is fixed in the more southern two populations (Congo and South Africa). The two are about equally frequent in the United States populations.

The two other variants are comparatively rare and are

Figure 3.15

HinfI digests of D. simulans type sII mtDNA. The full restriction pattern of the four variants is shown by the appropriate lanes on both gels together. (A) Agarose gel (1%) showing the larger fragments. HindIII-cut lambda-DNA is used as size standard; fragment sizes are denoted in the margin. Lane 1, variants A and D. Lane 2, variant C. Lane 3, variant B. Lane 4 shows the single line with a size polymorphism in the upper fragment, presumably due to an insertion in the A+T-rich region. (B) 8.5% acrylamide gel showing the smaller fragments. HpaII-cut pBR322-DNA is used as size standard; the sizes of some of these fragments are denoted in the margin. Lane 1, variant A. Lane 2, variants C and D. Lane 3, variant B.

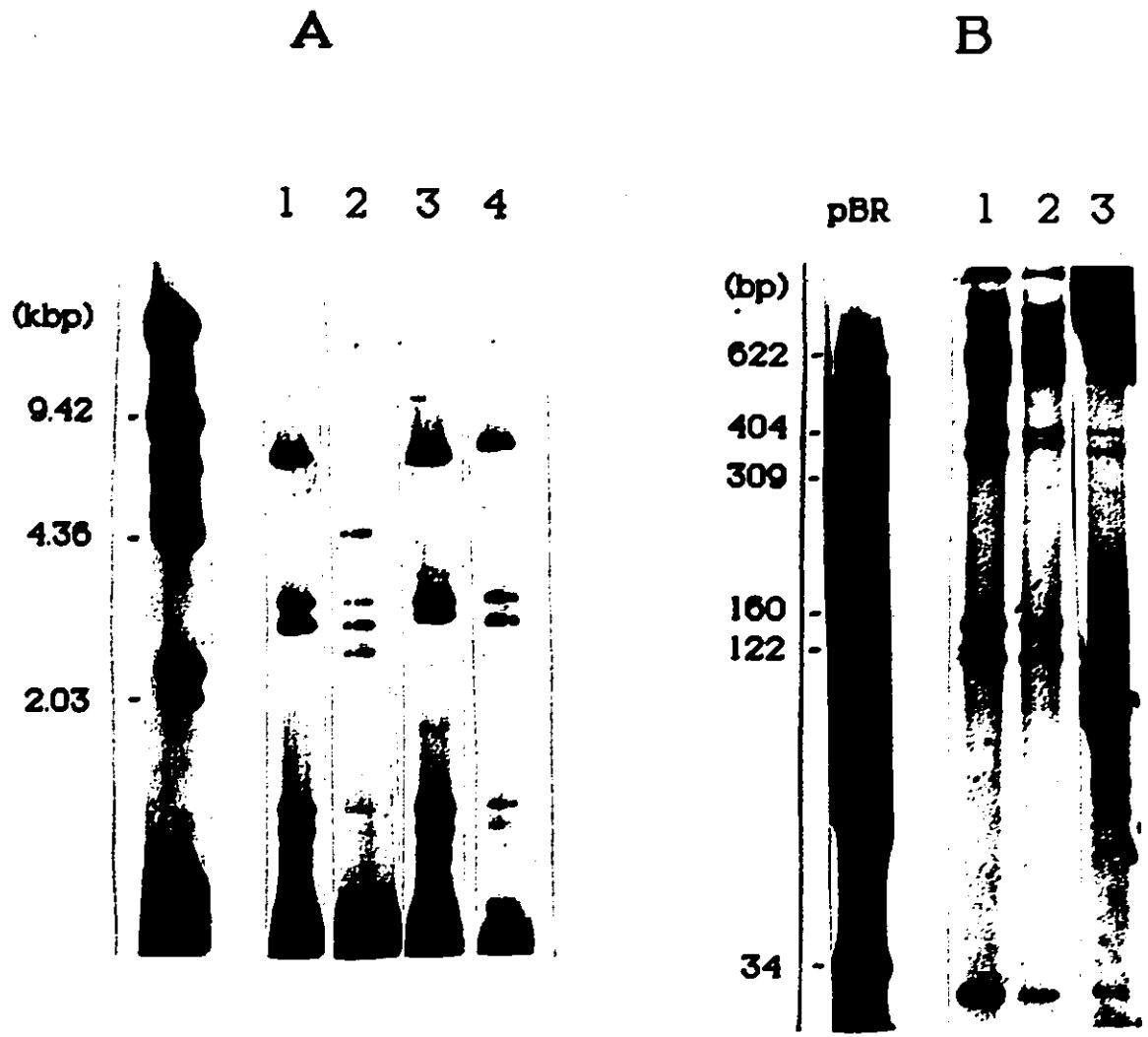

A
1 2 3 4
(kbp)
9.42
4.36
2.03
B
pBR
1 2 3
(bp)
622
404
309
160
122
34

Figure 3.16

Evolutionary relationships between the three HinfI variants observed in D. simulans type sII mtDNA. The presence (1) or absence (0) of the three variable sites is shown below each.

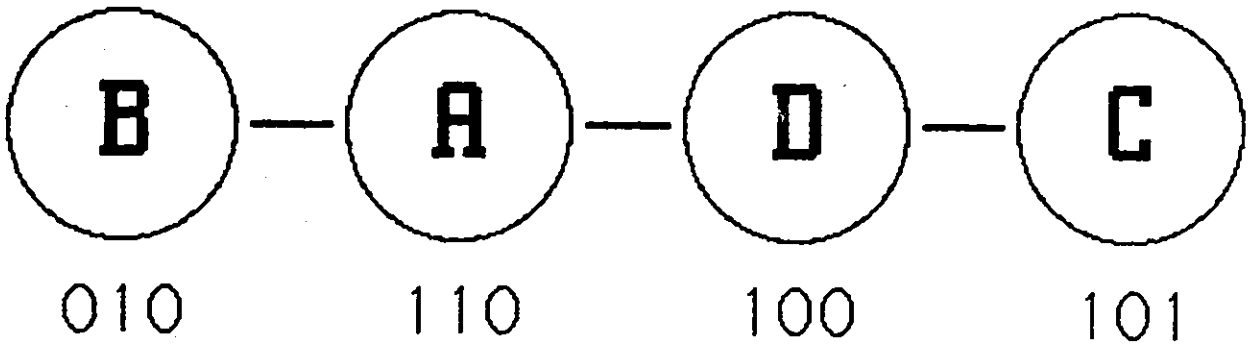

B
A
D
C
010
110
100
101

Table 3.11

Distribution of *D. simulans* restriction haplotypes (based on HinfI variant) by population[a]

	Haplotype			
Population	A	B	C	D
France	8	2	2	-
Tunisia	8	-	-	2
Congo	-	14	-	-
Cape Town	-	12	-	-
Australia				
Melbourne	-	-	4	-
Sydney	2	-	2	-
United States				
Florida	1	-	-	-
North Carolina	1	1	-	2
Maryland	-	1	-	-
Rhode Island	1	-	-	-
Massachusetts	1	-	-	-
California	2	3	-	-

a - Letter designations of HinfI variants are as in Fig. 1. Locations of populations are as described in text.

each found in only two populations. In each case, however, the two populations are geographically quite distant. The 'C' haplotype is found in Australia and France, and the 'D' haplotype is found in Tunisia and North Carolina. In addition, they were observed in more than one line, suggesting that they were not spurious migrants and that their presence was probably commensurate with the expansion of the species. By contrast, 14 of 23 D. melanogaster haplotypes were restricted (ie. unique) to a single locality.

Despite the lack of 'unique' haplotypes, the differences in haplotype composition between populations of D. simulans are distinct. The difference between the northern and southern parts of the Euro-African transect is particularly sharp. Also, the major haplotype in Australia is one of the two rarer types ('C'), which is otherwise found only at low frequency in France.

3.4.1.1 mtDNA polymorphism and mating incompatibility

Hoffman et al (1986) and Hoffman and Turelli (1988) have described a system of unidirectional incompatibility of populations of D. simulans. The two mating types, referred to as 'W' and 'R', are such that a 'W' female crossed to an 'R' male will produce very few progeny. The reciprocal cross has normal numbers of progeny. This is a maternal cytoplasm effect involving a rickettsia-like microorganism of the genus Wolbachia. All progeny from the 'R' female to 'W' male cross are 'R'. Population studies of this system have been

restricted to D. simulans populations in California. North and central California populations are all 'W' type, while those of southern California are 'R' type. Single 'R' lines have also been found in Ottawa and Italy.

Populations of D. simulans from Central to Southern California were examined in an experiment to analyze the relationship between mtDNA population structure and the distribution of the R/W mating incompatibility types. Incompatibility types for each isofemale line were determined by Dr. A. Hoffman (La Trobe University, Australia). Lines were checked for the HinfI type and for other enzymes in hopes of finding diagnostic polymorphism, particularly between dispersed 'R' lines. The other enzymes were EcoRI, HpaII and ClaI (as rare polymorphism has been observed from sII D. simulans by other researchers), and MboI and TaqI (because of the high number of sites cleaved).

The results are shown in Table 3.12. The central lines (Davis) were all type 'W' and were monomorphic for HinfI-A, while the southern lines (Highgrove) were all type 'R' and were monomorphic for HinfI-B. Populations in the dividing range harbored both mating types; mtDNA haplotype depended strictly on mating type. The 'R' lines were monomorphic for HinfI-B, while the 'W' lines were polymorphic for HinfI-A and HinfI-B. There was no other polymorphism for any other enzyme, including no detectable A+T-rich region size variation. The association of the 'R' strain with HinfI-B is highly significant ($p < 0.01$; G-test).

Table 3.12

Distribution of HinfI restriction haplotypes in *D. simulans* populations surveyed for mating incompatability type.

Region	Population	Mating type	n	HinfI B	HinfI A	HinfI D
Northern California	Davis	W	6	0	5	1
Dividing range	Santa Ynez	W	12	6	6	-
		R	9	9	0	-
"	Filmore	W	4	2	2	-
		R	4	4	0	-
Southern California	Highgrove	R	6	6	0	-
Ottawa		R	1	1	0	-
Italy		R	1	1	0	-

3.4.2 Inter-type and inter-species divergence

The two *D. simulans* types studied here (sI and sII) are sufficiently divergent to preclude inference of the precise number of mutational events between them, particularly with regard to r = 4 enzymes. Therefore, distances between these two types, and between each and *D. melanogaster*, were calculated by the apparent number of shared restriction fragments.

The calculated distances are shown in the upper diagonal of Table 3.13. The lower distance between *D. simulans* type sII and *D. melanogaster* than between the two *D. simulans* types suggests some sort of substitution rate heterogeneity among the conspecific types. The higher distance between *D. simulans* type sI and *D. melanogaster* further suggests that this heterogeneity would be in the form of an increased rate in the sI lineage. Such an increase rate could be indicative of more pronounced random genetic drift in the sI lineage due to a smaller effective population size (N_e). However, it must be emphasized that calculations of genetic distance based on shared fragments are much more prone to error than calculations based on shared restriction sites, particularly when patterns are strongly diverged, since similarly sized fragments may have arisen in a convergent fashion. The lower diagonal of Table 3.13 shows the distances calculated by Solignac et al (1986b) using shared restriction sites. These results give no indication of rate heterogeneity in *D. simulans* mtDNA.

Table 3.13

Nucleotide distances between *D. simulans* mtDNA types sII and sI, and *D. melanogaster* as calculated by two different studies[a]

	D. melanogaster	*D. simulans* sI	*D. simulans* sII
D. melanogaster		0.0931	0.0576
D. simulans sI	0.0491		0.0621
D. simulans sII	0.0504	0.0329	

a) - above the diagonal: distances calculated by shared fragment data (this study); below the diagonal: distances calculated by shared restriction site data (Solignac et al 1986).

4

DISCUSSION

4.1 Evolution of Drosophila mtDNA

4.1.1 Nucleotide substitution

The dynamics of nucleotide substitution in mtDNA are complex. The rate of base substitution in mtDNA has been the subject of much controversy. Originally, it was believed that the rate of evolution of mtDNA was universally higher than that observed in single copy nuclear DNA (scnDNA), as the proportion of mtDNA sequence polymorphism among primate species was found to be about 10 times that observed for scnDNA (Brown et al 1979). Subsequent analyses, however, have indicated that the rate of mtDNA substitution relative to scnDNA is quite variable across taxonomic groups (Vawter and Brown 1986). For instance, the scnDNA and mtDNA rates in Drosophila are apparently quite similar (Caccone et al 1988); an apparent exception are the Hawaiian Drosophila species where the rate of mtDNA substitution is significantly higher (DeSalle et al 1986a). Vawter and Brown also suggest that rate variability occurs primarily in scnDNA, and that the absolute rates of substitution in mtDNA are quite similar across taxonomic groups. Therefore, the comparatively low degree of mtDNA divergence in Drosophila may not be a result of a lower rate of mtDNA substitution.

One of the assumptions of many models of DNA sequence evolution, including mtDNA, is that the rate of evolution is the same at all sites and that the probability of substitution to each alternative nucleotide is the same (Kimura 1983). However, it has become increasingly obvious that this does not hold in all situations. First, the rate of substitution can differ due to different constraints placed on particular regions. Different genes will have higher rates by virtue of functional constraints, introns will have higher rates than exons, and third codon positions will have higher rates than first or second positions. In mtDNA such differential constraints are seen between the large non-coding region which has a higher rate, the rRNA genes which have much lower rate, and the protein and tRNA genes which are intermediate (Moritz et al 1987). The paucity of variable restriction sites in this study makes assessments of how these constraints are operating within D. melanogaster difficult. The only obvious heterogeneity in the distribution of variable sites in D. melanogaster is the occurrence of 5 of the 20 sites in the ATPase6 gene (unfortunately, the ambiguity in the position of some variable sites does not allow the evaluation of statistical significance of this observation). Yet on the basis of inter-specific comparisons of nucleotide sequence data, the ATPase6 gene is not more diverged than other mitochondrial genes.

At another level, there is evidence that the rate of substitution may be elevated at particular nucleotide sites.

The genealogical analysis revealed that while most restriction sites have remained monomorphic, some have been substituted twice and even three times, in numbers greater than expected under the Poisson distribution ($p < 0.01$: X^2-test, 2 d.f.). The extent of such 'parallel evolution' of sites is approximately 33% (30 steps in the phylogeny - 20 variable sites; 10 / 30 = 33%). Such parallel evolution has been observed in other taxa at levels of 44% in Equus species (George and Ryder 1986), and 18% in Mus (Ferris et al 1983). There have also been other reports of 'hypervariable' sites (ones that change at least three times within a species) in Peromyscus (Lansman et al 1983a), and Drosophila pseudoobscura (Hale and Beckenbach, 1985).

The reasons underlying such cases of hyper-variability of sites is very unclear, but likely involves recurrent substitutions at single bases (Aquadro and Greenberg 1983). The practical effect for evolutionary studies are that they complicate phylogenetic analysis. Templeton (1983) has pointed out that phylogenies should be selected that minimize parallel gains of sites. Parallel gain is the least likely type of parallel change as it requires the same substitution at the same base, while loss of a site occurs with any substitution to any of the bases within the recognition sequence. The phylogeny in Fig. 3.12 invokes only 2 parallel gains among the 10 parallel substitutions.

A second way in which the idealized model of nucleotide substitution does not hold is that some types of

substitutions often do not occur with equal probabilities. Strong transition/transversion biases have been observed in many molecules. In vertebrate mtDNA, transitions occur far more often than transversions (Aquadro and Greenberg 1983, Brown et al 1982). The situation differs somewhat in Drosophila in comparisons between species subgroups, where transversions greatly outnumber transitions (Wolstenholme et al 1985). However, this is not seen over more recently diverged species. Of the four variable sites in D. melanogaster mtDNA for which a nucleotide changes can be inferred, two are transitions and two are transversions. In addition, there is a clear bias toward codons ending in A or T in Drosophila mtDNA (Wolstenholme and Clary 1985) indicating that substitutions to A or T are somehow favoured mechanistically over substitutions to G or C.

4.1.2 Size variation

The other manner in which restriction patterns of mtDNA can vary in intra-specific comparisons is due to variation in the total size of the mtDNA molecule. Early studies did not focus clearly on this (Avise et al 1979, Lansman et al 1983b) since homeothermic animals, which were the subject of most of the early studies, tend to have the most stable mtDNA molecules in terms of size (Sederoff 1984). However, as the range of animals that have been studied for intra-specific variation of mtDNA has increased to include a wide array of homeotherms and poikilotherms (Densmore et al

1985, Powers et al 1986, Wallis 1987), it has become increasingly obvious that there is considerable size variation of mtDNA within species. A+T-rich region length polymorphism in Drosophila was first detected in D. melanogaster by Reilley and Thomas (1980) and in D. simulans and D. mauritiana by Fauron and Wolstenholme (1980 a,b). Heteroduplex reannealing analysis failed to suggest a mechanism for the generation of size variation, but it was observed that the A+T-rich region does become more different in sequence between species more rapidly than other parts of the mtDNA molecule. As indicated, D. melanogaster mtDNA varies in size due to addition/deletions to the non-coding (A+T-rich) region. As well, apparent size alterations occur at least three separate points within the coding region.

4.1.2.1 Non-coding region size variation

Size variation in the A+T-rich region of D. melanogaster is extensive. In addition, Baba-Aissa et al (1988) have found some size variants within sII D. simulans. Of 93 lines studied, 75 were of a single class. Among the remaining 18 lines there were two smaller size classes, the smallest one being 0.23 kbp smaller than the dominant class, and five larger size classes, the largest one being 0.55 kbp larger than the dominant class. Therefore, the distribution was skewed toward the smaller type, but not as sharply as in D. melanogaster. Baba-Aissa et al (1988) also observed proportionately less size variation in the sI type mtDNA. 45

of 61 lines harboured a dominant size class that is about 0.2 kbp larger than the dominant one in sII type. Most of the other sI size classes are smaller.

The most frequent sort of size variation yet recognized in animal mtDNA is due to differences in copy number of a tandemly repeating unit in the non-coding region. In Gryllus crickets, a 220 bp repeat within the non-coding region (D-loop region) is found in several copies (Rand and Harrison 1989). The repeat unit is bounded at either end by a G+C-rich 14 bp sequence with dyad symmetry. The secondary structure of this short sequence could present a signal for the slippage of replication, leading to the generation of an extra copy of the unit, or the deletion of one. Rand and Harrison (1989) also proposed that the presence of such repeats could present the opportunity for intra-molecular recombination, using the sequence homology of the repeat units.

The size of the repeat is clearly not critical. Snyder et al (1987) identified in the mtDNA of Placopecten magellanicus a non-coding region tandem repeat, present in two to eight copies. At 1.2 kbp in length, it is much longer than that reported for Gryllus. At the other end of the scale, Densmore et al (1985) identified an 86 bp tandem repeat in Cnemidophorous lizards.

Solignac et al (1986a) have shown the presence of repeating units within the A+T-rich region of Drosophila. A 470 bp repeat is found in a copy number of one to six copies.

In the mtDNA of melanogaster species complex species (D. melanogaster, D. simulans, D. sechellia, D. mauritiana) there are at least three copies of this repeat unit. These repeat units occur within the A+T-rich region, somewhat closer to the URF-2 side. To the URF-2 side of the repeat cluster are 750-1000 bp of DNA, while to the rRNA side of the cluster are 2150-3000 bp of DNA.

The distribution of size classes observed in this research with D. melanogaster is multimodal between 18.6 kbp and 19.9 kbp. The highest mode is at 18.6 kbp, with smaller ones at 19.1, 19.5 and 19.9 kbp. These data can easily be interpreted in the context of the observations and model of Solignac et al (1986a) suggested for other members of the melanogaster species complex. The sizes corresponding to the smaller frequency modes are ones that would be observed if there were three successive additions of the 470 bp repeat unit to the dominant 18.6 kbp class. Mechanisms acting to create size variation to either side of the repeat cluster could then produce the continuity of variation seen around each modal size class.

The processes that might be acting to cause size variation outside the repeat cluster is not known. The other mechanism by which size variation has been generated is the occurrence of long runs of the same nucleotide (homopolymer runs). Hauswirth et al (1984) reported that in the non-coding region D-loop region of bovine mtDNA, tracts of up to 19 C-residues are polymorphic in length. The extremely high A+T

content of the A+T-rich region in Drosophila suggests that long runs of A's and T's may be commonplace. The extra size classes may therefore be a result of just such polymorphic homopolymer runs. Alternatively, there may be a second smaller tandemly repeating unit to the rRNA side of the 470 bp repeat cluster that may be causing the extra size variation such as that observed by Densmore et al (1985) in Cnemidophorous. Varying copy number of such a smaller repeat unit like this could result in the range of sizes seen in D. melanogaster.

4.1.2.2 Small-scale size (mobility) polymorphism

Of the (at least) three cases of small-scale size polymorphism in D. melanogaster mtDNA, the most variable is within the 0.26 kbp MboI fragment. The frequency at which changes are made to the mobility of this fragment may be much higher than the A+T-rich region size variants. For instance, the isofemale lines from Japan show variation in the 0.26 kbp MboI fragment, even though they all were of the same restriction site haplotype and no A+T-rich region size variation was observed. The size class distribution of 0.26 kbp MboI fragment variants is not skewed as it is for A+T-rich region variants. Rather, the distribution more closely resembles a normal distribution, which is what would be expected under random small-scale addition/deletion with no transmission selection. Polymorphisms in the other mobility variant fragments occur far less frequently, and are restricted to a few haplotypes that are found largely in

France.

As indicated in the results section, however, it is difficult to ascribe the small-scale mobility variation of the 0.26 kbp MboI fragment to actual size variation of the fragments. This is due to the complete lack of concordance between variation at the 0.26 kbp MboI fragment and any of the TaqI fragments. As actual size variation would virtually demand the presence of inter-genic DNA, the only sequence that could correspond to the 0.26 kbp MboI fragment would be the 243 bp MboI sequence. Since this sequence overlaps extensively with the 330 bp TaqI sequence (including both segments of inter-genic DNA), variation due to insertion/deletion should have resulted in concordance between fragments on the gel.

An alternate explanation for the small-scale mobility polymorphisms may be that they are conformational mutants of the fragments introduced by base substitutions. Singh et al (1987a) reported that substitutions at three different nucleotide sites could cause mobility shifts in human mtDNA fragments that were 255 bp and 297 bp in length. They only occur in medium-to-small sized fragments, do not occur in all such fragments, and are seen only on polyacrylamide gels, not agarose gels. Similar findings have been reported for nuclear DNA of other systems (Bossi and Smith 1984, Zahn and Blattner 1985, Snyder et al 1986)). Singh et al (1987a) suggest that many of the size polymorphisms reported by Cann et al (1983) may actually have been such conformational mutations.

There is indirect evidence in this study that fragment

conformation plays a role in the mobility of Drosophila mtDNA fragments on acrylamide gels, in that mobilities observed in non-denaturing gels departed somewhat from the predictions of the sequence analysis. Denaturing gels confirmed the correctness of the pattern predicted by the published sequence. Unfortunately, the gels intended to check for conformational affects in the 0.26 kbp MboI fragment were not successful as the bands were much less sharp than those seen on non-denaturing gels. Attempts to clone and sequence some of the fragments showing small-scale mobility polymorphism were also unsuccessful. Therefore, there is presently no data to determine whether the conformational mutation hypothesis applies in the present study. In the absence of such data, the conformational mutation hypothesis is the most attractive explanation. One would hypothesize in the present case that the 0.26 kbp fragment is particularly sensitive to conformation alteration. Other fragments would be comparatively insensitive to such changes, with the clear majority of small-to-medium sized fragments showing no such alteration.

4.1.3 Are observed mtDNA variants selectively neutral?

As indicated at the outset, neutrality of intra-specific variants of mtDNA is a key premise behind this work. Yet the shape of the frequency distribution of A+T-rich region size variants implies some sort of advantage to a smaller size, raising questions about the strict neutrality of extant

mtDNA variants. The hypothesis of selective neutrality of mtDNA variants has been largely conjectural. The gene products encoded by mtDNA (ie. involved in ATP production and the electron transport chain) are of critical importance to the organism, seemingly leaving little room for variation in enzyme function. That is, it is assumed that any substitution in the gene product that affects metabolic function will have sufficiently deleterious effects on the organism that they will be rapidly eliminated.

Examples of the potential effect of mitochondrial genome change on the organism have been presented in the medical literature (reviewed in Merril and Harrington 1985, and Poulton 1988). Many mitochondrially-based disorders come under the classification of 'mitochondrial myopathy', as they are associated with structurally abnormal mitochondria and defects in the electron transport chain. The effects on the individual are variable, but most often rather severe. This suggests that should similar defects occur in natural populations of other species, the affected individual would likely not survive, and thus the mtDNA variants associated with them would be only transient. As yet, however, there has been no link between such abnormalities and a mtDNA sequence (base substitution) variant.

Support for the view that mtDNA sequence variants in natural populations are selectively neutral originally came from the observation of a great preponderance of synonymous over non-synonymous nucleotide substitutions in closely

related species (Brown and Simpson 1982). However, comparisons of more distantly related species indicate 1) that amino acid substitution in mitochondrially encoded polypeptides is not rare, and 2) that the rate of nucleotide and amino acid substitution is variable among mitochondrial genes (Brown 1983, Clary and Wolstenholme 1985). This indicates that some gene products are not as tightly conserved as others.

In addition, it is becoming increasingly apparent that 'synonymous' does not necessarily mean 'neutral'. Codon usage biases can lead to certain substitutions being preferred over others. In Drosophila this has led to a preponderance of codons ending in A's and T's (Clary and Wolstenholme 1985). Also, the observation in this study that two restriction enzymes (HinfI and DdeI) have significantly fewer sites than that predicted from the nucleotide composition of the mtDNA, suggests that something may be operating to remove some of these sites when they arise in the mtDNA molecule. A more obvious possibility is the local conformation introduced by the interrupted palindrome of HinfI and DdeI recognition sequences.

Another key aspect of the nature of mitochondrial gene products are that most of them are subunits of enzymes for which the other subunits are encoded in the nucleus. This leaves the obvious potential for strong cyto-nuclear interactions based on the association between subunits. Much effort has gone into describing the theoretical effects of such interactions (Clark 1984, Gregorius and Ross 1984,

Asmussen et al 1987, Arnold et al 1988), but empirical studies are still few in number. Clark (1985) and Clark and Lyckegaard (1988) have studied cytoplasmic effects on second chromosomal segregation in strains of D. melanogaster from geographically disperse source populations. They observed a significant cytoplasmic effect on chromosomal transmission when disperse lines are crossed, but not when lines from the same population are crossed. Even when there is an effect it is not consistent as the maternal cytoplasm can favour the chromosome from its own population or from the other. Clark and Lyckegaard (1988) point out that 'cytoplasmic effects' are not necessarily a product of organelles or their DNA. The 'R' and 'W' cytoplasm of D. simulans which is defined by the presence/absence of a cytoplasmically borne microorganism (Hoffman and Turelli 1988) stands as an example.

Finally, MacRae and Anderson (1988) have presented results of cage experiments in which two distinct mtDNA variants of Drosophila pseudoobscura were placed in competition. In one cage, one type of mtDNA increased steadily in frequency to an apparently stable level. However, this result was not reproduced in any other cage. Further, one of the fly strains comes from a population (Bogota) that is geographically isolated from the rest of the species. Flies from Bogota are partially reproductively isolated from those from mainland populations (Prakash 1972) and show non-random mating preference (Singh 1983). Singh and Hale (1989) have argued that this could have affected the transmission of mtDNA

in MacRae and Anderson's (1988) experiment.

In summary, the question of strict neutrality of extant intra-specific sequence variants remains open. While there is little hard evidence to support it, there is perhaps less to dispute it. For instance, the finding of variable rates of evolution among mitochondrial genes was based on observations between diverged species, and may not reflect the situation at the intra-specific level. A final answer will probably require some analysis at the DNA sequence level of the comparative levels of intra-specific polymorphism vs. inter-specific divergence (Hudson et al 1987, Kaplan et al 1988).

4.1.3.1 Heteroplasmy and the non-neutrality of size variants

Heteroplasmy for size variation in D. melanogaster mtDNA was found at a level commensurate with that found in other species. Although no heteroplasmic lines for D. simulans were observed in this study, Baba-Aissa et al (1988) have observed some in both sI and sII type mtDNA. Four heteroplasmic sII lines carried the dominant type and a somewhat larger one. The three heteroplasmic sI lines all contained the dominant type, with one also harbouring a smaller type, and two also harbouring a larger type.

Clark (1988) has commented on the observation of the skewed distribution of A+T-rich region size variation in D. melanogaster mtDNA (Hale and Singh 1986, and Fig. 3.16). He

argues that the distribution could be a result of the mechanisms underlying the generation of size variation, rather than selection of the smaller size variant. Specifically, if the mode of generation of size polymorphism does not allow deletions, then the distribution of variants generated from an initially size monomorphic population would approach that seen in Fig. 3.16, as a transitory stage toward a predicted normal distribution. He has presented a simple mathematical model to show how this would proceed.

However, the alternative explanation of Clark (1988) misses two very important features of the distribution of size classes into restriction site haplotypes. First, all haplotypes which are found in more than a single line harbour size variation. This indicates the high rate of mutation for new size variants. Second, all of these restriction site haplotypes include the 18.6 kbp size class among the sizes they harbour. Taken together, these points require that in order for phylogenetically distant restriction variants to include the 18.6 kbp size variant in the absence of any mechanism to generate deletions, or constraints on size, each new restriction variant would have to arise from a 18.6 kbp sized molecule. This is unlikely. Any originally size monomorphic population would become polymorphic for new, larger size variants relatively quickly. Since the ancestral state could not be re-attained, the frequency of the original smaller class will become less and less frequent. Eventually, it becomes extremely unlikely that phylogenetically distant

haplotypes would arise in the original smaller size class.

Assuming that the distribution and frequencies of size variants is not due to the dynamics of mutation alone, what sort of selection could produce the skewed distribution seen? Selection at the level of the organism seems unlikely. As the polymorphism is in the non-coding region, none of the gene products could be structurally altered. Alternatively, it is conceivable that changes to the super-helical conformation of the mtDNA introduced by a change in size could alter the interaction of the transcriptional machinery to the site of transcriptional initiation, thereby altering the efficiency of transcription. If the association were somehow hampered by the conformation found in larger sizes, then the production of transcript, and possibly the production of gene products would be affected. However, the enzymes involved in the electron transport system are not abundant, and there are several copies of the gene for each enzyme (one on each mtDNA molecule) within each mitochondrion. It is unlikely, therefore, that small alterations in transcriptional efficiency would alter the quantity of the enzymes to the point that it would effect the efficiency of the respiratory process, and thereby affect organismal fitness.

The other possibility is that selection is at the intra-cellular level; that is, it is a competition between mtDNA molecules. As the origin of replication is within the non-coding region of animal mtDNA, it is possible that variation in the A+T-rich region may alter the efficiency at

which the molecule is replicated. Again, changes in mtDNA conformation introduced by the changes in size could affect the interaction between the DNA molecule and the replication machinery. In order to understand this, it is helpful to visualize how new mtDNA variation is inherited and incorporated into the population.

In order for a new mtDNA variant to become fixed within a female germ-cell lineage (which is effectively equivalent to fixation within an individual due to maternal inheritance), the new variant must become fixed within the mitochondrion, and the variant-bearing mitochondrion must become fixed within the lineage. The dynamics of this process can be described in terms of Wright-Fisher sampling theory. In this, fixation or loss of a new variant is determined by intra-lineage sampling as germ-line cells go through cell division through the life of the organism. Replication of mitochondria and mtDNA within a cell is not discrete. Rather than copying itself once each cell generation, mtDNA molecules are randomly selected for replication throughout the cell cycle (Bogenhagen and Clayton, 1977), such that the copy number roughly doubles just before cell division. The variance of replication among mtDNA molecules will allow new variants to go to fixation within a mitochondrial lineage. Like the situation for neutral variants in a diploid Mendelian population, the probability of fixation of neutral mitochondrial variants within a germ-cell lineage will largely depend on the rate of neutral mutation.

If however the replication of a new variant were hampered slightly, then the probability of fixation would be decreased. Conversely, if the replication of a new variant was enhanced relative to the existing copies, then the probability of fixation would be enhanced. One would expect, therefore, depending on the magnitude of the replicative advantage/disadvantage, to see a predominance of one type across all otherwise neutral categories. This I suggest is the case in D. melanogaster mtDNA. mtDNA molecules become replicatively disadvantaged as they become progressively larger than the 18.6 kbp class, with the result that the 18.6 kbp class is predominant. Experimental support for this is mixed. Rand and Harrison (1986) found a transmission bias in Gryllus crickets, whereby offspring of females heteroplasmic for size classes tended to show a higher frequency of the smaller size class. A somewhat different result was found by Solignac et al (1987). Using D. mauritiana females heteroplasmic for size variant mtDNAs, they found that eggs laid by young adults had a lower proportion of the larger mtDNA size class, but that the proportion increased as the female aged, until the larger sized mtDNA was at a higher proportion than that in the young female. The authors point out that eggs laid in young adult Drosophila are the products of stem cell division in development (i.e. embryonic and larvonymphal stages), while those laid in older adults are the product of stem cell division in the adult. The implication is that any relative advantage/disadvantage to different sized

mtDNAs may depend on the state of cellular differentiation. In this case, smaller mtDNAs would be at an advantage in stem cells during development, but larger mtDNAs would have an advantage in the adult. In a resource limited environment where reproduction is limited to individuals early in their reproductive competency, the gradual loss of large sized mtDNAs from natural populations would result.

One other issue raised by the distribution of A+T-rich region size variants is the virtual absence of variants below the dominant 18.6 kbp size class. Species of the melanogaster species complex differ from other Drosophila species in having a clearly larger A+T-rich region. Two distinct structural changes to the mtDNA A+T-rich region of these species seem to have occurred. The first is the duplication of the 470 bp repeat to at least three copies; there is only one copy of this sequence in species outside the complex. The other change is an increase in size of the portion of the A+T-rich region to the rRNA gene side of the 470 bp tandem repeat cluster by about 1 kbp. As no restriction sites have yet been found in this portion of the A+T-rich region, it is not known if this increase is due to a different type of repeat unit or to a single insertion event.

The conservation of restriction sites within copies of the 470 bp repeat in D. simulans, D. sechellia and D. mauritiana suggests a mechanism of concerted evolution involving loss and re-duplications of repeat units. Therefore it is possible to generate variants smaller than 18.6 kbp

within the species of the melanogaster species complex. That this occurs in D. melanogaster is evidenced by the observation of two mtDNAs at 18.2 kbp. Yet the distribution is very asymmetrical, indicating that there is some sort of functional minimum to the 18.6 kbp size. The nature of this functional barrier is very difficult to determine since smaller mtDNAs in Drosophila species outside the complex are clearly viable.

While the probability of fixation of a new variant is dependent on the rate of neutral mutation under a simplified Wright/Fisher model, the time to fixation is dependent on the number of mitochondria per germ cell lineage, mtDNA molecules per mitochondrion, and number of cell divisions per organismal generation. Solignac et al (1984) showed mathematically that as the number of mitochondria per germ-cell increases, so does the number of generations required for alternate mtDNA variants to completely segregate. In this view, heteroplasmy is strictly a transitional state in the process of fixing a new variant, not a state that is being actively maintained. The transition will seem more stable when the number of mitochondria per cell is high. This seems to be the case in Drosophila, and in other insects. Solignac et al (1984) estimate that there are about 400 mitochondria per female germ-cell (oocyte) in D. mauritiana. With an estimated 10 germ-cell divisions, it would take hundreds of generations for complete sorting to occur.

Size heteroplasmy has not been observed in the several studies with mammalian mtDNA. This is not too surprising since

there are more cell divisions per generation and fewer mitochondria per germ-cell; this leads to the prediction that fixation of new variants will generally be rapid in mammalian systems. It is more surprising that there have been virtually no observations of individuals heteroplasmic for site variant mtDNAs. A line heteroplasmic for a HaeIII polymorphism was observed in the early part of this research (Hale and Singh 1986), and one has been inferred to have occurred in a cow lineage (Hauswirth and Laipis 1982). With the apparent long time to segregation of variants, it might have been expected that some site heteroplasmic lines would have been detected. However, as pointed out by Bermingham et al (1986), site heteroplasmy would be easily confused with partial digestion as the heteroplasmy could only be detected with the single enzyme rather than all enzymes as with size heteroplasmy. If the lack of cases of site heteroplasmy is genuine, and not an artifact of lack of observational rigour, then this may be indicative of the much higher rate of size mutation over site mutation.

No lines were observed to be heteroplasmic for small-scale mobility polymorphisms of the 0.26 kbp MboI fragment. This could be considered somewhat odd, since the frequency of generation of variants in this fragment is apparently quite rapid. However, the limit of resolution of the non-denaturing acrylamide gels was 3 bp. It may be that the shift in mobility for any given substitution was simply insufficient to allow the different form to be resolved on

these gels.

4.2 Mitochondrial DNA in Population Genetics

Natural selection can confound any study of species history by favoring alleles that give the population a reproductive advantage. There is considerable indirect evidence that the population structure of D. melanogaster has been heavily influenced by various forms of natural selection. Several traits show variations due to latitude, often resulting in parallel clines on different continents. Again this is particularly clear in allozyme studies where at least eighteen different loci show parallel latitudinal variation on three different continents. Similarly, latitudinal clines have been observed for chromosomal arrangements, some morphological traits, egg production, among others (reviewed in Lemeunier et al, 1986).

The utility of mtDNA in distinguishing historical influences on such clines is shown with populations of the horseshoe crab Limulus polyphemus (Saunders et al 1985). Allozyme studies had shown moderate divergence between Atlantic coast and Florida Gulf coast populations. mtDNA analysis showed there to be a sharp discontinuity in the species at the point where the Florida/Georgia border meets the ocean. This demonstrated that a constriction in gene flow at this point may be in part responsible for the differentiation observed at the allozyme level.

4.2.1 Other molecular markers

Obviously, mtDNA is not the only neutral marker in the animal genome. It is increasingly clear that much nuclear DNA variation is neutral as well. The problem, however, has been how to distinguish the specific loci for which variants are effectively neutral, and also how to work out genealogies for the variants of such loci. Identifying neutral markers has been hampered by the inability to assign a measure of divergence between within-species variants at a single locus. Such a measure is important, as a key prediction of neutral theory is that both within-species heterozygosity and between-species divergence will be a direct function of the rate of neutral mutation at each locus. As a result, tests of neutrality have been oriented toward degrees of departure from neutral expectations of sample heterozygosity for allozyme alleles (eg. Watterson 1977; Lewontin & Krakauer 1973, 1975), but these have little statistical power (Li 1979, Ewens 1979). Alternatively, studies have been directed toward the predicted mean and variance of the molecular divergence between species (Gillespie, 1986), as such measures can usually be assigned when long periods of time are involved. However, conclusions based on the long term evolution of a molecule cannot be used to predict neutrality below the species level. Long-term constancy of the rate of evolution is an average value, and may include periods of selection within particular species.

The application of molecular biology to population studies is still in relative infancy. Kreitman (1983)

demonstrated the potential power of sequence data to population genetics when he sequenced the Adh locus from eleven D. melanogaster chromosomes isolated from nature. One of his findings was that the amino acid substitution responsible for the fast/slow enzyme electrophoretic polymorphism was the only amino acid substitution in the entire sample. Also, heterogeneity in the frequencies of synonymous substitutions and substitutions in non-coding regions (eg. introns) cast doubt on the assumption that synonymous substitutions are strictly neutral.

Unfortunately, population genetic applications of DNA sequencing have not been fast in coming. Sequencing is a time-consuming and laborious procedure because of the requirement to clone the desired gene from each individual sampled (however, the polymerase chain reaction (PCR) sequencing technique promises to change this). As such, sequencing studies have been mostly directed toward inter-specific variation, as small sample sizes will yield sufficient information about the relative abundance of different sorts of base substitution that conclusions can be reached with some confidence.

Studies at the population level have largely utilized restriction analysis, particularly by four-cutter enzymes, since the cloning step is not needed for each chromosome isolated. Since each restriction site is expected to appear every 256 bp (assuming an equal GC:AT ratio), usage of several of such enzymes will result in a significant proportion of all

base substitutions being detected. A major price for technical ease, however, is that it usually cannot be determined what kinds of substitutions have taken place. Studies of restriction variation of nuclear genes have been carried out for populations of D. melanogaster with Adh (Kreitman 1983; Kreitman & Aguade 1986; Aquadro et al 1986; Aguade 1988), the white locus (Langley and Aquadro, 1987), rosy (Aquadro et al 1988) and the Zeste-tko locus (Aguade 1989).

Data such as these are conducive to testing predictions of the neutral theory, regarding both polymorphism and divergence. Hudson, Kreitman and Aguade (1987) have used this method to compare polymorphism and divergence within and between D. melanogaster and D. sechellia at the Adh locus, for which considerable sequence data is available. They found that there was approximately four-fold more polymorphism in the coding region of the Adh gene than would be expected on the basis of observed divergence. As there was approximately equal divergence between species for both the coding region and the 5'-flanking region, the authors interpreted their findings as being indicative of balancing selection in the coding region.

Even if a single neutral allozyme locus can be identified, however, inferences of species history can be hampered by intra-genic recombination of variants, which makes genealogical reconstruction difficult. For example, there is no linkage disequilibrium for polymorphic restriction sites in the Zeste-tko genes (Aguade et al 1989) and no phylogeny for this locus can reasonably be derived. The occurrence of

segregating polymorphism for each variable restriction site among populations indicates that the two North American populations are more closely related than either is to the Japanese sample. Yet without genealogical information, it is impossible to infer whether unique variants are endemic or due to recent immigration. Some progress is being made in methods of genealogical reconstruction for recombining genes, but there is still much to be done. The non-recombinogenic transmission of mtDNA makes such inferences not only possible, but rather simple.

4.2.2 mtDNA and intra-specific phylogeography

The importance of genealogical information together with population structure in the study of species history (termed 'intra-specific phylogeography' by Avise et al (1987)) can be appreciated by considering a hypothetical example in which allopatric populations of a species are polymorphic yet completely diverged for a given marker. If the types within each population form a monophyletic branch within the genealogy, then it argues that there has been historically little contact between the populations. Increasing degrees of departure from monophyly indicates increasing degrees of contact between the populations.

Avise et al (1987) have generalized the outcomes of joint studies of intra-specific mtDNA genealogy and population structure into broad categories based on alternate states of two criteria. The first criterion is phylogenetic continuity

of variants. A 'continuous' phylogeny means that the variance of the number of site differences between adjacent haplotypes on the genealogical network is small. In contrast, a 'discontinuous' phylogeny will see groups of haplotypes separated by a several mutational events, yielding a large variance in site differences between haplotypes. The theory of coalescents states that in a stable population, independently evolving gene lineages will be lost stochastically, and that at any given time all the gene haplotypes in the population will trace back to a single common ancestral type (reviewed in Hartl and Clark 1989). When an equilibrium determined by the mutation rate and the population size is reached, the elapsed time since the last common ancestral type will be constant, and so will the observed total variability. The resultant genealogy will be continuous and will remain so. Periodic bottlenecks will decrease the effective population size, and make the time interval since the last common ancestor much smaller.

If the population becomes subdivided into two or more isolated groups, then the stochastic loss of lineages present at the time of isolation will be severely hindered, as drift will occur independently in the sub-populations. This will lead to a discontinuous phylogeny as stochastic loss of lineages will occur <u>within</u> populations, but the last ancestor common to all populations will remain; the distance to the extant types will increase with time (Avise et al 1984).

The other criterion is the distribution of

monophyletic groups of haplotypes across geographic regions, which is taken to indicate levels of recent gene flow. Such groups can either co-occur in regions indicating very high gene flow, or be spatially separate and non-overlapping indicating very little gene flow.

Avise et al (1987) describe four categories based on extreme alternative states of each criterion: continuous vs. discontinuous genealogy, and spatially separate vs. overlapping geographic distribution. The literature documents species for which the mtDNA phylogeographic distributions fit all of the distinct categories. For instance, southeastern U.S. populations of the bowfin fish Amia calva show phylogenetic discontinuity and spatial isolation of haplotypes. These populations can be divided into eastern and western assemblages, between which haplotypes are separated by at least four restriction site changes. Within assemblages, haplotypes are separated by only one or two changes. At the other end, mtDNA haplotypes in populations of the American eel Anguilla rostrata along the American eastern seaboard are phylogenetically continuous and spatially co-occurring, as each haplotype was found in geographically distant locations. These eels spawn en masse in the mid-Atlantic, and therefore for purposes of reproduction they approximate a single panmictic population, rather than a sub-divided one.

In addition, Avise et al (1987) describe an intermediate category that appears fairly regularly in their surveys, in which pylogenetically continuous haplotypes are

neither entirely spatially separate or co-occurring. Rather, there are pockets of phylogenetically close haplotypes against a 'background' of dispersed ones. An example is within the eastern assemblage of A. calva; a single haplotype is found throughout the range, while others have a much more restricted distribution. The phylogeographic distribution of D. melanogaster is most like this intermediate category. The genealogy is largely continuous, although it could be argued that there is a gap between haplotype #1 the trifurcation leading to haplotypes #12, #19 and #23. Many haplotypes are widespread in distribution, and there are nests of related haplotypes (eg. #17 & #21 in West Africa, #5 & #18 in southeast Asia).

The situation with D. simulans can be viewed in two ways. If the sII type data is considered separate from types sI and sIII, then the phylogeographic distribution is continuous and co-occurring. Phylogenetically adjacent haplotypes are separated by a single site, and all haplotypes are found in geographically distant locations. If, however, the total species is considered, then the distribution would be better described as non-continuous and spatially separated. The sI and sIII mtDNA types are separated from sII and each other by many steps, and are found in isolated locations in allopatry to the others. This immediately suggests that the history of D. simulans is complex.

4.2.3 mtDNA and population genetic theory

In 1983, Birky et al proposed that in a Mendelian population the effective copy number of organelle genes would be one-fourth that of nuclear genes. This was based on strict maternal transmission of mtDNA, and rapid (ie. one organismal generation) vegetative segregation of new variants. The copy number of mtDNA genes is equal to the number of females, while the copy number of nuclear genes is equal to twice the number of males and females. Unfortunately, standard population genetic formulas for nuclear genes (eg. $F=1/4Nem+1$) cannot simply be re-expressed for mtDNA by making a simple 'factor-of-four' correction, since certain assumptions underlying this idealized difference may not hold. For instance, inheritance may not be strictly uniparental, a common occurrence with plant organelles (Palmer 1985a). Although tests for biparental transmission in animal systems have yielded negative results (Reilley and Thomas 1980, Lansman et al 1983b), it cannot be entirely ruled out. Also, relative effective copy numbers will be sensitive to male/female biases in breeding and migration. A female breeding bias will lead to a smaller difference between nuclear and mitochondrial numbers, and contrarily so with a male bias. Similarly, a female migration bias will mean a less pronounced difference with regard to mitochondrial vs. nuclear gene flow, with a male bias leading to a more pronounced difference.

Perhaps more critical is the assumption of rapid

vegetative segregation. The observation of stable heteroplasmy in Drosophila (this study, and Solignac et al, 1984) and other insect species (Rand and Harrison 1986, 1989) is clear evidence that this is not occurring. The effect of heteroplasmy on organelle population genetic theory is to increase the effective copy number, since the female is transmitting two, and sometimes three (Rand and Harrison 1989, Solignac et al 1986) alleles. Birky et al (1983) formalized this as part of a series of hierarchical diversity indices, which measures the 'allelic' diversity at several levels. Two of these indices are of interest in the present context. One of these indices, K_a, represents the probability that two mtDNA molecules sampled from a single adult cell will be different. K_a in the germ cell is related to the corresponding measure in the zygote, K_z (i.e. the probability that two copies of mtDNA sampled from a single zygote will be different), by a matrix function involving the rate of neutral mutation, the number of mtDNA copies in a cell, and the number of cell divisions in a cell lineage per animal generation. K_a will be K_z in the subsequent generation.

Birky et al (1989) explored theoretically the consequences of fluctuating K_a on interpopulational diversity (G_{st}). They found that K_a above a certain level (the critical level varies somewhat with mutation rate) will result in significant error in the estimates of some population genetic measures, notably interpopulational differentiation. Below these critical values, the error is very small. K_a can be low

for two reasons; either the segregation of variants is rapid, or the mutation rate is very small. Size heteroplasmy in D. melanogaster mtDNA is evidence that segregation is not rapid. In examining the preliminary data of mtDNA size variation of this species (Hale and Singh 1986), Birky et al (1989) calculated a K_a of 0.033-0.091. However, studies on the dynamics of thermal renaturation of Drosophila mtDNA points to a rate of base substitution of only 1.7×10^{-8} (Caccone et al 1988). This makes it unlikely that for sequence (site) variants the K_a in Drosophila species will be above the critical value to affect observations of interpopulational diversity. Therefore, the slow segregation of variants indicated by size heteroplasmy is not an important concern.

While the difference in effective copy number between the two genomes may or may not be precisely 1:4, it is clear that the copy number of mitochondrial genes is smaller. This has two important effects. First, mtDNA will be more susceptible to stochastic effects, particularly when the effective population size is small. Random genetic drift will be more pronounced for mtDNA variants. Second, the copy number of mitochondrial genes transmitted by migrants will be lower than the copy number of nuclear genes. Birky et al (1989) show that under the assumptions of approximately equal breeding and migrating sex ratios, equal mutation rates and negligible K_a, the ratio of population subdivision (as measured by G_{st}) for nuclear to mitochondrial genes will be approximately 1:4. This agrees well with the data for D. melanogaster. Overall, the

ratio is approximately 1:5 (Table 4.1), while the ratios are somewhat higher for two of the three regions. The theoretical ratio (1:4) is that predicted for a neutral mutation/drift equilibrium situation, a situation that has not been reached species-wide in D. melanogaster. However, the G_{st} values of mtDNA are predicted to approach their equilibrium quite quickly, even as inter- and intra-populational diversity is changing.

Populations will be less affected by gene flow in terms of mtDNA variability than they will be for nuclear genes. It is possible that, at a moderate-to-low level of migration, populations can appear subdivided for mtDNA, yet panmictic for allozyme variation. This has two key ramifications. First, the lower sensitivity of mtDNA to homogenization through migration means that it is less likely that historical relationships between populations will be lost as a result of gene flow. This is important since these Drosophila species show evidence of high gene flow (Singh and Rhomberg 1987a). Second, it makes comparative inferences of selection or neutrality of allozyme population structure difficult to prove. mtDNA differentiation does not necessarily mean low gene flow between populations, and an accompanying lack of allozymic population structure would not necessarily be due to similar selection regimes in these populations. However, the problem arises only when one is dealing with observations on only one set of populations, and as we will show later, the problem can be removed with parallel

Table 4.1

A summary of variation statistics for mtDNA and allozymes from D.melanogaster[a]

Region	Data type	Average intra-populational diversity (H_s)	Regional diversity (H_t)	Fixation index (G_{st})
Euro-African	mtDNA	0.106*	0.196*	0.459
	allozymes	0.130	0.132	0.106
Western Hemisphere	mtDNA	0.033*	0.059*	0.441
	allozymes	0.101	0.127	0.079
Far East	mtDNA	0.056*	0.203*	0.724
	allozymes	0.098	0.112	0.097
World-wide	mtDNA	0.064*	0.188*	0.660
	allozymes	0.102	0.138	0.126

a) - Allozyme data from Rhomberg and Singh (1987a) and unpublished.

* = x 10^{-2}

observations of mtDNA and scnDNA variation with independent sets of populations.

4.3 mtDNA population structure in *D. melanogaster* and *D. simulans*

4.3.1 Nuclear genetic structure and possible causes

The widespread distribution of *D. melanogaster* and *D. simulans* is believed to be due to man-mediated transport since, outside of Africa, they are generally found in human-influenced situations, such as urban or agricultural areas. As a consequence, they are regarded as 'commensal' species (Parsons 1975, 1983). They both occur in temperate and tropical environments, although *D. simulans* is somewhat less successful in temperate regions (McKenzie and Parsons 1974). *D. simulans* therefore has a more restricted range than *D. melanogaster*. Neither species is able to over-winter in natural settings in temperate areas, although temperate populations do carry-over to the spring by inhabiting human dwellings. Coyne (1987) has also found that temperate populations may also be partially restocked by migration from warmer areas. Ecologically, the two species appear to be very similar in niche space; morphologically, they are almost identical.

Despite the remarkable similarity in range, ecology, and morphology, there are distinct differences between the two species in terms of geographic genetic structure. Generally, *D. simulans* has been found to be less variable than

D. melanogaster, both in terms of heterozygosity within populations and differentiation between populations (reviewed in Lemeunier et al 1986). Choudhary and Singh (1987) reviewed a series of possible explanations for these differences in population structure. One of these, the mutator-gene hypothesis, proposes that there is a higher rate of gene-mutation in D. melanogaster due to a greater number of movable (transposable) sequences (Brookfield et al 1984) and mutator loci (Green 1976), which are far less common in D. simulans. However, this hypothesis fails to explain the differences in geographic structure between the two species.

Another hypothesis is based on the idea that much of the differences between populations is due to differences in species history. The recent colonization hypothesis suggests that the lack of geographic differentiation in D. simulans is due to a recent expansion of D. simulans from ancestral stocks in Africa. D. simulans would simply be further from equilibrium than D. melanogaster, as the rates of gene flow for the two species appear to be similar (Singh and Rhomberg 1987a). Previously, a bottleneck of the ancestral D. simulans stocks was suggested to have occurred previous to expansion, in order to explain the comparatively low variability (Singh et al 1987b). The results of Choudhary and Singh (1987), showing the effectively equal variability in D. simulans, places the need for that in doubt.

Two hypotheses invoke natural selection as the agent of population structure. One of these, the niche-width

hypothesis, is based on the proposal that the amount of genetic variability found in a species will be in proportion to the range of resources that it can utilize and the range of environmental conditions that it can tolerate (VanValen 1965). Experimental support for this view comes from both natural (Steiner 1977) and laboratory (Powell 1971, Powell and Wistrand 1978) sources. The implication of this hypothesis, therefore, is that despite their similar distribution and apparently similar resource utilization, D. simulans is more restricted in some environmental parameter than D. melanogaster. Choudhary (1988) has suggested that the critical parameter might be temperature tolerance, based on the lower success of D. simulans in temperate regions and its tendency for later-season onset of reproduction.

The other proposal rooted in natural selection is that D. simulans has developed the genetic strategy of a 'general purpose genotype' (Baker 1965). Under this hypothesis, the genetic constitution of D. simulans allows adaptation to different environments to proceed without appreciable alteration in allele frequencies. By contrast, D. melanogaster would be considered to have adopted a strategy of fine-tuning to particular environments, with more distinct shifts in allele frequency with changing environments. That there are fewer latitudinal clines of allele frequencies in D. simulans that in D. melanogaster has been cited as supporting this view (Choudhary 1988).

mtDNA was used to help assess the degree to which

allozyme population structure has been influenced by history and natural selection. The two options differ sharply in the predicted effects on mtDNA as a defined 'neutral' marker. Under the historical 'recent expansion' hypothesis, mtDNA variants in D. simulans should also be relatively unstructured as historical events will affect all genetically determined characters. Alternatively, if mtDNA variants are strongly differentiated, then it argues for a role of selection in reducing geographic differentiation.

4.3.2 mtDNA structure and history in D. melanogaster

There are few studies on intra-specific variability in other Drosophila species. Yet from what is available, it appears that there is generally less variation in D. melanogaster mtDNA than in other Drosophila species. Latorre et al (1988) report an average nucleotide diversity in D. subobscura of 1.1×10^{-2}, with the most diverged haplotype pair having a diversity value of 3.3×10^{-2}. DeSalle et al (1986b) report that two species of Hawaiian Drosophila, D. silvestris and D. heteroneura, have intra-specific mtDNA diversities of 4.9×10^{-2} and 3.3×10^{-2}, respectively. There is some data for D. pseudoobscura, D. persimilis and D. miranda (Powell 1983, Hale and Beckenbach 1985), but these are based on only a few r = 6 enzyme, and may lead to a seriously erroneous estimate (Latorre et al 1986, 1988).

As pointed out earlier, genetic structure can be observed as geographical or temporal. Geographical patterns

of differentiation can occur with regard to latitude, longitude, or altitude. Populations of D. melanogaster show clines of all four types among various genetic factors, of which latitudinal variation has been the most fully described. In this study, latitudinal and longitudinal variability was assessed against mtDNA.

One of the key findings of this work is the differentiation of the species range into three distinct regions, separated along longitudinal lines, by mtDNA data alone. The three regions (Euro-Africa, Far East, and Western Hemisphere) were distinguished on the basis of levels of intra-populational variability and inter-populational divergence. Euro-African populations are highly polymorphic and moderately diverged, Far East populations are much less polymorphic but strongly diverged, and Western Hemisphere populations are nearly monomorphic and very similar to each other. It would seem therefore that these regions have very different histories. In the results section, I proposed a colonization history based solely on mtDNA that inferred Euro-African populations to be the oldest, and therefore ancestral to the other two regions. The Far East populations were inferred to be due to multiple independent invasions from a polymorphic population, or due to local bottlenecks fixing alternate variants. The Western Hemisphere populations were inferred to result from a single founder group of flies into the Western Hemisphere.

Since history will affect all genetic determinants,

there should be some evidence that these regions have different histories at other levels of genetic organization. As it turns out, there is a rather remarkable agreement between the mtDNA data and that of other markers. David and Capy (1988) have summarized findings from chromosomal arrangements, allozymes, morphometric and physiological traits, and proposed a history of D. melanogaster based on the sum of these data. Like the present study, they divide the species into three categories. 'Ancestral' populations are those found in tropical Africa, and were so designated because allozyme diversity is greater and because more unique allozyme alleles are found there. Also, other melanogaster sub-group species are endemic to this region.

Populations in Europe are not part of the ancestral populations, but are included in the second category, 'ancient' populations. The mtDNA data cannot distinguish the two European populations (England and France) from the two African populations. Both show similar levels of diversity as in the African populations, and harbour high frequency unique haplotypes. These are among the criteria by which the Euro-African region was distinguished as being ancestral.

Also among David and Capy's (1988) 'ancient' populations are the Far East populations of this study. Tessier (1957) distinguished Japanese D. melanogaster from those in Europe on the basis of differences in body size and ovariole number. Other differences that have been found are DDT resistance (Melou 1961), phototactic response (Medioni

1958), and growth rate (David et al 1976). The range of flies showing these same properties extend into Taiwan, Indochina and Sri Lanka, and led David et al (1976) to describe them as a Far East 'race'. The range of this Far East race encompasses much of the Far East region described in the present study, with the exception of Australia.

Far East populations are not appreciably less polymorphic than their counterparts in Europe and Africa for several other traits (e.g. allozymes, chromosomal arrangements). However, while there is considerable inter-populational variability for many of the traits that distinguish the Far East race (developmental growth rate, adult weight and ovariole number), no correlation was found between these traits and latitude. This divergence between populations parallels the observation of strong differentiation for mtDNA variants.

David et al (1976) interpret the sharp differences between the Far East and Euro-African flies as suggesting that the Far East populations must be quite old, pre-dating human mass-transport. They suggest that because there is a continuous land mass between the African continent and the Far East, that D. melanogaster could have spread there naturally, with small isolated populations at the leading front, continually undergoing drift, and fixing the new variants typical of the Far East race.

The implication for mtDNA is that the founder population would likely be monomorphic for a single mtDNA

variant, as mtDNA is particularly susceptible to founder effects. Therefore, any endemic types should form a monophyletic group on the genealogy. However, the genealogy in Fig. 3.12 shows that this is not quite the case. While many of the endemic haplotypes derive directly from haplotype #1, haplotype #13 is derived from #12, which is found at high frequency in France. Therefore, in order for the haplotype composition of the Far East populations to be due solely to divergence from a single colonization stock, the founding group would probably have had to harbour both haplotypes #1 and #12, instead of just a single one. It is notable, however, that #13 is restricted to a single population, Korea, in which was also found a singleton of another haplotype also found in France. It may be, therefore, that Far East populations are old, and that the effects of sequential introduction, of the type suggested in the results section, could be restricted to Korea.

If the regional diversity seen by mtDNA is due to a build up of new mtDNA variants, why the low inter-populational diversity? David and Capy (1988) suggest that the Far East race may have arose before the last glacial period. As a result, the conditions may have been right for a species bottleneck in the region. If remnants were able to survive in isolated pockets, then each would harbour a different mtDNA variant, thus preserving the total diversity in the face of low numbers of flies.

Absent from the Far East race of David and Capy (1988)

is the Australian population. This was placed in the third category, that of 'new' populations. 'New' populations are defined as those introduced by humans, presumably within the last 200-300 years, and includes North American populations. Australian D. melanogaster were classified as distinct from the Far East race as they do not display any of the distinguishing traits of the Far East race. The mtDNA data is wholly consistent with this. The Australian population is more variable for mtDNA than any other Far East population, harbouring four haplotypes among six lines sampled. The two unique haplotypes (# 11,19) do not cluster phylogenetically with other Far East haplotypes. The two non-unique haplotypes (# 7,12) are both found in Euro-African populations.

The other set of 'new' populations (that overlap with this study) are the North American ones. The observation that all North American populations are almost fixed for the same mtDNA variant indicates that all these populations derive from a single founder group with haplotype #7 reaching its high frequency by founder effects. However, David and Capy (1988) suggest that the colonization of North America by D. melanogaster was not a one-step process. The original introduction of D. melanogaster into the Western Hemisphere would be concomitant with the slave trade, which began in the early 1700's, in which tropically acclimated flies from Africa were introduced. The African origin of these flies is indicated by the similarity of tropical North American flies and African flies in the composition of cuticular

hydrocarbons, and in overall genetic composition. Further, they suggest that an inability to acclimate to colder temperatures initially prevented the spread of D. melanogaster into temperate regions of North America. This would have required an introduction of a separate cold-acclimated stock. Again, evidence come from the compostion of cuticular hydrocarbons in temperate North American and European flies.

The mtDNA data, however, does not show any evidence of such a two-stage process. Temperate populations of North American D. melanogaster (Ottawa, Hamilton, Massachusetts) are virtually identical to those in tropical zones (Florida, Texas, California), as all populations are almost fixed for the same variant. This similarity strongly suggests that both temperate and tropical populations derive from the same original introduction(s) of flies into the Western Hemisphere. Haplotype #7 is apparently a rare variant in Africa, and therefore it is somewhat surprising that it has been effectively fixed in the Western Hemisphere. Barring the possibility that there are isolated pockets of flies with a high frequency of #7 in both Europe and Africa, it is extremely unlikely that two independent colonizing introductions from Europe and Africa into the Western Hemisphere would lead exclusively to #7.

Eanes et al (1989) have also argued against a two-stage colonization of North America on the basis of restriction analysis of the G6PD locus from flies from temperate North America (New York State) and Europe (France).

They found that haplotypes characterized by a 4.2 kbp insertion, which are at high frequency in Europe, did not appear in the temperate North American sample, and suggest that this is inconsistent with large-scale introduction of European genes into North America.

The cuticular hydrocarbons of insects are complex, and therefore the similarities between Europe and temperate North America are not likely to be the result of convergent evolution for this trait. The challenge, therefore, is to reconcile the molecular genetic evidence with that of cuticular hydrocarbons. If the particular composition of the cuticular hydrocarbons of the temperate region could be shown to be advantageous in that environment, then appropriate hypotheses could be proposed in the context of either one- or two-stage colonizations. For instance, if a single original colonizing stock was polymorphic for genes determining cuticular hydrocarbons, then the spread of the flies north would be accomplished by groups harbouring higher frequencies of those alleles advantageous in temperate zones, possibly in response to gradually colder conditions. Alternatively, if the temperate zones were colonized by European flies and those populations were small, then migration of flies from tropical North America could result in the displacement of mtDNA and nuclear alleles to that found in the tropical populations by stochastic effects, while maintaining the temperate zone cuticular hydrocarbons by selection. Cuticular hydrocarbons are known to play roles in protection against dessication

(Edney 1977) and male courtship induction (Antony and Jallon 1982), but it is not immediately clear how such traits could confer an advantage in temperate environments. Further research into the physiology of cuticular hydrocarbons and molecular analysis of other nuclear genes in flies from temperate zones is necessary to fully resolve this.

The Argentina population was not classified by David and Capy (1988) since so little is known about it. The two possibilities for introduction of D. melanogaster into South America are a separate introduction from Euro-African populations (as suggested by those authors), or dispersal of flies from North America. Since the Argentinean haplotype (#8) is phylogenetically very close to #7, it seems more likely that it would be the result of a spread from North America. The presence of a continuous land bridge and the monomorphism of the population supports this option.

The history described by David and Capy (1988) for D. melanogaster is certainly more comprehensive than the one based on mtDNA alone. The inability of mtDNA to distinguish between Europe and African populations in terms of ultimate ancestry stands as an example. However, the history of David and Capy (1988) is based on many different studies of many different genetic determinants, involving a great deal of time and effort. For a single character, the mtDNA provides a remarkably good 'first approximation' (Fig. 3.14). By contrast, none of the other traits, by themselves, could provide such an approximation.

The situation with allozymes is a clear case in point. Although there is much evidence through other characters that Far East populations of D. melanogaster are far older than those of the Western Hemisphere, there is nothing in the allozyme data, either in terms of increased heterozygosity or a greater abundance of unique alleles, to indicate this. Similarly, there is nothing to suggest that Western Hemisphere populations are particularly new. The only allozyme trait that sets Western Hemisphere populations apart is that the 6PGD locus is polymorphic in North American populations with two alleles, while monomorphic elsewhere. Also, allozymes fail to distinguish the Australian population as distinct from other Far East populations. In terms of genetic divergence or genetic identity, Australia tends to be somewhat closer to nearby Far East populations (Vietnam and Taiwan) than to the Euro-African populations from which the recent colonizers came from.

4.3.3 mtDNA structure and species history in D. simulans

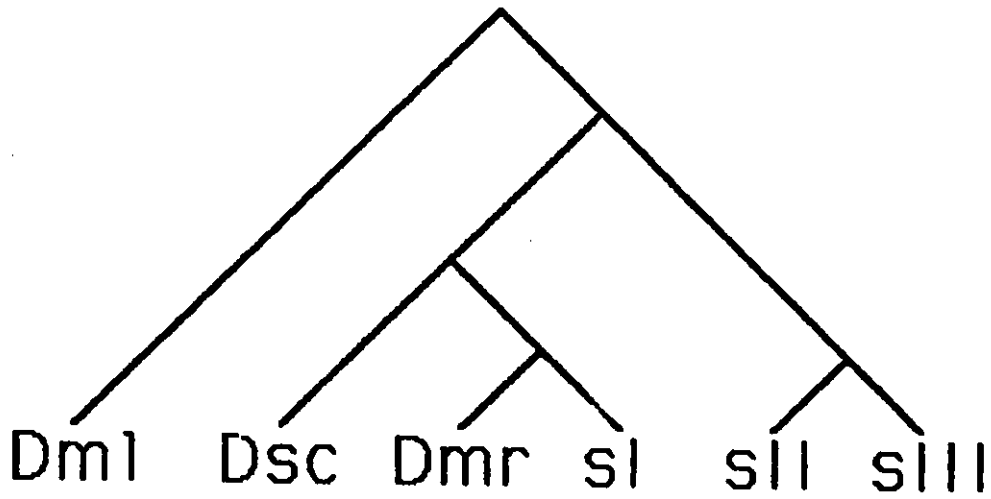
On the basis of the sharp discontinuity of the whole species mtDNA phylogeny alone, one is compelled to infer that the past history of D. simulans must have included the presence of long-standing isolated ancestral populations with strong barriers to mitochondrial gene flow, with each ancestral population leading to a different mtDNA type. This question has been reviewed and addressed in some detail by

Lachaise et al (1987). They suggest that D. simulans split from D. melanogaster on the African continent, but then retreated to the island of Madagascar, just off the east coast of the continent. A propagule located in the Seychelles would then have led to an sI ancestral population; both the sII and sIII mtDNA lineages are hypothesized to have arisen in Madagascar. The phylogeny of mtDNA types within the melanogaster species complex (Fig. 4.1) indicates that the divergence of the sII and sIII lineages was more recent than the sI divergence. Therefore, it is possible that sII and sIII shared a larger, highly polymorphic ancestral population, but subsequently split into two founder groups, each fixing a different diverged mtDNA type. Of these, the sII returned to the mainland, while sIII remained endemic to Madagascar.

Superimposed on this situation is the evolution of the other two species of the melanogaster species complex, D. mauritiana and D. sechellia. D. sechellia is endemic to the Seychelles, but allopatric to the sI D. simulans population. D. mauritiana is endemic to the island of Mauritius. These two are very close to D. simulans, as evidence by the homosequential arrangements of their chromosomal banding patterns (Lemeunier and Ashburner 1984). The mtDNA of D. sechellia and D. mauritiana form a monophyletic grouping with the sI mtDNA type of D. simulans (Fig. 4.1), which may suggest that they diverged from D. simulans after the establishment of an sI ancestral population. Within this triad, D. sechellia appears as an outgroup, suggesting that D. mauritiana is the

Figure 4.1

Phylogenetic relationship between the three major *D. simulans* mtDNA types and those of related species, as demonstrated by Solignac and Monnerot (1986). *D. simulans* types **sI**, **sII**, and **sIII** are as described in the text. **Dmr** is the type specific to *D. mauritiana*, **Dsc** is the type specific to *D. sechellia*, and **Dml** is a type in *D. melanogaster*.


Dml
Dsc
Dmr
sl
sll
slll

younger of the two island species, having split from D. simulans at a later date. Yet other genetical data often present very different relationships between the three species. Each potential arrangement of the species triad is supported by at least two types of data (reviewed in Singh 1989). Unfortunately, many of these data do not include D. simulans samples from the Seychelles population. A definitive phylogeny for these species is yet to come.

D. mauritiana harbours a second type of mtDNA, which has a restriction patterns identical to that of sIII type D. simulans mtDNA. The presence of sIII mtDNA in D. mauritiana would seem to indicate inter-specific hybridization between D. mauritiana and sIII-harbouring D. simulans, which are found sympatrically in Mauritius. F_1 female hybrids of D. mauritiana and D. simulans are fertile (Lachaise et al 1986), so a cross between female sIII D. simulans and male D. mauritiana, with backcross to D. mauritiana, would theoretically result in introgression of D. simulans mtDNA into D. mauritiana. Cage experiments show that the introgression of D. simulans mtDNA into D. mauritiana can be quite efficient, while the reverse introgression seems to be much less so (Aubert and Solignac 1988).

4.3.3.1 Continental populations of D. simulans

In contrast to the comparatively complex arrangement of mtDNA variants in D. melanogaster, the distribution of variants in continental populations of D. simulans would seem

to be more straightforward. The key features of the distribution of sII mtDNA variants are that 1) all observed haplotypes coalesce over a very short genetic distance, much shorter than that for D. melanogaster, and 2) these haplotypes are distributed in such a way that there are no 'unique' variants.

The phylogeny is that which would be expected from a local population that was recovering from a constriction in numbers that reduced the mtDNA complement to a single type. This indicates unequivocally that continental populations of D. simulans are recently derived from a single source stock. Yet it is not clear whether all of the four HinfI haplotypes were part of the source population. I have suggested that the observation that none of the haplotypes are 'unique' to a single locality argues that they were distributed from the source during the mass colonization by D. simulans, with local founder effects accounting for the difference in frequency between populations. However, it is also possible that one, two, or even three of the four sII haplotypes were derived after colonization, and were dispersed by recurrent migration between populations. In particular, if haplotype A or B arose post-colonization, the implication is that migration and gene flow can have a very strong influence on the mtDNA population structure in D. simulans as both have attained high frequency in dispersed localities.

The evidence pertaining to this point is inconclusive. The observation that the 'R' mating incompatibility type is

exclusively associated with haplotype B in both Europe and North America (Table 3.12), even though haplotype B is at low frequency in Europe (Table 3.11), suggests a single origin for the incompatability type, and therefore that long-distance recurrent migration is occurring. However, since this phenotype is being 'driven' by the outcome of asymmetric mating fertility, its dynamics may not be representative of other cytoplasmic traits. Also, the results from D. melanogaster, for which a similar level of gene flow has been inferred on the basis of rare allozyme allele analysis (Singh and Rhomberg 1987a), indicates that recurrent gene flow has been generally ineffective on mtDNA population structure. Nevertheless, there is some evidence for such effects in the Western Hemisphere populations of D. melanogaster (esp. BC, TEX) where haplotypes clearly unrelated to #7 are found at low frequency.

It is clear from the evidence with mtDNA that the species histories of D. melanogaster and D. simulans are very different. The world-wide colonization by D. melanogaster has been a comparatively slow, gradual process, with different histories in different geographical regions: Europe in prehistoric time, Far East in the historic time, and the Western Hemisphere in more recent time. The colonization of continental populations of D. simulans has, on the other hand, been quite recent and rapid. In fact, the dynamics of colonization by D. simulans can be seen as analogous to the colonization of the Western Hemisphere by D. melanogaster, as

both spread quickly from a single founding stock. These events may have been roughly concomitant with the advent of mass-transport by humans. Therefore, it terms of geographic distribution, it has taken D. simulans a comparatively short time to do what it took D. melanogaster a long time to do. With this idea in place, it remains to analyze the amount and distribution of allozymic (and other) nuclear genetic variants in terms of species history.

4.3.4 Comparison with allozyme data

4.3.4.1 Population structure

Many groups have undertaken surveys of allozyme variation in natural populations of D. melanogaster (Anderson and Oakeshott 1981, Berger 1970, Cavener and Clegg 1981, Franklin 1981, Oakeshott et al 1981, Singh and Coulthart 1982, Singh et al 1982, Voelker et al 1977) and, to a somewhat lesser extent, in D. simulans (Berger 1970, Cabrera et al 1982, DeAlbequerque and Napp 1981, Gonzalez et al 1982, O'Brien and MacIntyre 1969, Triantaphyllidis et al 1973, 1982, Watada et al 1986). The most comprehensive ones have been those of Singh and Rhomberg (1987b) in D. melanogaster and Choudhary and Singh (1987) in D. simulans, in which 117 and 114 loci, respectively, were surveyed in an attempt to reach more precise estimates of heterozygosity and differentiation.

In D. melanogaster, 15 populations were sampled, of which nine are the same as used in this mtDNA survey. Two main issues are raised in comparing the two systems; the overall

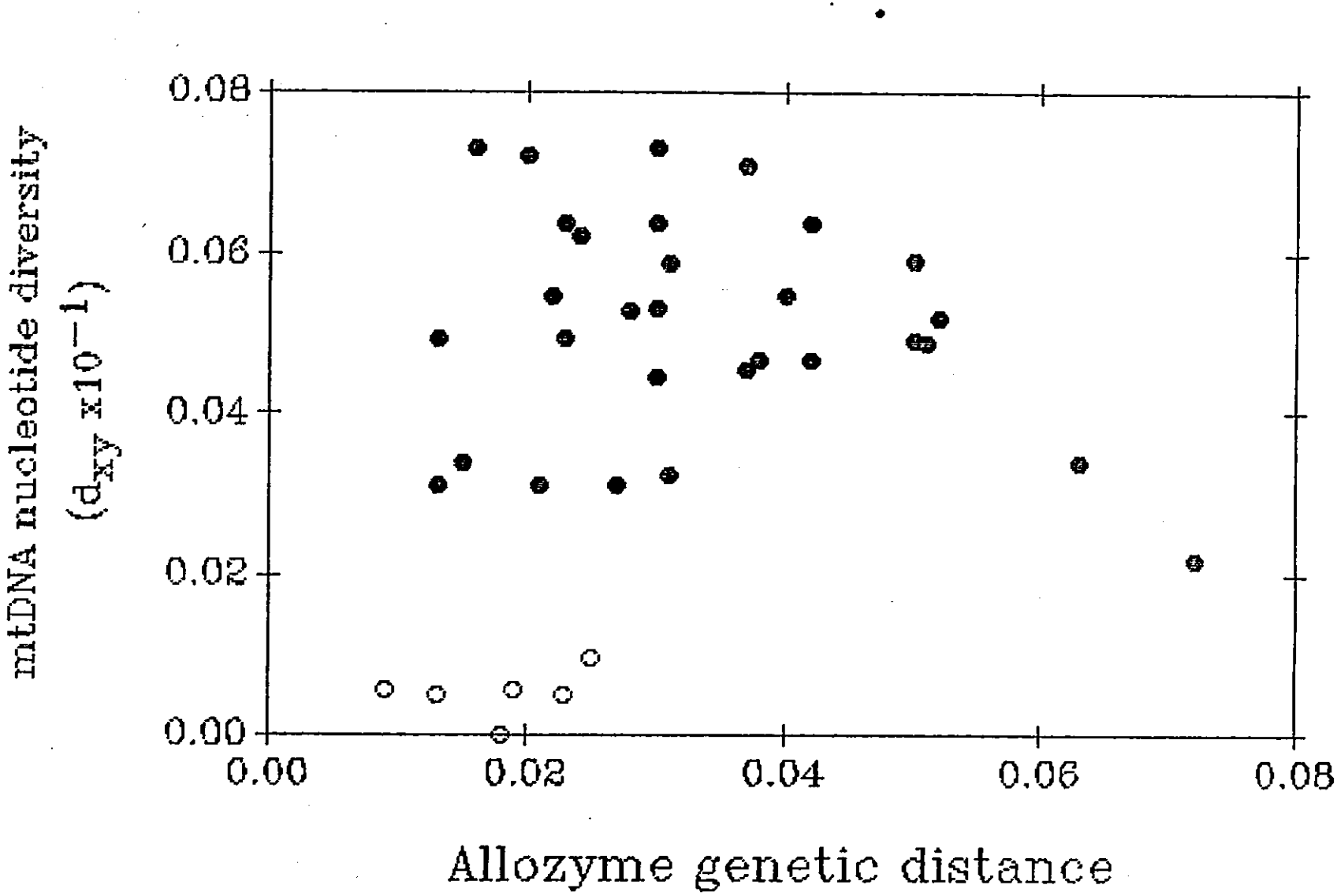
geographic differentiation and structure of populations, and the comparative abundance and pattern of distribution of variants at individual loci.

Table 4.1 gives a summary of relevant measures with respect to population structure of allozymes and mtDNA. Two points stand out. First, there is comparatively little difference between regions for the means of intra-populational heterozygosity over loci. Populations which are monomorphic for mtDNA (such as Hamilton and Taiwan) are equally as heterozygous for allozymes as populations (such as France and West Africa) which have highly polymorphic mtDNA complements. Second there is little difference in regional H_t or F_{st}. Specifically, the F_{st} in Far East populations is not appreciably higher than in other regions, and the regional H_t is not appreciably lower in the Western Hemisphere (the accompanying standard deviation terms do not allow statistical test of the differences).

These comparisons indicate that the inter-population differentiation of mtDNA is largely independent of allozyme variation. This is also seen in Fig. 4.2, which is a plot of mtDNA distance (d_{xy}) vs. allozyme distance for population pairs. Initially, there is a significant positive correlation between these measures ($p < 0.05$). However, all of the comparisons of populations within the Western Hemisphere, which are hypothesized to have shared a recent common history, cluster together near the origin. If these comparisons are omitted, then a non-significant negative correlation is the

Figure 4.2

Correlation of inter-populational genetic distance from mtDNA and allozyme analysis for nine D. melanogaster populations. Open circles denote comparisons within the Western Hemisphere, which have been omitted from the calculation of regression and correlation (see text). $Y = 0.0579 - 0.21X$. Correlation coefficient (r^2) = 1.4% ($p > 0.05$).


mtDNA nucleotide diversity
(d_{xy} x10^{-1})
0.08
0.06
0.04
0.02
0.00
0.00
0.02
0.04
0.06
0.08
Allozyme genetic distance

result ($p > 0.05$). In the absence of any explanation for a negative correlation, it is probably safe to infer that there is no meaningful correlation at all.

Sixty-one of the 117 gene-enzyme loci in D. melanogaster were found to be polymorphic. The abundance and distribution of variation at polymorphic loci differs considerably from locus to locus. The distribution of values for inter-populational differentiation of loci (F_{st}) is strongly skewed. Although the range of values is from 0 to 0.68, there is a prominent mode within a bell-shaped distribution around $F_{st} = 0.08$, and then a long tail of higher values. Within the mode are loci with little total heterozygosity and ones with high heterozygosity. As such, differentiation is not a simple function of variability. In fact, loci with intermediate values of total heterozygosity are somewhat more differentiated between populations than those with a higher H_t.

Among those allozyme loci with a higher F_{st} are 18 loci that show parallel clinal variation along temperate/tropical transects in each of the three regions (North America: OTT-TEX; Euro-African: FRA-WAF; Far East: TAI-VIE). Singh and Rhomberg (1987b) showed that although there is a positive correlation between allozyme distance vs. great circle geographic distance, it is weak because temperate populations are quite similar in their allozyme complement, while tropical populations tend more to differentiate in correlation with geographic distance. The results for mtDNA

are quite different. First, there is no detectable correlation between mtDNA distance vs. geographic distance (Fig. 4.3) (like Fig 4.2, there is a non-significant negative correlation if the population pairs within the Western Hemisphere are omitted). Second, all of the temperate populations are completely diverged from one another, and similarly so with the tropical populations. There are no shared haplotypes with each set of three populations. In addition, the pairs of populations show sharply contrasting degrees of differentiation, concomitant with the general results for each region. The Western Hemisphere pair is almost identical, the Euro-African pair is moderately diverged, and the Far East pair is alternately fixed. As the temperate and tropical populations all have different histories, the latitudinal allozyme clines observed in each region are not historical in nature. Natural selection to some latitudinally changing environmental variable is the likely agent of establishing these clines.

Choudhary and Singh (1987) analyzed 114 loci in D. simulans populations along a Euro-African transect (France, Tunisia, Congo and South Africa) and from a population on the Seychelles Islands. A summary of some relevant statistics is presented in Table 4.2.

Like D. melanogaster, populations of D. simulans do not differ much in intra-populational heterozygosity. The mean values of heterozygosity also do not significantly differ between the two species. This is reflected in the distribution

Figure 4.3

Correlation of inter-populational mtDNA genetic distance and geographic distance between nine D. melanogaster populations. Open circles denote comparisons within the Western Hemisphere, which have been omitted from the calculation of regression and correlation (see text). $Y = 0.0689 - 2.7 \times 10^{-6} X$. Correlation coefficient (r^2) = 10.4% ($p > 0.05$)

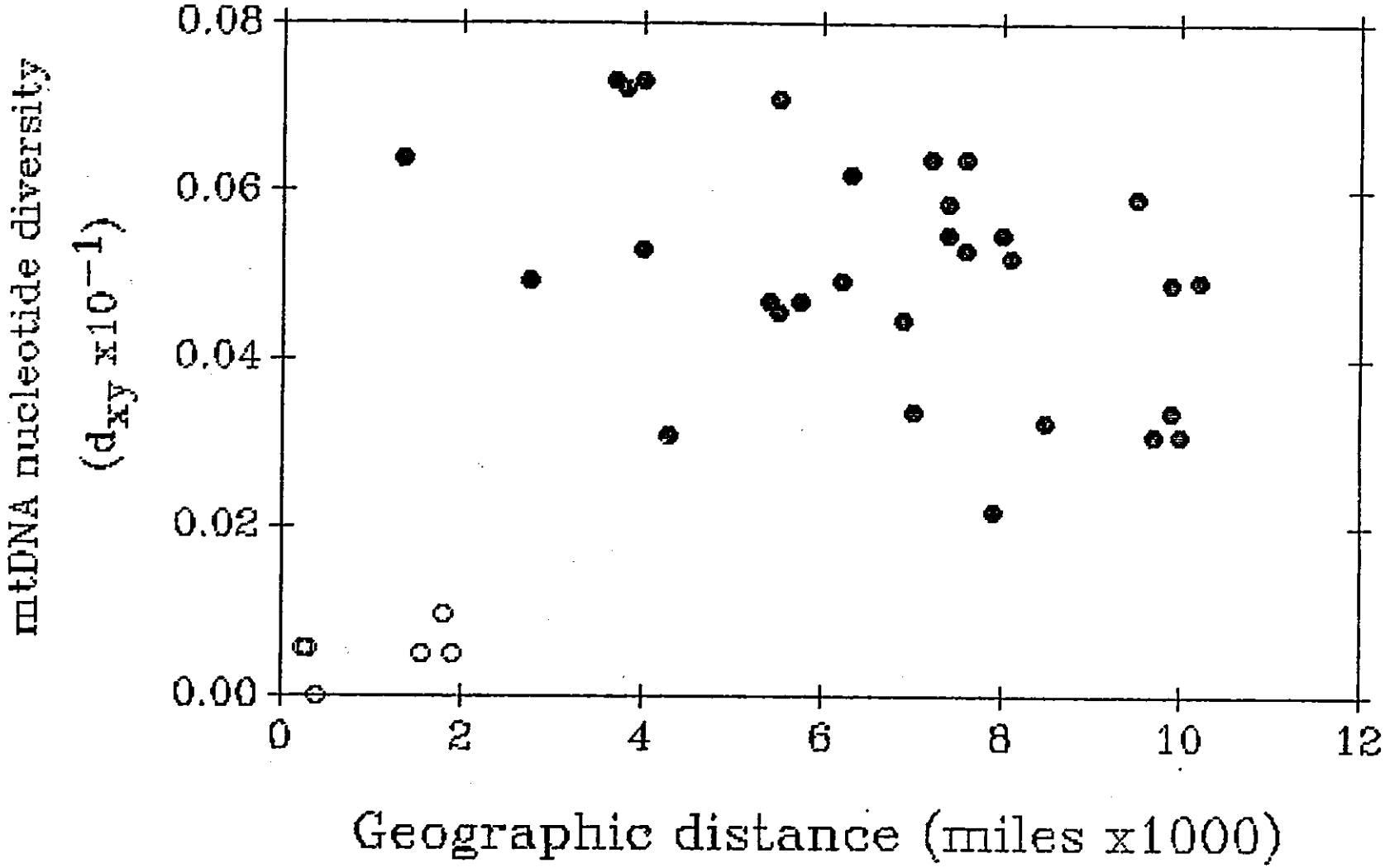

mtDNA nucleotide diversity
(d_{xy} x10^{-1})
0.08
0.06
0.04
0.02
0.00
0
2
4
6
8
10
12
Geographic distance (miles x1000)

Table 4.2

A summary of variation statistics for mtDNA and allozymes from *D. simulans*[a]

Data type	Average intra-populational diversity	Total diversity	Fixation index
mtDNA	0.026*	0.058*	0.550
allozymes	0.096	0.105	0.025

a) - Allozyme data from Choudhary and Singh (1987a) and Choudhary (1988).

* = x 10^{-2}

of single locus heterozygosity values in D. simulans which, like D. melanogaster, are widely distributed over a range of values up to 0.65. This stands in sharp contrast to mtDNA, where over the same continental populations the mtDNA variability is much smaller than that seen in D. melanogaster.

A sharp difference between the two species is the fixation indices. For climatically comparable populations of the two species, the fixation index is significantly lower ($p < 0.05$) in D. simulans. For individual loci, the distribution of F_{st} values is more sharply skewed than for D. melanogaster, with only a handful of loci having F_{st} over 0.2. If the Seychelles population, which is allozymically distinct from the continental populations, is not considered then no locus has F_{st} exceeding 0.28.

As indicated at the outset, the lower geographic differentiation of D. simulans could be explained due to a form of selection or to purely historical effects. Nuclear genes are less susceptible to the effects of random genetic drift than mtDNA because of the larger effective copy number. If the distribution of mtDNA is due primarily to recent expansion, then the nuclear variability found in the ancestral population will be similarly distributed. A weaker effect of random genetic drift will mean that there would be less post-colonization population structure. Therefore selection, in the form of a narrow niche-width or a generalist genotype, need not be invoked in the case of D. simulans. It is more parsimonious to simply hypothesize that this species is

further from ecological equilibrium than D. melanogaster.

4.3.4.2 Levels of diversity

These two species show very contrasting levels of total heterozygosity at three levels: mtDNA, allozymes, and nuclear genes (i.e. DNA sequence). The sharp difference in mtDNA diversity vs. allozymes has been described in detail. At the nuclear gene level, Aquadro et al (1988) report that in a sample of of each species taken from a single North American locality, the Xdh locus shows 4-6 times more polymorphism in D. simulans than in D. melanogaster. This latter result may not apply in other genes or in other populations (C.F. Aquadro, pers. comm.). Nevertheless, there is at least the prospect that, relative to D. melanogaster, cosmopolitan populations of D. simulans will have less mtDNA diversity, approximately equal allozyme diversity, and greater nuclear gene diversity.

Finding a hypothesis that will reconcile these observations must necessarily start with the pre-expansion state of ancestral populations of each species. What are the implications of a sub-divided ancestral species structure on polymorphism at all levels? As outlined earlier, a consequence of the coalescent model of gene evolution is that the total neutral variability of a gene will increase as long as the total species remains subdivided into isolated populations, even though it will not increase in individual populations. Therefore, a greater degree of total variability in D.

simulans would be expected at all levels. D. melanogaster by contrast would have been maintained as a single panmictic ancestral population after it split from D. simulans, in which both nuclear and mitochondrial variation would have been subject to stochastic loss of gene lineages, and the accompanying loss (or at least much slower increase, relative to D. simulans) of variability with time. In the case of nuclear genic and mtDNA variation, this expectation is met. But there are several points of caution that must be addressed.

The first point concerns why perhaps all of the nuclear variability has been included in the world-wide colonization of D. simulans, while only a single mtDNA type has. The virtual complete allopatry of the major mtDNA types argues for virtually no gene flow between the ancestral populations. However, as previously mentioned, mtDNA is less sensitive to inter-populational homogenization of variation by gene flow than are nuclear genes because of the lower effective copy number. Therefore, it is conceivable that a reasonably low level of renewed migration between diverged ancestral populations would stably introduce diverged nuclear variants from one population into the next, without resulting in mixing of mtDNA. Therefore, the ancestral population suggested to be responsible for the world-wide colonization, the sII population, would be able to incorporate the diverged nuclear variants from the sI and sIII ancestral populations, but not their mtDNA variants. The colonization stock,

therefore, would consist of much of the total variation of D. simulans, but only one of the mtDNA types.

Alternatively, it could be that the level of allozyme and nuclear DNA variability seen in continental populations of D. simulans is wholly representative of that which naturally arose in the sII ancestral stock. The lack of mtDNA variability (i.e. all four haplotypes coalesce over a very short distance) indicates that the effective population size of the sII ancestral stock is very small, and therefore the levels of allozyme and nuclear DNA variability would be expected to be low as well. However, as indicated earlier, the lower effective copy number of mtDNA means that levels of intra-populational mtDNA variability will be more susceptible to population bottlenecks than nuclear genes. It is worth recalling that Western Hemisphere populations of D. melanogaster are not significantly less heterozygous for allozymes, despite the strong founder effect on the mtDNA.

One other means by which mtDNA could be reduced in the ancestral population is by the prior action of a microorganism-based mating incompatability system. The results with the R-W system in D. simulans (Table 3.12) suggests that such a system could reduce mtDNA to monomorphism with little or no effect on nuclear genetic variation or population structure.

A second point of caution is whether the results of Aquadro et al (1988) reporting higher genic variation in D. simulans are generally true or represent an isolated case. If

it is not a general result, then why is the total mtDNA variation higher in D. simulans over D. melanogaster, while the two species are approximately equally heterozygous at both the allozyme and nuclear genic levels? One possibility is that low effective population size in the the ancestral populations of D. simulans may have led to a higher apparent rate of base substitution in mtDNA. There was some evidence for this based on the proportion of shared fragments between mtDNA types (Table 3.13), although the possibly more accurate data of Solignac et al (1986b) did not support this claim.

A final point of caution is the approximately equal levels of allozyme heterozygosity in these two species should the result of Aquadro et al (1988) turn out to be generally true. There are two possible explanations, both invoking purifying selection. Aquadro et al (1988) suggested that D. simulans has a higher N_e than D. melanogaster. More D. simulans would then be subject to purifying selection as Ohta (1976) has shown that the level of neutral heterozygosity will be proportional to $N_e s$ (effective population size times selection coefficient). An increase in N_e will mean a lowering of the critical value of s, and therfore fewer alleles will be effectively neutral. They suggest that this increased degree of purifying selection will keep the allozymic heterozygosity from being higher than in D. melanogaster, but allow neutral (synonymous) nuclear genic polymorphism to remain at higher levels in D. simulans. However, the paucity of mtDNA variability in continental populations seems to

indicate that N_e for continental D. simulans should be lower than D. melanogaster, not higher. Although allopatry and remixing of ancestal populations would produce an N_e for nuclear genes somewhat higher than N of the colonizing stock, the required increase seems to be just too high.

Another possibility is that the greater abundance of segregating chromosomal inversions, insertion/deletion sequences, and transposable elements in D. melanogaster may be influencing the degree of nucleotide polymorphism in that species. Purifying selection against chromosomes affected by insertion/deletion sequences and TE's, and directional selection acting on the linked gene combinations of chromosomal inversions would all act to decrease the amount of nucleotide polymorphism. As such elements and inversions are much rarer in D. simulans, such processes would not act on nucleotide polymorphism in that species.

SUMMARY AND CONCLUSIONS

The purpose of the present study was to analyze the utility of mtDNA restriction polymorphism to the study of species history, and to use this information to reach some conclusions about the probable role of natural selection in the maintenance of gene-enzyme variation in natural populations.

In D. melanogaster, the mtDNA data were able to give a very good approximation of a history inferred from the combined data of a great deal of different genetically determined traits. The species range could be divided into three longitudinal regions. European and African populations were inferred to be ancestral, with Far East populations being somewhat younger, and those of the Western Hemisphere very recently established, likely from a single founder introduction. The population structure of continental D. simulans is quite different in that there were few variants, of which none were unique to a single population. This indicates that continental populations of D. simulans are the likely result of a recent world-wide expansion of that species from a single ancestral group.

These results were used to draw conclusions about the nature of population genetic structure of nuclear gene-enzymes (allozymes) in these two species. In D. melanogaster, parallel latitudinal clines observed for several allozyme loci were determined to be a result of selection, rather than species

history, since temperate populations of each longitudinal region had very different mtDNA complements, and similarly so with tropical populations. The inference that continental populations of D. simulans are a result of a recent expansion led to the conclusion that the lower degree of geographical structure of allozymes, relative to D. melanogaster, is a result of D. simulans being further away from an ecological equilibrium, rather than a result of selection for a narrower niche-width or a general purpose genotype.

Finally, questions about the mechanisms underlying the sharply contrasting levels of mtDNA, allozymes, and nuclear gene DNA sequence variation cannot yet be fully answered. A good deal more work on comparative levels of variation, particularly on nuclear DNA variation, in the two species are needed.

The utility of mtDNA in analyses of populations is based in technical ease of isolation, non-recombinogenicity, and in the concept of a 'neutral marker'. All of these have been advantages over nuclear gene sequence analysis, which recombine and are under varying degrees and types of natural selection. However, all of the barriers to nuclear gene analysis are quickly disappearing. Statistical methods are being developed that can identify neutral nuclear markers (Hudson, Kreitman and Aguade 1987), and can derive genealogies in the face of intra-genic recombination by coalescent models (Hudson and Kaplan 1988). Also, the advent of the polymerase chain reaction (PCR) allows sequencing of genes isolated from

populations without the need to clone the gene of each individual sampled. PCR utilizes a set of primers to specifically amplify the gene sequence of an organism in such quantities that sequencing can be carried out. Sequence data in sample sizes that make population inferences statistically meaningful are within reach.

Nevertheless, maternal inheritance of mtDNA ensures a continued niche in population and evolutionary genetic research. This feature allows analysis of differential sex ratios of migrants, and direction of hybridization between species in hybrid zones. In addition, its inheritance means that it is somewhat isolated from the nuclear genome and will be relatively unaffected by the hitchhiking effects of linkage to favorable nuclear mutations. In this sense, it will remain a valuable marker with which to compare nuclear genetic variation in populations.

REFERENCES

Aguade, M. (1988) Restriction map variation of the Adh locus of Drosophila melanogaster in inverted and noninverted chromosomes. Genetics 119: 135-140.

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

Aloni, Y. and G. Attardi (1971) Expression of mitochondrial genome in Hela cells. I. Evidence for complete transcription of mitochondrial DNA. J. Mol. Biol. 55: 251-270.

Anderson, P.R. and J.G. Oakeshott (1981) Geographic clines and climatic associations of Adh and a-Gpdh gene frequencies in Drosophila melanogaster. pp. 237-250, In: Genetic Studies of Drosophila Populations. (ed. by J.B. Gibson and J.G. Oakeshott), Australian National University, Canberra.

Antony, C. and J.-M. Jallon (1982) The chemical basis for sex recognition in Drosophila melanogaster. J. Insect. Physiol. 28: 873-880.

Aquadro, C.F., S.F. Deese, M.M. Bland, C.H. Langley and C.C. Laurie-Ahlberg (1986) Molecular popualtion genetics

of the alcohol dehydrogenase gene region of Drosophila melanogaster. Genetics 114: 1165-1190.

Aquadro, C.F. and B.D. Greenberg (1983) Human mitochondrial DNA variation and evolution: analysis of nucleotide sequences from seven individuals. Genetics 103: 287-312.

Aquadro, C.F., K.M. Lado and W.A. Noon (1988) The rosy region of Drosophila melanogaster and D. simulans. I. Contrasting levels of naturally occurring DNA restriction map variation and divergence. Genetics 119: 875-888.

Arnold, J., M.A. Asmussen and J.C. Avise (1988) An epistatic mating system model can produce permanent cytonuclear disequilibria in a hybrid zone. Proc. Nat. Acad. Sci. U. S. A. 85: 1893-1896.

Asmussen, M.A., J. Arnold and J.C. Avise (1987) Definition and properties of disequilibrium statistics for associations between nuclear and cytoplasmic genotypes. Genetics 115: 755-768.

Attardi, G. (1985) Animal mitochondrial DNA: An extreme example of genetic economy. Int. Rev. Cytol. 93: 93-145.

Aubert, J. and M. Solignac (1988) Experimental study of cytoplasmic genes introgression in Drosophila. Genome 30: s322. (Abstract)

Avise, J.C., J. Arnold, R.M. Ball, E. Bermingham, T. Lamb, J.E. Neigel, C.A. Reeb and N.C. Saunders (1987) Intraspecific phylogeography: The mtDNA bridge between population genetics and systematics. Ann. Rev. Ecol. Syst. 18: 489-522.

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

Avise, J.C., J.E. Neigel and J. Arnold (1984) Demographic influences on mitochondrial DNA lineage survivorship in animal populations. J. Mol. Evol. 20: 99-105.

Ayala, F.J. and J.R. Powell (1972) Allozymes as diagnostic characters of sibling species of Drosophila. Proc. Nat. Acad. Sci. U. S. A. 69: 1094-1096.

Baba-Aissa, F. and M. Solignac (1984) La plupart des populations de Drosophila simulans ont probablement pour ancetre une femelle unique dans un passe recent. C. R. Acad. Sci. Paris 299 (III): 289-292.

Baba-Aissa, F., M. Solignac, N. Dennebouy and J.R. David (1989) Mitochondrial DNA variability in Drosophila simulans; quasi-absence of polymorphism within each of the three cytoplasmic races. Heredity 61: 419-426.

Baker, H.G. (1965) Characteristics and modes of origin of weeds. pp. 147-172, In: The Genetics of Colonizing Species. (ed. by H.G. Baker and G.L. Stebbins), Academic Press, New York.

Berger, E.M. (1970) A comparison of gene-enzyme variation between Drosophila melanogaster and D. simulans. Genetics 66: 677-683.

Bermingham, E. and J.C. Avise (1986) Molecular zoogeography of freshwater fishes in the southeastern United States. Genetics 113: 939-965.

Bermingham, E., T. Lamb and J.C. Avise (1986) Size polymorphism and heteroplasmy in the mtDNA of lower vertebrates. J. Hered. 77: 249-252.

Birky Jr., C.W., P. Fuerst and T. Maruyama (1989) Organelle gene diversity under migration, mutation, and drift: equilbrium expectations, approach to equilibrium, effects of heteroplasmic cells, and comparison to nuclear genes. Genetics 121: 613-627.

Birky Jr., C.W., T. Maruyama and P. Fuerst (1983) An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. Genetics 103: 513-527.

Bock, I.R. (1980) Current status of the Drosophila melanogaster species-group (Diptera). Syst. Entomol. 5: 341-356.

Bogenhagen, D. and D.A. Clayton (1977) Mouse L cell mitochondrial DNA molecules are selected randomly for replication throughout the cell cycle. Cell 11: 719-727.

Borst, P. (1972) Mitochondrial nucleic acids. Ann. Rev. Biochem. 41: 333-376.

Bossi, L. and D.M. Smith (1984) Conformational change in the DNA associated with an unusual promoter mutation in a transfer-RNA operon of Salmonella. Cell 39: 643-652.

Brookfield, J.F.Y., E. Montgomery and C.H. Langley (1984) Apparent absence of transposable elements related to the p elements of Drosophila melanogaster in other species of Drosophila. Nature 310: 330-332.

Brown, G.G. and M.V. Simpson (1982) Novel features of animal mtDNA evolution as shown by sequences of two rat cytochrome oxidase subunit II genes. Proc. Nat. Acad. Sci. U. S. A. 79: 3246-3250.

Brown, W.M. (1983) Evolution of Animal mitochondrial DNA. pp. 62-88, In: Evolution of Genes and Proteins. (ed. by M. Nei and R.K. Koehn), Sinauer, Sunderland, Mass..

Brown, W.M. (1985) The Mitochondrial Genome of Animals. pp. 95-130, In: Molecular Evolutionary Genetics. (ed. by R.J. MacIntyre), Plenum, New York.

Brown, W.M., M. George Jr. and A.C. Wilson (1979) Rapid evolution of animal mitochondrial DNA. Proc. Nat. Acad. Sci. U. S. A. 76: 1967-1971.

Brown, W.M., E.M. Prager, A. Wang and A.C. Wilson (1982) Mitochondrial DNA sequences of primates: Tempo and mode of evolution. J. Mol. Evol. 18: 225-239.

Brown, W.M. and J. Vinograd (1974) Restriction endonuclease cleavage maps of animal mitochondrial DNAs. Proc. Nat. Acad. Sci. U. S. A. 71: 4617-4621.

Cabrera, V.M., A.M. Gonzalez, J.M. Larruga and A. Gullon (1982) Electrophoretic variability in natural populations of Drosophila melanogaster and D. simulans. Genetica 59: 191-201.

Caccone, A., G.D. Amato and J.R. Powell (1988) Rates and patterns of scnDNA and mtDNA divergence within the Drosophila melanogaster subgroup. Genetics 118: 671-683.

Cann, R.M. and A.C. Wilson (1983) Length mutations in human mitochondrial DNA. Genetics 104: 699-711.

Carr, S.M., S.W. Ballinger, J.N. Derr, L.H. Blankenship and J.W. Bickham (1986) Mitochondrial DNA analysis of hybridization between sympatric white-tailed and mule deer in west Texas. Proc. Nat. Acad. Sci. U. S. A. 83: 9576-9580.

Cavener, D.R. and M.T. Clegg (1981) Temporal stability of allozyme frequencies in a natural population of Drosophila melanogaster. Genetics 98: 613-623.

Choudhary, M. (1987) Genetic structure and species divergence between popualtions of Drosophila melanogaster and its sibling species, Drosophila simulans. Ph.D. thesis, McMaster University, Hamilton, Ontario.

Choudhary, M. and R.S. Singh (1987) A comprehensive study of genic variation in natural populations of Drosophila melanogaster. III. Variations in genetic structure and their causes between D. melanogaster and its sibling species D. simulans. Genetics 117: 697-710.

Chun, E.H.L., M.H. Vaughan and A. Rich (1963) The isolation and characterization of DNA associated with chloroplast preparations. J. Mol. Biol. 7: 130-141.

Clark, A.G. (1984) Natural selection with nuclear and cytoplasmic transmission. I. A deterministic model. Genetics 107: 679-701.

Clark, A.G. (1985) Natural selection with nuclear and cytoplasmic transmission. II. Tests with Drosophila from diverse populations. Genetics 111: 97-112.

Clark, A.G. (1988) A deterministic theory of heteroplasmy. Evolution 42: 621-626.

Clark, A.G. and E.M.S. Lyckegaard (1988) Natural selection with nuclear and cytoplasmic transmission. III. Joint analysis of segregation and mtDNA in Drosophila melanogaster. Genetics 118: 471-482.

Clary, D.O., J.M. Goddard, S.C. Martin, C.M.-R. Fauron and D.R. Wolstenholme (1982) Drosophila mitochondrial DNA: a novel gene order. Nucl. Acids Res. 10: 6619-6637.

Clary, D.O. and D.R. Wolstenholme (1985) The mitochondrial DNA molecule of Drosophila yakuba. Nucleotide sequence, gene organization and genetic code. J. Mol. Evol. 22: 252-271.

Clayton, D.A. (1982) Replication of animal mitochondrial DNA. Cell 28: 693-705.

Clayton, D.A. (1984) Transcription of the mammalian mitochondrial genome. Ann. Rev. Biochem. 53: 573-594.

Coulthart, M.B. and R.S. Singh (1988a) Differing amounts of genetic polymorphism in testes and male accessory glands of Drosophila melanogaster and D. simulans. Biochem. Genet. 26: 153-164.

Coulthart, M.B. and R.S. Singh (1988b) Low genic variation in male-reproductive-tract proteins of Drosophila melanogaster and D. simulans. Mol. Biol. Evol. 5: 167-181.

Coyne, J.A. (1987) Long distance migration of Drosophila: Dispersal of D. melanogaster alleles from a Maryland orchard. Am. Nat. 130: 70-82.

Darwin, C. (1859) The Origin of Species by Means of Natural Selection. John Murray, London.

David, J.R., C. Bocquet and E. Pla (1976) New results on the genetic characteristics of the Far East race of Drosophila melanogaster. Genet. Res. 283: 253-260.

David, J.R. and P. Capy (1988) Genetic variation of Drosophila melanogaster natural populations. Trends Genet. 4: 106-111.

David, J.R. and L. Tsacas (1981) Cosmopolitan, subcosmopolitan, and widespread species: different strategies in the drosophilid family. C. R. Soc. Biogeog. 57: 11-26.

Dawid, I. (1972) Evolution of mitochondrial DNA sequences in Xenopus. Dev. Biol. 29: 139-151.

De Salle, R. and L.V. Giddings (1986a) Disocordance of nuclear and mitochondrial DNA phylogenies in Hawaiian Drosophila. Proc. Nat. Acad. Sci. U. S. A. 83: 6902-6906.

De Salle, R., L.V. Giddings and A. Templeton (1986b) Mitochondrial DNA variability in natural populations of Hawaiian Drosophila. I. Methods and levels of

variability in D. silvestris and D. heteroneura populations. Heredity 56: 75-85.

DeAlbuquerque, C.M.R. and M. Napp (1981) Genetic variation at the esterase 6 locus in natural populations of Drosophila simulans in relation to environmental heterogeneity. Genetics 98: 399-407.

deBruijn, M.H.L. (1983) Drosophila melanogaster mitochondrial DNA, a novel organization and genetic code. Nature 304: 234-241.

Densmore, L.D., J.W. Wright and W.M. Brown (1985) Length variation and heteroplasmy are frequent in mitochondrial DNA from parthenogenetic and bisexual lizards (genus Cnemidophorus). Genetics 110: 689-707.

Devereux, J., P. Haeberli and O. Smithies (1984) A comprehensive set of sequence analysis programs for the VAX. Nucl. Acids Res. 12: 387-397.

Dobzhansky, T. (1955) A review of some fundamental concepts and problems of population genetics. Cold Spring Harbor Symp. Quant. Biol. 20: 1-15.

Dobzhansky, T. (1973) Nothing in biology makes sense except in the light of evolution. Amer. Biol. Teacher 35: 125-129.

Eanes, W.F., J.W. Ajioka, J. Hey and C. Wesley (1989) Restriction-map variation associated with the G6PD

polymorphism in natural populations of Drosophila melanogaster. Mol. Biol. Evol. 6: 384-397.

Edney, E.B. (1977) Water Balance in Land Arthropods. Springer, Berlin, W. Germany.

Engels, W.J. (1981) Estimating genetic divergence and genetic variability with restriction endonucleases. Proc. Nat. Acad. Sci. U. S. A. 78: 6329-6333.

Ephrussi, B., H. Hottinguer and J. Tavlitzki (1949) Action de l'acriflavine sur les levures. Ann. Inst. Pasteur (Paris) 76: 419-450.

Ewens, W.J. (1977) Population genetic theory in relation to the neutralist-selectionist controversy. Adv. Hum. Genet. 8: 67-134.

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

Fauron, C.M.-R. and D.R. Wolstenholme (1976) Structural heterogeneity of mitochondrial DNA molecules within the genus Drosophila. Proc. Nat. Acad. Sci. U. S. A. 73: 3623-3627.

Fauron, C.M.-R. and D.R. Wolstenholme (1980) Extensive diversity among Drosophila species with respect to nucleotide sequences within the adenine + thymine rich

region of mitochondrial DNA molecules. Nucl. Acids Res. 11: 2439-2453.

Fauron, C.M.-R. and D.R. Wolstenholme (1980) Intraspecific diversity of nucleotide sequences within the adenine + thymine rich region of mitochondrial DNA molecules of Drosophila mauritiana, D. melanogaster and D. simulans. Nucl. Acids Res. 8: 5391-5410.

Felsenstein, J. (1978) Cases in which parsimony and compatability methods will be positively misleading. Syst. Zool. 27: 401-410.

Felsenstein, J. (1983) Parsimony in systematics: biological and statistical issues. Ann. Rev. Ecol. Syst. 14: 313-333.

Felsenstein, J. (1985) Confidence limits on phylogenies: An approach utilizing the bootstrap. Evolution 39: 783-791.

Ferris, S.D., R.D. Sage, E.M. Prager, U. Ritte and A.C. Wilson (1984) Mitochondrial DNA evolution in mice. Genetics 105: 681-721.

Fisher, R.A. (1930) The Genetical Theory of Natural Selection. Oxford University Press, Oxford.

Fitch, W.M. and C.H. Langley (1976) Protein evolution and the molecular clock. Fed. Proc. 35: 2092-2097.

Franklin, I. and R.C. Lewontin (1970) Is the gene the unit of selection? Genetics 65: 707-734.

Franklin, I.R. (1981) An analysis of temporal variation at isozyme loci in Drosophila melanogaster. pp. 217-236, In: Genetic Studies of Drosophila Populations. (ed. by J.B. Gibson and J.G. Oakeshott), Australian National University, Canberra.

Garesse, R. (1988) Drosophila melanogaster mitochondrial DNA: gene organization and evolutionary considerations. Genetics 118: 637-648.

George Jr., M. and O.A. Ryder (1986) Mitochondrial DNA evolution in the genus Equus. Mol. Biol. Evol. 3: 535-546.

Gillespie, J.H. (1986) Variability of evolutionary rates of DNA. Genetics 113: 1077-1091.

Gonzalez, A.M., V.M. Cabrera, J.M. Larruga and A. Gullon (1982) Genetic distance in the sibling species of Drosophila melanogaster, D. simulans, and D. mauritiana. Evolution 36: 517-522.

Gray, M.W. and W.F. Doolittle (1982) Has the endosymbiont hypothesis been proven? Micro. Rev. 46: 1-42.

Green, M. (1976) Mutable and mutator loci. In: The Genetics and Biology of Drosophila. (ed. by M. Ashburner and E. Novitski), Vol. 1(b): 929-944.

Gregorius, H.-R. and M.D. Ross (1984) Selection with gene-cytoplasm interactions. I. Maintenance of cytoplasm polymorphisms. Genetics 107: 165-178.

Gyllensten, U., D. Wharton and A.C. Wilson (1985) Maternal inheritance of mitochondrial DNA during backcrossing of two species of mice. J. Hered. 76: 321-324.

Hale, L.R. and A.T. Beckenbach (1985) Mitochondrial DNA variation in Drosophila pseudoobscura and related species in Pacific Northwest populations. Can. J. Genet. Cytol. 27: 357-364.

Hale, L.R. and R.S. Singh (1986) Extensive variation and heteroplasmy in size of mitochondrial DNA among geographic populations of Drosophila melanogaster. Proc. Nat. Acad. Sci. U. S. A. 83: 8813-8817.

Hale, L.R. and R.S. Singh (1987) Mitochondrial DNA variation and genetic structure in populations of Drosophila melanogaster. Mol. Biol. Evol. 4: 622-637.

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

Harrison, R.G., D.M. Rand and W.C. Wheeler (1987) Mitochondrial DNA size variation in Field Crickets across a narrow hybrid zone. Mol. Biol. Evol. 4: 144-158.

Hartl, D.L. and A.G. Clark (1989) Principles of Population Genetics. Sinauer, Sunderland, Mass..

Hauswirth, W.W. and P.J. Laipis (1982) Mitochondrial DNA polymorphism in a maternal lineage of Holstein cows. Proc. Nat. Acad. Sci. U. S. A. 79: 4686-4690.

Hauswirth, W.W., M.J. VandeWalle, P.J. Laipis and P.D. Olivo (1984) Heterogeneous mitochondrial DNA D-loop sequences in bovine tissue. Cell 37: 1001-1007.

Hoffman, A.A. and M. Turelli (1988) Unidirectional incompatability in Drosophila simulans: Inheritance, geographic variation and fitness effects. Genetics 119: 435-444.

Hoffman, A.A., M. Turelli and G.M. Simmons (1986) Unidirectional incompatability between populations of Drosophila simulans. Evolution 40: 692-701.

Hubby, J.L. and R.C. Lewontin (1966) A molecular approach 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. (1982) Estimating genetic varibility with restriction endonucleases. Genetics 100: 711-719.

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. (1942) Evolution: The Modern Synthesis. George Allen and Unwin, London.

Jordanova, L.J. (1984) Lamarck. Oxford University Press, Oxford.

Jukes, T.H. and C.R. Cantor (1969) Evolution of protein molecules. pp. 21-132, In: Mammalian Protein Metabolism. (ed. by H.N. Munro), Academic Press, New York.

Kaplan, N. (1983) Statistical analysis of restriction enzyme map data and nucleotide sequence data. pp. 75-106, In: Statistical Analysis of DNA Sequence Data. (ed. by B.S. Weir), Marcel Dekker, New York.

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

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

King, J.L. and T.H. Jukes (1969) Non-Darwinian evolution. Science 164: 788-791.

Kluge, A.G. and J.S. Farris (1969) Quantitative phyletics and the evolution of anurans. Syst. Zool. 18:1-32.

Klukas, C.K. and I. Dawid (1976) Characterization and mapping of mitochondrial ribosomal RNA and mitochondrial DNA in Drosophila melanogaster. Cell 9: 615-625.

Koehn, R.K., A.J. Zera and J.G. Hall (1983) Enzyme polymorphism and natural selection. pp. 115-136, In:

Evolution of Genes and Proteins. (ed. by M. Nei and R.K. Koehn), Sinauer, Sunderland, Mass..

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

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

Lachaise, D., M.-L. Cariou, J.R. David, F. Lemeunier, L. Tsacas and M. Ashburner (1987) Historical biogeography of the Drosophila melanogaster species subgroup. Evol. Biol. 22:159-255.

Lachaise, D., J.R. David, F. Lemeunier and L. Tsacas (1986) The reproductive relationships of Drosophila sechellia with D. mauritiana, D. simulans, and D. melanogaster from the Afrotropical region. Evolution 40: 262-271.

Lagerkvist, U. (1978) Two out of three; Alternative model for codon reading. Proc. Nat. Acad. Sci. U. S. A. 75: 1759-1762.

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

Lansman, R.A., J.C. Avise, C.F. Aquadro, J.F. Shapira and S.W. Daniel (1983a) Extensive genetic variation in

mitochondrial DNA's among geographic populations of the deer mouse, Peromyscus maniculatus. Evolution 37:1-16.

Lansman, R.A., J.C. Avise and M.D. Huettel (1983b) Critical experimental test of the possibility of "paternal leakage" of mitochondrial DNA. Proc. Nat. Acad. Sci. U. S. A. 80: 1969-1971.

Lansman, R.A., R.O. Shade, J.F. Shapira and J.C. Avise (1981) The use of restriction endonucleases to measure mitochondrial sequence relatedness in natural populations. III. Techniques and potential applications. J. Mol. Evol. 17: 214-226.

Latorre, A., E. Barrio, A. Moya and F.J. Ayala (1988) Mitochondrial DNA evolution in the Drosophila obscura group. Mol. Biol. Evol. 5: 717-728.

Latorre, A., A. Moya and F.J. Ayala (1986) Evolution of mitochondrial DNA in Drosophila subobscura. Proc. Nat. Acad. Sci. U. S. A. 83: 8649-8653.

Lemeunier, F. and M. Ashburner (1984) Relationships within the melanogaster species subgroup of the genus Drosophila (Sophophora). IV. The chromosomes of two new species. Chromosoma 89: 343-351.

Lemeunier, F., J.R. David, L. Tsacas and M. Ashburner (1986) The melanogaster species group. pp. 147-256, In: The Genetics and Biology of Drosophila. (ed. by M.

Ashburner, H.J. Carson and J.N. Thompson Jr.), Academic Press, New York.

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

Lewontin, R.C. and J. Krakauer (1973) Distribution of gene frequency as a test of the theory of the selective neutrality of polymorphisms. Genetics 74: 175-195.

Lewontin, R.C. and J. Krakauer (1975) Testing the heterogeneity of F values. Genetics 80: 397-398.

Lewontin, R.C., J.A. Moore, W.B. Provine and B. Wallace (1981) Dobzhansky's Genetics of Natural Populations, I-XLIII. Columbia University Press, New York.

Li, W.-H. (1979) Maintenance of genetic variability under the pressure of neutral and deleterious mutations in a finite population. Genetics 92: 647-667.

MacRae, A.F. and W.W. Anderson (1988) Evidence for non-neutrality of mitochondrial DNA haplotypes in Drosophila pseudoobscura. Genetics 120: 485-494.

Maniatis, T., E.F. Fritsch and J. Sambrook (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbour Laboratory, Cold Spring Harbour, N.Y..

Mayr, E. (1982) The Growth of Biological Thought: Diversity, Evolution, and Inheritance. Belknap, Harvard University Press, Cambridge, Mass..

Mayr, E. and W.B. Provine (1980) The Evolutionary Synthesis: Perspectives on the Unification of Biology. Harvard University Press, Cambridge, Mass..

McKenzie, J.A. and P.A. Parsons (1974) Numerical changes and environmental utilization in natural populations of Drosophila melanogaster and D. simulans. Aust. J. Zool. 22: 175-187.

Medoni, J. (1958) Le comportement de Drosophila melanogaster Meig. dans un appareil a choix lumineux: Etude comparative de souches sauvages de provenances geographiques diverses. C. R. Seances Soc. Biol. Ses Fil. 152: 1004-1007.

Melou, J.P. (1961) Etude du nombre d'ovarioles chez diverses souches francaises et japonaises de Drosophila melanogaster. Ann. Genet. 3: 25-28.

Merril, C.R. and M.G. Harrington (1985) The search for mitochondrial inheritance of human diseases. Trends Genet. 1: 140-144.

Merten, S.H. and M.L. Pardue (1981) Mitochondrial DNA in Drosophila: an analysis of genome organization and transcription in Drosophila melanogaster and D. virilis. J. Mol. Biol. 153: 1-21.

Moritz, C. and W.M. Brown (1986) Tandem duplication of D-loop and ribosomal RNA sequences in lizard mitochondrial DNA. Science 233: 1425-1427.

Moritz, C., T.E. Dowling and W.M. Brown (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Ann. Rev. Ecol. Syst. 18: 269-292.

Muller, H.J. (1927) Artificial transmutation of the gene. Science 66: 84-87.

Muller, H.J. (1950) Our load of mutations. Am. J. Hum. Genet. 2: 111-176.

Murphy, W., B. Attardi, C. Tu and G. Attardi (1975) Evidence for complete symmetrical transcription in vivo of mitochondrial DNA in Hela cells. J. Mol. Biol. 99: 809-814.

Nei, M. (1982) Evolution of human races at the gene level. pp. Human Genetics, part A: The Unfolding Genome. (ed. by B. Bonne-Tamir, P. Cohen and R.N. Goodman), Alan R. Liss, New York.

Nei, M. (1987) Molecular Evolutionary Genetics. Columbia University Press, New York.

Nei, M. and W.-H. Li (1979) Mathematical model for studying genetic variation in terms of restriction endonucleases. Proc. Nat. Acad. Sci. U. S. A. 76: 5269-5273.

Nei, M. and F. Tajima (1981) DNA polymorphism detectable by restriction endonucleases. Genetics 97: 145-163.

Nei, M. and F. Tajima (1983) Maximum likelihood estimation of the number of nucleotide substitutions from restriction sites data. Genetics 105: 207-217.

Nevo, E. (1984) The evolutionary significance of genic diversity: Ecological, demographic and life history correlates. pp. 13-213, In: Evolutionary Dynamics of Genetic Diversity. (ed. by G.S. Mani), Lecture notes in Biomathematics.

O'Brien, S.J. and R.J. MacIntyre (1969) An analysis of gene-enzyme variability in natural populations of Drosophila melanogaster and D. simulans. Am. Nat. 103: 97-113.

Oakeshott, J.G., G.K. Chambers, J.B. Gibson and D.A. Willcocks (1981) Latitudinal relationships of esterase 6 and phosphoglucomutase gene frequencies in Drosophila melanogaster. Heredity 47: 385-396.

Ohta, T. (1976) Role of very slightly deleterious mutations in molecular evolution and polymorphism. Theor. Pop. Biol. 10: 254-275.

Osborn, H.F. (1894) From the Greeks to Darwin. Macmillan, New York.

Palmer, J.D. (1985a) Evolution of Chloroplast and Mitochondrial DNA in Plants and Algae. pp. 131-240,

In: Molecular Evolutionary Genetics. (ed. by R.J. MacIntyre), Plenum, New York.

Palmer, J.D. (1985b) Comparative organization of chloroplast genomes. Ann. Rev. Genet. 19: 325-354.

Parsons, P.A. (1975) The comparative evolutionary biology of the sibling species, Drosophila melanogaster and D. simulans. Q. Rev. Biol. 50: 151-169.

Parsons, P.A. (1983) The Evolutionary Biology of Colonizing Species. Cambridge University Press, Cambridge.

Place, A.R. and D.A. Powers (1984) Purification and characterization of the lactate dehydrogenase (LSH-B4) allozymes of Fundulus heteroclitus. J. Biol. Chem. 259: 1299-1308.

Poulton, J. (1988) Mitochondrial DNA and genetic disease. Arch. Dis. Childhood 63: 883-885.

Powell, J.R. (1971) Genetic variation in natural populations of animals. Science 174: 1035-1036.

Powell, J.R. (1983) Interspecific cytoplasmic gene flow in the absence of nuclear gene flow: Evidence from Drosophila. Proc. Nat. Acad. Sci. U. S. A. 80: 492-495.

Powell, J.R. and H. Wistrand (1978) The effect of heterogeneous environments and a competitor on genetic variation in Drosophila. Am. Nat. 112: 935-947.

Powers, T.O., E.G. Platzer and B.C. Hyman (1986) Large mitochondrial genome and mitochondrial DNA size polymorphism in the mosquito parasite, Romanomermis culicivorax. Curr. Genet. 11: 71-77.

Prakash, S. (1972) Origin of reproductive isolation in the absence of apparent genic differentiation in a geographic isolate of Drosophila pseudoobscura. Genetics 72: 143-155.

Provine, W.B. (1971) The Origins of Theoretical Population Genetics. University of Chicago Press, Chicago.

Rand, D.M. and R.G. Harrison (1986) Mitochondrial DNA transmission genetics in crickets. Genetics 114: 955-970.

Rand, D.M. and R.G. Harrison (1989) Molecular population genetics of mtDNA size variation in crickets. Genetics 121: 551-569.

Reilley, J.G. and C.A. Thomas (1980) Length polymorphism, restriction site variation and maternal inheritance of mitochondrial DNA of Drosophila melanogaster. Plasmid 3: 109-115.

Rhomberg, L.R. and R.S. Singh (1989) Evidence for a link between local and seasonal clines in gene frequencies and latitudinal gene clines in a cyclic parthenogen. Genetica 78: 73-79.

Rochaix, J.D. (1985) Genetic organization of the chloroplast. Int. Rev. Cytol. 93: 57-91.

Sager, R. and M.R. Ishida (1963) Chloroplast DNA in Chlamydomonas. Proc. Nat. Acad. Sci. U. S. A. 50: 2725-2730.

Saunders, N.C., L.G. Kessler and J.C. Avise (1985) Genetic variation and geographic differentiation in mitochondrial DNA of the horseshoe crab, Limulus polyphemus. Genetics 112: 613-627.

Sederoff, R.R. (1984) Structural variation in mitochondrial DNA. Adv. Genet. 22: 1-108.

Shah, D.M. and C.H. Langley (1977) Complex mitochondrial DNA in Drosophila. Nucl. Acids Res. 4: 2949-2960.

Shah, D.M. and C.H. Langley (1979) Inter- and intraspecific variation in restriction maps of Drosophila mitochondrial DNAs. Nature 281: 696-699.

Singh, G., N. Neckelmann and D.C. Wallace (1987) Conformational mutations in human mitochondrial DNA. Nature 329: 270-272.

Singh, R.S., M. Choudhary and J.R. David (1987) Contrasting patterns of geographic variation in cosmopolitan sibling species Drosophila melanogaster and D. simulans. Biochem. Genet. 25: 27-40.

Singh, R.S. (1983) Genetic differentiation for allozymes and fitness characters between mainland and Bogota

populations of Drosophila pseudoobscura. Can. J. Genet. Cytol. 25: 590-604.

Singh, R.S. (1989) Population genetics and evolution of species related to Drosophila melanogaster. Ann. Rev. Genet. (in press).

Singh, R.S. and M.B. Coulthart (1982) Genic variation in abundant proteins of Drosophila melanogaster and D. pseudoobscura. Genetics 102: 437-453.

Singh, R.S. and L.R. Hale (1989) Are mitochondrial DNA variants selectively non-neutral? (submitted to Genetics).

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

Singh, R.S. and L.R. Rhomberg (1987) 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 L.R. Rhomberg (1987) 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.

Slatkin, M. (1987) Gene flow and the geographic structure of natural populations. Science 236: 787-792.

Sneath, P.H.A. and R.R. Sokal (1973) Numerical Taxonomy. Freeman, San Francisco.

Snyder, M., A.R. Buchman and R.W, Davis (1986) Bent DNA at a yeast autonomously replicating sequence. Nature 324: 87-89.

Snyder, M., A.R. Fraser, J. LaRoche, K.E. Gartner-Kepkay and E. Zouros (1987) Atypical mitochondrial DNA from the deep-sea scallop Placopecten magellanicus. Proc. Nat. Acad. Sci. U. S. A. 84: 7595-7599.

Sober, E. (1983) Parsimony in systematics: philosophical issues. Ann. Rev. Ecol. Syst. 14: 335-357.

Solignac, M., J. Genermont, M. Monnerot and J.-C. Mounolou (1984) Genetics of mitochondria in Drosophila: mtDNA inheritance in heteroplasmic strains of D. mauritiana. Mol. Gen. Genet. 197: 183-188.

Solignac, M., J. Genermont, M. Monnerot and J.-C. Mounolou (1987) Drosophila mitochondrial genetics: Evolution of heteroplasmy through germ-line cell divisions. Genetics 117: 687-696.

Solignac, M. and M. Monnerot (1986) Race formation, speciation, and introgression within Drosophila melanogaster, D. mauritiana, and D. sechellia inferred

from mitochondrial DNA analysis. Evolution 40: 531-539.

Solignac, M., M. Monnerot and J.-C. Mounolou (1986) Concerted evolution of sequence repeats in Drosophila mitochondrial DNA. J. Mol. Evol. 24: 53-60.

Solignac, M., M. Monnerot and J.-C. Mounolou (1986) Mitochondrial DNA evolution in the melanogaster species subgroup of Drosophila. J. Mol. Evol. 23: 31-40.

Spolsky, C. and T. Uzzell (1984) Natural interspecies transfer of mitochondrial DNA in amphibians. Proc. Nat. Acad. Sci. U. S. A. 81: 5802-5805.

Spolsky, C. and T. Uzzell (1986) Evolutionary history of the hybridogenetic frog Rana esculenta as deduced from mtDNA analysis. Mol. Biol. Evol. 3: 44-56.

Steiner, W.W.M. (1977) On niche-width and genetic variation in Hawaiian Drosophila. Am. Nat. 111: 1037-1045.

Teissier, G. (1957) Discriminative biometrical characters in French and Japanese Drosophila melanogaster. In: Proc. Int. Genet. Symp., 1956, Cytologia, Suppl., pp. 502-505.

Templeton, A. (1983) Phylogenetic inference from restriction endonuclease cleavage site maps with particular reference to the evolution of man and the apes. Evolution 37: 221-244.

Triantaphyllidis, C.D. (1973) Allozyme variation in populations of Drosophila melanogaster and D. simulans from northern Greece. J. Hered. 64: 69-72.

Triantaphyllidis, C.D., Z.G. Scouras, J.N. Panourgias and G.C. Ionnidis (1982) Allozyme variation in Greek wild populations of Drosophila melanogaster and D. simulans along a north-south gradient. Genetica 58: 129-136.

Vanlerberghe, F., P. Boursot, J. Catalan, S. Gerasimov, F. Bonhomme, B.A. Botev and L. Thaler (1988) Analyse genetique de la zone d'hybridization entre les deux sous-especes de souris Mus musculus domesticus et M. m. musculus en Bulgarie. Genome 30: 427-437.

Vanlerberghe, F., P. Boursot, J.T. Nielsen and F. Bonhomme (1988) A steep cline for mitochondrial DNA in Danish mice. Genet. Res. 52: 185-193.

VanValen, L. (1965) Morphological variation and the width of the ecological niche. Am. Nat. 99: 377-389.

Vawter, L. and W.M. Brown (1986) Nuclear and mitochondrial DNA comparisons reveal extreme rate variation in the molecular clock. Science 234: 194-196.

Voelker, R.A., T. Mukai and F.M. Johnson (1977) Genetic variation in populations of Drosophila melanogaster from the western United States. Genetica 47: 143-148.

Wallis, G.P. (1987) Mitochondrial DNA insertion polymorphism and germ line heteroplasmy in the Triturus cristatus complex. Heredity 58: 229-238.

Warrior, R. and J. Gall (1985) The mitochondrial DNA of Hydra attenuata and Hydra littoralis consists of two linear molecules. Arch. Sc. Geneve. 3: 439-445.

Watada, M., Y.N. Tobari and S. Ohba (1986) Genetic differentiation in Japanese populations of Drosophila simulans and D. melanogaster. I. Allozyme polymorphisms. Jpn. J. Genet. 61: 253-269.

Watson, J.D. and F.H.C. Crick (1953) A structure for deoxyribose nucleic acid. Nature 171: 737-738.

Watterson, G.A. (1977) Heterosis or neutrality. Genetics 85: 789-814.

White, M.M., S.D. Mane and R.C. Richmond (1988) Studies of esterase 6 in Drosophila melanogaster. XVIII. Biochemical differences between the slow and fast allozymes. Mol. Biol. Evol. 5: 41-62.

Wilson, A.C., S.S. Carlson and T.J. White (1977) Biochemical evolution. Ann. Rev. Biochem. 46: 573-639.

Wolstenholme, D.R. and D.O. Clary (1985) Sequence evolution of Drosophila mitochondrial DNA. Genetics 109: 725-744.

Wolstenholme, D.R., D.O. Clary, J.L. MacFarlane, J.A. Wahleithner and L. Wilcox (1985) Organization and

evolution of invertebrate mitochondrial genomes. pp. 175-182, In: Acheivments and Perspectives of Mitochondrial Research. (ed. by E. Quagliariello, E.C. Slater, F. Palmieri, C. Saccone and A.M. Kroon), Elsevier, Amsterdam.

Wolstenholme, D.R. and I. Dawid (1968) A size difference between mitochondrial DNA molecules of urodele and anuran amphibians. J. Cell Biol. 39: 222-228.

Wolstenholme, D.R., J.M. Goddard and C.M.-R. Fauron (1979) Stucture and replication of mitochondrial DNA from the genus Drosophila. pp. 409-425, In: Proc. 8th Ann. ICN/UCLA Symp. Molecular and Cellular Biology: Extrachromosomal DNA. (ed. by D. Cummings, I.B. Dawid, P. Borst, S. Weissman and F. Fox), Academic Press, New York.

Wright, S. (1931) Evolution in Mendelian populations. Genetics 16: 97-159.

Zahn, K. and R.F. Blattner (1985) Sequence induced DNA curvature at the bacteriophage-lambda origin of replication. Nature 317: 451-453.

Zuckerkandl, E. and L. Pauling (1965) Evolutionary divergence and convergence in proteins. pp. 97-166, In: Evolving Genes and Proteins. (ed. by V. Bryson and H.J. Vogel), Academic Press, New York.

APPENDIX 1

TABLE A1

Observed restriction fragments in digests of D. melanogaster mtDNA. For each enzyme, the fragments obseved (in kbp) are denoted. For enzymes revealing polymorphism, variants are denoted by an upper-case letter (as described in the text), and the presence of fragments is denoted by an x. Fragments marked by an asterisk are ones showing A+T-rich region size variation.

	EcoRI	XbaI	HindIII	HpaII
	10.7*	9.4*	8.5	9.5*
	5.3	8.1	5.1*	3.7
	1.72	1.13	4.8	3.25
	0.90	0.29	0.38	2.15
Σ	18.6	18.9	18.8	18.6

	AvaII				
	A	B	C	D	E
14.7*			x		
8.7	x	x		x	x
7.9				x	
6.2					x
5.8	x	x			
2.08	x	x	x		x
1.90			x		
1.65	x		x	x	x
0.33	x			x	
0.23	x		x	x	x
Σ	18.8	18.5	18.7	18.8	18.9

Appendix 1 (continued)

HaeIII

	A	B	C	D
8.7*	x	x	x	x
6.2		x		
5.3	x		x	
3.7	x	x		x
3.6				x
2.85			x	
1.70				x
0.90	x		x	x
0.85			x	
Σ	18.6	18.6	18.6	18.6

DdeI

	A	B	C
6.6*	x	x	x
3.7	x		
2.08	x	x	x
2.00		x	x
1.67		x	
1.30	x	x	x
1.21	x	x	x
1.15			x
1.01	x	x	x
0.89	x	x	x
0.70	x	x	x
0.62	x	x	x
0.52			x
0.50	x	x	x
Σ	18.6	18.6	18.6

Appendix 1 (continued)

MboI

	A	B	C	D	E	K
5.2*	x	x	x		x	x
4.6				x		
1.86	x	x	x	x	x	x
1.50						x
1.38	x	x	x	x	x	
1.32	x	x	x	x	x	x
1.19	x	x	x	x	x	x
1.09	x	x	x	x	x	x
1.04	x		x		x	
0.90	x	x	x	x	x	x
0.81	x	x	x	x	x	x
0.73			x			
0.72	x		x		x	
0.69	x	x		x	x	x
0.58				x		
0.51	x	x	x	x	x	x
0.41	x	x	x	x	x	x
0.41	x	x	x	x	x	x
0.38	x	x	x	x	x	x
0.35		x		x		x
0.35	x	x	x	x	x	x
0.26	x	x	x	x		x
0.245	x	x	x	x	x	x
0.240	x	x	x	x		x
0.181	x	x	x	x	x	
0.156					x	
0.138	x	x	x	x	x	x
0.097					x	
0.079	x	x	x	x	x	x
0.077	x	x	x	x	x	x
0.049	x	x		x	x	x
0.023	x	x	x	x	x	x
0.022	x	x	x	x	x	x
0.020	x	x	x	x	x	x
Σ	18.8	18.8	18.8	18.8	18.8	18.7

Appendix 1 (continued)

	TaqI		
	A	B	E
5.5*	x		x
3.9		x	
1.93	x	x	x
1.49		x	
1.17			x
1.00	x	x	x
0.91	x	x	
0.86	x	x	x
0.85	x	x	x
0.63	x	x	x
0.54	x	x	x
0.49	x	x	x
0.49	x	x	x
0.47	x	x	x
0.45	x	x	x
0.42	x	x	x
0.42	x	x	x
0.41	x	x	x
0.36	x	x	x
0.35	x	x	x
0.33	x	x	x
0.31	x	x	x
0.27	x	x	x
0.26	x	x	x
0.25	x	x	x
0.238	x	x	x
0.238	x	x	x
0.195	x	x	
0.184	x	x	x
0.113	x	x	x
0.055	x	x	x
0.055	x	x	x
0.050	x	x	x
0.041	x	x	x
0.021	x	x	x
0.021	x	x	x
Σ	18.8	18.7	18.9

Appendix 1 (continued)

	HinfI				
	A	B	C	D	F
7.0*	x	x	x	x	x
3.25	x	x	x	x	x
2.25					x
1.75	x	x			
1.70			x	x	
1.26	x	x	x	x	x
1.18	x	x	x		x
1.11	x	x	x	x	x
0.70	x				
0.60				x	
0.58	x	x	x	x	x
0.55	x	x	x	x	x
0.52				x	
0.38		x	x	x	x
0.36	x	x	x	x	x
0.35		x	x	x	x
0.33	x	x	x	x	x
0.061	x	x	x	x	x
0.061	x	x	x	x	x
0.037			x	x	
0.025	x	x	x	x	x
Σ	18.7	18.7	18.7	18.7	18.7

APPENDIX 1

TABLE A2

Observed restriction fragments in digests of D. simulans mtDNA, for both sI and sII types. The presence of fragments (in kbp) in each type is denoted by an x. In cases of within-type polymorphism, the variant patterns having the fragment are indicated by their upper-case letter designation.

EcoRI

	sI	sII
8.6		x
8.4	x	
4.1	x	x
2.48	x	
1.77	x	x
1.28		x
1.05		x
0.96	x	
0.90	x	x
0.63		x
0.43		x
Σ	18.6	18.6

XbaI

	sI	sII
9.3		x
8.3	x	x
5.9	x	
3.2	x	
0.97	x	x
0.35	x	x
Σ	18.7	18.9

AvaII

	sI	sII
8.5		x
5.8	x	x
4.7	x	
3.9	x	
2.06	x	x
1.65	x	x
0.33	x	x
0.24	x	x
Σ	18.7	18.6

HaeIII

	sI	sII
6.7	x	x
6.1	x	x
6.1	x	x
Σ	18.9	18.9

Appendix 1 (continued)

HindIII

	sI	sII
8.7	x	
8.3		x
5.1	x	x
4.8	x	x
0.42	x	x
Σ	18.6	18.6

HpaII

	sI	sII
17.0		x
15.0	x	
2.20	x	x
1.65		x
Σ	19.2	18.9

HinfI

	sI	sII
7.1		ABC
5.0	x	
4.4		C
4.3	x	
3.2		x
3.0		B
2.9		ACD
2.6	x	
2.50		C
2.23	x	
1.18	x	
1.11		x
0.98	x	x
0.89	x	
0.66		x
0.63		x
0.63		x
0.59	x	x
0.55	x	x
0.41		x
0.38		CD
0.36		AB
0.28	x	
0.20	x	
0.152	x	x
0.126		ACD
0.050	x	
0.032		x
0.022		AB
Σ	19.0	18.8

DdeI

	sI	sII
7.3	x	
7.1		x
5.4	B	
5.0	A	
3.5		x
2.6		x
1.74	x	
1.60		x
1.55	x	
1.01	x	x
0.88	x	
0.70	x	x
0.62		x
0.59		x
0.55		x
0.40	A	
0.220		x
0.078		x
Σ	18.6	18.6

Appendix 1 (continued)

	MboI	
	sI	sII
5.2	x	
5.1		x
1.57	x	x
1.32	x	x
1.19	x	x
1.19		x
1.09		x
1.04	x	x
0.90	x	x
0.80	x	x
0.78		x
0.73	x	x
0.67	x	
0.64	x	
0.58	x	x
0.47	x	
0.43		x
0.42	x	
0.39	x	x
0.37	x	
0.36		x
0.35		x
0.33	x	x
0.32		x
0.30	x	x
0.28		x
0.26	x	x
0.181	x	x
0.132	x	x
0.098	x	
0.077	x	x
0.077	x	x
Σ	18.6	18.7

	TaqI	
	sI	sII
7.0	x	
6.5		x
1.93		x
1.36	x	
1.28	x	x
1.00	x	
0.86	x	x
0.85	x	x
0.83		x
0.74	x	x
0.72		x
0.63	x	x
0.54		x
0.49	x	x
0.49		x
0.47		x
0.45	x	x
0.43		x
0.41	x	x
0.38		x
0.38	x	
0.36	x	x
0.35		x
0.35		x
0.33		x
0.29		x
0.27		x
0.26	x	x
0.238		x
0.238		x
0.225		x
0.184	x	x
0.113		x
0.105		x
0.055		x
Σ	18.9	18.8