NOTICE

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

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

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

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

AVIS

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

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

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

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

Polymorphism and Heteroplasmy of Mitochondrial DNA in the Scallop *Placopecten magellanicus* (Gmelin)

by

Kathleen M. Fuller

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia

January,1991

ISBN 0-315-64536-9

To Thos. B.

Table of Contents

List of Figures

List of Tables

Abstract

A survey of mitochondrial DNA variation in a sample of 280 individuals from six geographic populations of the scallop species *Placopecten magellanicus* revealed extensive length polymorphism. Variable regions in the molecule were mapped, and almost all size polymorphism was found to occur at three discrete loci.

The mitochondrial genome has a consensus size of 35 kilobase pairs, and varies between 32 and 42 kb. The three major length variable loci exhibit variation both within (heteroplasmy) and among (polymorphism) individuals. Locus I is a tandem array of two to eight copies of a 1.45 kb repeat; locus II and locus III have increments of change of about 250 base pairs and 50 bp, with six and more than ten size classes respectively. Small deletions of about 200 bp in three different restriction fragments between locus II and locus III were also observed. Contiguous fragments totalling over 15 kb located between locus I and locus III showed no variation in length.

Two methods were used to measure the diversity of variation. For locus I and locus II, size class diversity was estimated as 64% and 19% respectively. For locus III genotypic rather than allelic diversity was estimated, yielding a conservative value of 91%, and this statistic also gave a restriction morph diversity of 37%. The frequency of heteroplasmy ranged from 3.4% for locus II to 69% for locus III. A number of individuals exhibited simultaneous heteroplasmy for more than one marker. Four cases of restriction site heteroplasmy were observed, and three of them were associated with locus I heteroplasmy.

For each of the major length polymorphic loci significant differences in the distribution of variation were found between populations, but there was no consistent pattern of genetic differentiation. It was concluded that the high backward and forward mutation rate among a limited number of size classes that characterizes mtDNA length polymorphism precludes the use of this kind of variation as a population marker.

List of Abbreviations

Aat	aspartate aminotransferase
ANOVA	analysis of variance
bp	nucleotide base pairs
BP	before present
BSA	bovine serum albumin
cpm	counts per minute
CO I,II,III	cytochrome c oxidase subunits
dATP	deoxyadenosine-5'-triphosphate
dCTP	deoxycytidine-5'-triphosphate
dGTP	deoxyguanosine-5'-triphosphate
dH_2O	distilled water
DTT	dithiothreitol
EDTA	ethylenediaminetetraacetic acid
HEPES	N-2-hydroxyethylpiperazine-N'-2-ethanesulfonic acid
kb	kilobase pairs (1000 base pairs)
LrRNA	large subunit of mitoribosomal RNA
mA	milliampere
Mpi	mannosephosphate isomerase
mRNA	messenger RNA
mtDNA	mitochondrial DNA
mtRNA	mitochondrial RNA
My	million years
ND	mitochondrial NADH dehydrogenase subunits
NS	statistically not significant ($p>0.05$)

Odh	octopine dehydrogenase
PEG	polyethylene glycol 8000
Pgd	6-phosphogluconate dehydrogenase
Pgi	phosphoglucose isomerase
Pgm	phosphoglucose mutase
rRNA	ribosomal RNA
scnDNA	single copy nuclear DNA
ssDNA	single stranded DNA
tRNA	transfer RNA
TTP	thymidine-5'-triphosphate
U	units
V	volts

Acknowledgements

I thank Eleutherios Zouros for supervising my research, and for the intellectual stimulation and insights he provided during the course of my work. The members of my supervisory committee, Jonathan Wright (Biology), Michael Gray (Biochemistry), and Michael Dadswell (Acadia University) each contributed his expertise and advice, and helped me to maintain a focus in bringing together molecular and evolutionary approaches. Jean Worms (DFO, Moncton), Sam Naidu (DFO, St. John's), National Sea Products, Ltd., Michael Dadswell, and the Biology Diving Club generously provided animals for research. Marty Snyder guided my first efforts in the laboratory, and together with Julie LaRoche was an unfailing source of support and encouragement throughout this study. Kami Ahmad ran the allozyme gels, and with Filip Volckaert and Branimir Gjetvaj contributed ideas and a sense of fun and discovery to the lab. Special thanks go to Doug Cook for his steadfast support, his technical skill and advice, and his friendship.

I am deeply grateful to my husband Tom for the sacrifices he made to enable me to pursue this course, and for the love he showed in sharing the pain and joy. The understanding and pride of my sons Giles and Toby helped me immeasurably. I owe much to all of my family.

An NSERC postgraduate scholarship and a Walter C. Sumner Fellowship provided financial support, and an NSERC grant to E. Zouros covered research expenses.

Chapter 1

General Introduction

1.1 Introduction

The mitochondrial genome of animals has captured the interest of biologists in a number of disciplines during the past decade. The simple genetic organization provides an ideal system for the study of gene expression and the coordinate regulation of organelle and nuclear genomes. For systematic biologists, the genome provides a record of molecular evolution that is generally believed to be free of the effects of biparental inheritance and recombination. Rates of sequence divergence within and among regions of the genome can be compared, both within and among taxa. Behaving as a clonal lineage, mitochondrial DNA (mtDNA) evolves as a function of the population dynamics within organelles, cells, individuals, and populations of individuals.

Two of the questions that have intrigued evolutionary biologists concern characteristics of animal mtDNA that contrast with those of other eukaryotic kingdoms. First, what are the constraints that led to and maintain this most economical genome, with a size that barely accommodates the mitochondrial genes and single control region? Second, what evolutionary interplay of forces can account for the empirical observation of predominantly intra-individual homogeneity, or homoplasmy, and inter-individual heterogeneity, or polymorphism? Both of these questions may be addressed in systems that depart from the metazoan norm with regard to genome size and intra-individual variation.

The mitochondrial genome of the scallop *Placopecten magellanicus* is one such system. The mtDNA molecule is much larger than required to

encode the standard set of mitochondrial genes. Further, the molecule is highly variable within and among individuals. I have undertaken to describe and where possible to quantify this variation at several hierarchical levels in order to gain some understanding of the evolutionary dynamics of the mitochondrial genome in this species. In Chapter 3, I present the basic molecular characterization of the genome obtained by restriction mapping and some heterologous hybridization. In Chapter 4, I describe the kinds and frequencies of polymorphism found in a survey of 280 scallops, and in Chapter 5 I extend this to include patterns of variation within individuals for several mitochondrial markers and for different tissues of the animal. I address the question of population subdivision by examining the distribution of polymorphic characters among the six geographic populations comprising the survey with special reference to mitochondrial length variation, and describe correlations among and between mitochondrial and other characters in Chapter 6.

1.2 Structure and evolution of animal mtDNA

The size, structure, gene content and organization of mtDNA varies within and among the four eukaryotic kingdoms, but in most species the genome contains a basic set of genes encoding large and small mitoribosomal RNA subunits, the tRNAs needed for translation of mitochondrial mRNA, and a number of polypeptides involved in the electron transport chain and oxidative phosphorylation (Gray 1989). The diversity of structure is greatest among unicellular organisms, where the genome may exist as a single linear molecule as in *Paramecium* or as the complex kinetoplast mini- and maxicircles of *Trypanosoma* (Sederoff 1984). Fungal mitochondrial

genomes vary almost 10-fold in size, and are distinctive by virtue of possessing an "extra" set of genes, encoded within introns of other genes, that function in transcript processing (Kotylak *et al.* 1985). The plant mitochondrial genome is typically very large and complex, often multipartite and comprising lengthy spacer regions and introns in addition to the usual mitochondrial genes (Sederoff 1984). In contrast, animal mtDNA is distinguished by its simplicity; the genome is usually the minimal size needed to encode the two rRNAs, 22 tRNAs, and 13 mRNAs which are virtually universal throughout the kingdom (Attardi 1985). With the single exception of *Hydra* (Warrior and Gall 1985), where the genome is in the form of two linear molecules, all characterized animal mitochondria contain multiple copies of a covalently closed circular molecule that is typically about 16 kilobases (kb) in size.

Complete nucleotide sequences have been obtained for human (Anderson *et al.* 1981), mouse (Bibb *et al.* 1981), frog (Roe *et al.* 1985), bovine (Anderson *et al.* 1982), *Drosophila* (Clary and Wolstenholme 1985), sea urchin (Jacobs *et al.* 1988), and nematode (Wolstenholme *et al.* 1987, Garey and Wolstenholme 1989) mtDNA, representing four phyla. These data have provided information on the gene content, structure, and organization in metazoan mtDNA. All characterized animal mtDNAs encode the same 37 transcripts with the exception of *Ascaris suum,* which lacks the ATPase8 gene; sequence data from another nematode, *Caenorhabditis elegans* , indicate that loss of ATPase8 may be a general feature in this group (Wolstenholme *et al.* 1987). Each mtDNA also includes a non-transcribed region, ranging from 121 base pairs (bp) in *Strongylocentrotus purpuratus* (Jacobs *et al.* 1988) to 3200 bp in *Xenopus laevis* (Roe *et al.* 1985), that

contains sequences that initiate replication and transcription. No intervening sequences have been found in any animal taxon, and intergenic sequences are absent or reduced to a few nucleotides; in many cases stop codons are added by post-transcriptional modification, and several overlapping genes have been described (Brown 1985).

These factors combined produce a very streamlined genome. With tightly packed genes and a single non-transcribed region, the molecule rarely exceeds 18 kb in size. Exceptions that have been documented occur in species where duplications or repeated sequences result in an enlarged genome. As more species have been characterized, more examples of "abnormal" genome size have been reported, but no correlation between size and taxonomic group has become apparent. However, alterations of gene arrangement do appear to be characteristic of higher taxa. Marked differences exist among vertebrate, insect, echinoderm, and nematode functional maps (Moritz *et al.* 1987)

In the vertebrate genome most templates are on the heavy strand, with one or a few tRNA genes separating other coding regions (Chang *et al.* 1985). This organization led Attardi (1985) to propose his punctuation model, suggesting that secondary structures assumed by tRNAs in the nascent polycistronic transcript serve as processing signals. In insects, the templates are more evenly distributed between strands and tRNAs are clustered (Clary and Wolstenholme 1985). Similar, but more extreme, clustering is found in the sea urchin, and it has been suggested that tRNA genes may be recombinogenic substrates (Moritz and Brown 1986, Cantatore *et al.* 1987, Jacobs *et al.* 1988), producing the rearrangements of blocks of genes seen in comparisons of vertebrate, echinoderm, and insect mtDNA. In all genomes

where functional regions have been mapped, the 12S rRNA gene lies adjacent to the control region, except in *S. purpuratus* where the putative control region is separated from the 12S rRNA gene by two tRNAs (Jacobs *et al.* 1988).

All evidence to date indicates that both light and heavy strand promoters lie within the control region (Bogenhagen *et al.* 1985). In vertebrates there is evidence for transcriptional attenuation, with higher steady state levels of the rRNAs adjacent to the promoter (Dubin *et al.* 1985), but a similar mechanism cannot be invoked for the sea urchin where the large and small subunits are separated and neither is directly adjacent to the control region. Although the possibility of separate promoters to produce partial and complete transcripts has not yet been excluded, Jacobs and co-workers (1988) have suggested that regulation of steady state levels is most likely post-transcriptional in *S. purpuratus*.

Replication is also initiated in the control region. In vertebrates and echinoderms, which have a displacement, or D-loop, structure (Brown 1983), heavy strand synthesis begins in the D-loop and proceeds about 67% of the way around the molecule before the light strand origin is uncovered (Clayton 1982). In insects initiation takes place in the A+T rich region, and heavy strand replication is almost complete before light strand synthesis begins (Clary and Wolstenholme 1985). Asynchronous replication requires that long segments of the genome be exposed as single-stranded DNA through much of the process. In mammals, Clayton (1982) has found that non-replicating molecules are generally triple stranded in the D-loop region and that turnover of the third D-loop strand is high, with as many as 95% of D-loop strands being lost. Whether the D-loop strand serves as a primer or is

displaced at replication initiation remains unclear, but the region has a much higher evolutionary rate than the rest of the molecule. Apart from several short conserved sequence blocks, the D-loop shows little sequence similarity even among congeners (Brown 1983).

The tempo and mode of mtDNA evolution differ at higher taxonomic levels. Length polymorphism shows very different patterns in comparisons between homeotherms and poikilotherms. Most variation in mammals involves deletions, generally with phenotypic effect (Holt *et al.* 1988a,1988b,1989, Schon *et al.* 1989, Wallace 1989), or different numbers of nucleotides in short homopolymer runs (Brown and DesRosiers 1983, Cann and Wilson 1983, Olivo *et al.* 1983). In contrast, the lower vertebrates and invertebrates are rich in mtDNA size variation. Tandem duplications and tandem arrays have been described in a wide range of taxa, from nematodes (Powers *et al.* 1986) to lizards (Densmore *et al.* 1985), with conspecifics differing by up to 9 kb. Both of these kinds of length mutation have been shown to occur at rates several orders of magnitude greater than nucleotide substitution (Solignac *et al.* 1984, Moritz and Brown 1987, Rand and Harrison 1989). Further, it now appears that rates of nucleotide substitution in mtDNA may vary widely between warm-blooded and cold-blooded animals, with the more metabolically active homeotherms showing rates 5-10 times higher (Miyata *et al.* 1982, Moritz *et al.* 1987). Substitution rates also vary among regions of the molecule, with the rRNA and tRNA genes being the most highly conserved. Among protein coding genes, conservation at the nucleotide level is highest in subunit 3 of the cytochrome oxidase (COIII) gene. Brown (1983) has suggested that sites that can tolerate

mutations become saturated after about 10 My, obscuring the linear relationship between sequence divergence and time.

The proportion of polymorphic restriction sites and the nucleotide diversity found within species varies widely. Under a neutral model the amount of sequence divergence is a function of effective population size and mutation rate. With good evidence for a high mutation rate in vertebrate mtDNA, the low diversity in human mtDNA is attributed to a relatively recent bottleneck (Cann *et al.* 1987). Similarly, Densmore *et al.* (1985) consider the extremely low diversity in *Cnemidophorus* mtDNA to be evidence of a founder event perhaps as recent as 10,000 years ago. Avise *et al.* (1988) suggest that the enormous diversity in menhaden fishes may be explained by their huge population sizes and in chuckwalla lizards by extensive population subdivision. The evolutionary rates of mtDNA and scnDNA appear to be similar in some invertebrate taxa however (Powell *et al.* 1986, Caccone *et al.* 1988), and although Vawter and Brown (1986) concluded that this difference in relative rates was mainly due to an elevated scnDNA mutation rate, Latorre *et al.* (1986) presented convincing evidence for significant heterogeneity in mtDNA evolutionary rates among *Drosophila* lineages.

Whether polymorphism is present for sequence or for size in a population it must first have arisen by mutation within a single germline, but reports of intra-individual polymorphism, or heteroplasmy, remain relatively rare. By far the greatest number of such cases involve size variation rather than restriction site or sequence variation. Only four examples of site heteroplasmy have been reported, in the cow (Hauswirth *et al.* 1984), *D. melanogaster* (Hale and Singh 1986), the American shad

(Bentzen *et al.* 1988), and the Icelandic scallop (Gjetvaj 1989), while almost every system displaying length variation includes some individuals that are heteroplasmic for size (Moritz *et al.* 1987).

Several techniques have been developed to study the evolution of populations using mtDNA sequence data. In humans, where diversity is low, sequencing of selected regions or high resolution restriction mapping guided by sequence have elucidated the genetic relationships of geographic populations (Cann *et al.* 1987). Estimates of sequence divergence based on the proportion of shared restriction sites have been used to describe population subdivision and phylogenetic relationships in a number of animal species (Avise *et al.* 1987). Resolution can be taken almost to the level of the individual by "fingerprinting" of hypervariable sites (Avise *et al.* 1989), although the observation that certain sites seem to mutate relatively frequently in independent lineages means that convergence may obscure some relationships.

Only recently has length variation been studied in the context of population diversity. In a survey of *D. melanogaster* , Hale and Singh (1986) found a total of 13 mtDNA size classes, with no geographic population having more than 6 classes. A similar survey of spawning populations of the American shad revealed two size classes, the larger of which was only observed in heteroplasmy with the smaller and showed a significant difference in distribution between northern and southern populations (Bentzen *et al.* 1988). In two species of cricket, Rand and Harrison (1989) found that the high level of diversity for mtDNA size classes was mainly attributable to variation among individuals rather than among populations or species.

1.3 *Placopecten magellanicus*

P. magellanicus is a member of the family Pectinidae, order Pterioidia, class Bivalvia. Bivalves first appeared in the early Cambrian (550 Mya), and the first members of the Pectinidae arose following the mass extinctions in the late Permian (230 Mya; Runnegar 1987). The fossil record indicates that the genus *Placopecten* first occurred about 20-25 Mya (Hertlein 1969). Today *P. magellanicus* is the single species of the genus.

P. magellanicus is a large and long-lived species, with individuals reaching 20 years of age and achieving shell heights of 20 cm or more. The soft parts account for about 40% of the body weight and comprise a single large adductor muscle, a single adjacent gonad, a digestive gland, paired gills, and a thick muscular mantle rimming the valves and having numerous sensory tentacles and eyes. *P. magellanicus* is a filter-feeder, utilizing primarily algae, bacteria, and detrital material (Shumway *et al.* 1987).

Like other pectinids, *P. magellanicus* is capable of swimming by clapping the valves and expelling a propulsive jet of water. Juveniles are good swimmers until they reach a shell height of about 10 cm; large scallops are capable of short bursts of swimming (Dadswell and Weihs 1990).

P. magellanicus is gonochoric, with fewer than 1% of individuals being hermaphroditic (Worms and Davidson 1986). Typically the gonad ripens in late spring, with spawning occurring in summer or early fall. Naidu (1970) described the occurrence of a minor and a major spawn in early summer and fall respectively in Newfoundland populations. In a single spawn a female may release $1x10^7$ small (60-90 μm) eggs. Reproductive output shows little variation between the sexes and is directly proportional to

the size of the gonad and age of the animal (MacDonald and Thompson 1985b). Fertilization is external, followed by a pelagic larval phase of 30 to 40 days before metamorphosis (Culliney 1974). Scallops become sexually mature following the first full summer of growth, at 50-55 mm (Naidu 1970).

The species range is in the Northwest Atlantic from Cape Hatteras, North Carolina in the south to the Strait of Belle Isle, Newfoundland in the north (MacDonald and Thompson 1985a). In general, *P. magellanicus* inhabits deeper waters in the south, up to 200 m, while at the northern limit it may be found at depths of just 2 m. Local populations or beds are frequently very large, numbering millions of individuals.

Chapter 2
Materials and Methods

2.1 Collection

Scallops were obtained from six localities in Atlantic Canada and transported live on ice or packed in seaweed to the Aquatron Laboratory at Dalhousie University, Halifax, Canada, where they were held in running sea water until processed. Samples from the lower Bay of Fundy (H) and upper Bay of Fundy (I) were collected by Dr. M. Dadswell (Acadia University); the sample from the Baie des Chaleurs (C) was collected by Dr. J. Worms [Department of Fisheries and Oceans (DFO),Government of Canada]; the sample from Ste. Pierre Bank (N) was collected by Dr. S. Naidu (DFO); the sample from Georges Bank (G) was a gift of National Sea Products Ltd.; and the sample from the Eastern Shore of Nova Scotia (S) was collected by divers from Ship Harbour.

Animals were scored for sex on the basis of gross appearance of the gonad whenever possible. The female gonad begins colouring when the ovary starts to ripen, and retains a pinkish tinge for some time after emptying (Naidu 1970). Shell height and length were measured to the nearest millimetre, and age was determined from growth rings on both shell and resilium (Merrill *et al.* 1966). Meat weight was recorded as wet weight of the adductor muscle. A small sample (200-400 mg) of the adductor muscle was stored at -70° C for allozyme analysis.

2.2 Allozyme analysis

Adductor muscle samples were ground in distilled water on ice in a 1.5 ml centrifuge tube, followed by centrifugation at 14,000 rpm in an

Eppendorf microfuge. The supernatant was used as an enzyme source. Enzymes scored were phosphoglucomutase (*Pgm*, EC 2.7.5.1), octopine dehydrogenase (*Odh*, EC 1.5.1.11), mannosephosphate isomerase (*Mpi*, EC 5.3.1.8), aspartate aminotransferase (*Aat* , EC 2.6.1.1), 6-phosphogluconate dehydrogenase (*Pgd* , EC 1.1.1.44), and glucosephosphate isomerase (*Pgi* , EC 5.3.1.9).

The gel and tray buffer systems used were Tris-citrate pH 8.0 (Selander *et al.* 1971) for *Mpi* , *Aat* , *Pgi* , and *Pgd* (run for 10 h at 90 mA), the gel containing 10 g of Electrostarch, 65 g of Sigmastarch and 600 ml of buffer, and Tris-citrate pH 7.0 (Ayala *et al.* 1972) for *Pgm* and *Odh* (run for 4 h at 90 mA and 240 V), the gel containing 8 g of Electrostarch, 67 g of Sigmastarch, and 590 ml of buffer. Stain recipes were from Selander *et al.* (1971), Schaal and Anderson (1974) and Siebenaller (1979). Electromorph mobilities were assigned values according to Foltz and Zouros (1984). Mr. Kamran Ahmad ran the gels and scored alleles (unpubl. honors thesis, Dalhousie University).

2.3 Isolation of mitochondrial DNA

Mitochondrial DNA was purified from the adductor muscle of freshly killed scallops. The muscle was subdivided into 3 to 4 g sections, each of which was minced with a scalpel and homogenized with a motor-driven teflon pestle (Potter-Elvehjem tissue grinder style #2) in 10 ml of Isotonic Solution (IS) (500 mM sucrose, 150 mM KCl, 2 mM EDTA, 25 mM HEPES, pH 7.4). The homogenate was transferred to a 15 ml Corex tube with protease type XXIII (Sigma Chemical Company) at a final concentration

of 400 U/ml and incubated at least 3 h in a 37 °C water bath to reduce viscosity.

Following incubation, nuclei and cellular debris were sedimented by centrifugation for 10 min at 2500 rpm in a benchtop centrifuge. The supernatant was transferred to a fresh Corex tube and further fractionated by centrifugation at 18,000 rpm in a Beckman JA-20 rotor for 20 min. The resulting mitochondrial pellets were resuspended in 3-5 ml of IS, and at this point mitochondrial fractions from the same animal were combined in a single tube. The mitochondrial suspension was layered on a sucrose step gradient (5 ml of 1.5 M sucrose, 10 mM Tris-HCl, 5 mM EDTA, pH 8.0 and 10 ml of 1.0 M sucrose, 10 mM Tris-HCl, 5 mM EDTA, pH 8.0) in a 35 ml Beckman Ultraclear tube, and subjected to centrifugation for 1 h at 25,000 rpm in a Beckman SW28 swinging bucket rotor.

The mitochondria banded at the 1.5 M/1.0 M interface, and were collected through a bent syringe placed beneath the interface. This fraction was diluted with 10 ml of IS and the mitochondria were sedimented by another centrifugation step at 18,000 rpm. The final pellet was resuspended in 0.5 ml of IS and 1.5 ml of lysis buffer (500 mM sodium acetate, 10 mM EDTA, 10 mM Tris-HCl, 0.5% Sarkosyl, pH 8.0), and incubated at 37°C for 1 h or 4°C overnight to lyse the mitochondria. Lysis was followed by extraction, first with an equal volume of redistilled phenol equilibrated with 0.1 M Tris-HCl pH 8.0, then with an equal volume of 1:1 (v/v) phenol and chloroform, and finally with an equal volume of chloroform alone. The solution was brought to 0.2 M NaCl and the nucleic acids were precipitated by 2 volumes of 95% ethanol at -20°C for at least 2 h. Following centrifugation for 30 min at 18,000 rpm, the precipitate was dried, then

dissolved in 100 µl of TE (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). The purified mtDNA was stored at 4°C for the duration of the study.

2.4 Restriction analysis

2.4.1 Restriction digestion and mapping

For each digest, 100-200 ng of mtDNA was used. Reactions were carried out in a total volume of 15 µl for single enzyme and 20 µl for double enzyme digests, in 1.5 ml Eppendorf tubes either suspended in a 37°C water bath or in racks in a 37°C incubator for 2 - 16 h. All restriction enzymes were supplied by Bethesda Research Laboratories, MD (BRL), and 5-10 U of each enzyme were used per digest, with either 1.5 or 2 µl of the recommended 10X reaction buffer supplied by BRL. In double digests where recommended reaction buffers differed, the BRL restriction enzyme/buffer system compatibility chart was used as a guide in choosing an appropriate buffer. Following incubation, 1 µl of a freshly prepared 0.01X dilution of stock (10 mg/ml) RNase A was added to each tube. A 15 minute incubation at room temperature was sufficient to digest the RNA that might interfere with visualization of bands less than 800 bp. Reactions were stopped with 3 µl of loading buffer (25% glycerol, 0.25% bromophenol blue, 50 mM Tris-HCl, 2 mM EDTA, pH 8.0).

Restriction patterns for the endonucleases Eco RI, Sal I, Stu I, and Sph I were obtained from single and/or double digests. Restriction fragments were separated by electrophoresis in 0.8% or 1.0% agarose in either minigels (6.5 x 8 x 0.4 cm) at 80 V for 1.5 h or maxigels (15 x 20 x 0.4 cm) at 30 V for 16 h or 60 V for 6 h, in 1X TAE buffer (40 mM Tris-acetate, 1 mM EDTA). The molecular weight markers included in each gel were

Hind III-digested λ-phage and the BRL kilobase ladder. The gels were stained for 30 minutes in a 0.5 µg/ml solution of ethidium bromide, destained in dH_2O for at least 2 h, and photographed on a Fotodyne transilluminator using Kodak Tri-X Pan Professional 4" x 5" sheet film. Fragment lengths were estimated by interpolation on standard curves for the molecular weight markers.

Restriction sites for the enzymes Eco RI, Bam HI, and Pst I were previously mapped (Snyder *et al.* 1987). Maps for Hpa I, Sal I, Stu I and Sph I were constructed by double digestion with each of the three originally mapped enzymes and confirmed by double digestion with each of the remaining three enzymes. The set of mapping digests for each enzyme was done with mtDNA from a single individual to avoid confusion due to length polymorphism in comparisons of cleavage products.

The placement of the 0.7 kb Pst I fragment was determined by fragment extraction. A direct test by hybridization was not possible since the fragment shows sequence similarity with other regions. Instead, the Eco RI fragment thought to contain this fragment was cut from a gel and extracted by the freeze-squeeze method (Thuring *et al.* 1975). Approximately 100 µl of buffer solution was squeezed from the gel slice, and this was extracted once with 100 µl of phenol, once with 100 µl of iso-amyl alcohol, and twice with ether, air-dried, and resuspended in 25 µl of TE. The isolated fragment was then digested with Pst I and subjected to electrophoresis at 80 V for 1.5 h in a 6.5 x 8 x 0.4 cm 1.0% agarose gel.

The placement of several small restriction fragments derived from double digests was verified by cleavage of end-labelled Eco RI-digested mtDNA. For 40 ng of mtDNA digested to completion with Eco RI in a total

volume of 20 µl, a fill-in reaction (Maniatis *et al.* 1982) with 5 µCi of [α-^{32}P]dATP and 1 U of the Klenow fragment of DNA polymerase I was allowed to proceed for 20 min at 25°C. The reaction was stopped by the addition of one half volume of 7.5 M ammonium acetate and three volumes of 95% ethanol, and the DNA was precipitated at -70°C for 20 min. The precipitate was sedimented by centrifugation, washed with 70% ethanol, resuspended in 25 µl of dH_2O, and reprecipitated to remove remaining unincorporated nucleotides. The final pellet was resuspended in 20 µl of dH_2O and the activity of labelled fragments was determined by Cerenkov counting of DNA precipitated by trichloroacetic acid (10% TCA, 1% sodium pyrophosphate). Aliquots of about 1000 cpm were then digested to completion with Sal I, Sph I, or Hpa I, and the fragments were separated by electrophoresis at 30 V for 18 h in 1.0% agarose.

Following electrophoresis, the gel was dried on Whatman 3MM paper and placed in a cassette with Kodak X-Omat XAR5 X-ray film for exposure overnight at -70°C.

2.4.2 Restriction site polymorphism

Eco RI digests were assayed in 0.8% or 1.0% agarose in either minigels (6.5 x 8 x 0.4 cm) at 80 V for 1.5 h or maxigels (15 x 20 x 0.4 cm) at 30 V for 16 h or 60 V for 6 h. The electrophoresis conditions for Sal I/Sph I digests were 1.0% agarose in maxigels at 30 V for 16 h. For Sph I and Sal I/Stu I digests, 1.0% agarose maxigels were set up with two 15-well combs placed to give 9 cm long upper and lower halves, allowing 24 samples to be loaded on each gel. These were subjected to electrophoresis at 70 V for 4 h.

Patterns for Sal I were obtained from Sal I/Stu I and Sal I/Sph I double digests. Patterns for Stu I were scored in Sal I/Stu I double digests. Variants for Sal I and Stu I were subsequently confirmed by the separation of single digest fragments by electrophoresis at 30 V for 16 h in 0.8% agarose maxigels.

Cleavage patterns for each enzyme were given letter designations, with A representing the most frequent morph. Composite morphs are thus described by a four letter code.

2.4.3 Length polymorphism

The four enzymes used in the survey for length polymorphism produced fragments small enough to reveal size changes in any region of the molecule. Size variation at locus I was scored according to the length of Eco RI-B. Following the complete survey, standardization of length estimates among gels was carried out by running nested single size class gels of samples drawn from a matrix of similarity. In this way, length estimates from each survey gel were tested for identity with each other gel. Size variants are designated by a number from 2 to 8, corresponding to repeat copy number. The inferred copy number for each individual in the Eco RI survey was confirmed by scoring the length of Sal I-B/Sph I-A. This fragment is smaller than Eco RI-B, and for most size classes is within the linear range of the gel.

Size variation at locus II was scored as the length of Sal I-B/Stu I-D. Where there was ambiguity in the assignment to a size class because of co-migration with another fragment, a Sal I/Pst I double digest was analyzed for clarification. Variants are designated by letter, with A representing the smallest and F the largest class.

Size variation at locus III was scored as the length of Sal I-A/Sph I-C. This fragment is small enough to permit discrimination of bands differing by as little as 50 bp. Because the number of size alleles was large and differences small, standardization among gels was not feasible and identity of alleles could not be assumed. Instead, alleles were lumped into three categories. All alleles of the fragment with length <4.0 kb were designated S, those with length ≥4.0 and ≤4.3 kb were designated M, and those with length >4.3 kb were designated L.

2.5 Southern transfer and hybridization

2.5.1 Isolated mtDNA

One hundred ng aliquots of sample mtDNAs were digested to completion and subjected to electrophoresis in 0.8% or 1.0% agarose maxigels at 30 V for 16 h or 60 V for 6 h, stained with ethidium bromide and photographed in ultraviolet light. Gels were then soaked for 20 min in 0.25 N HCl, twice for 20 min in 1.5 M NaCl / 0.5 M NaOH, and twice for 20 min in 3.0 M sodium acetate, pH 5.5. The DNA was transferred to ICN Biotrans™ nylon membranes by capillary transfer in 6X SSC (1X: 150 mM NaCl, 15 mM sodium citrate) for 16 h. Membranes were baked *in vacuo* at 80°C for 2 h and stored wrapped until hybridization.

Hybridization was performed according to the manufacturer's recommendations for Biotrans. Membranes were prehybridized for at least 4 h in hybridization fluid (50% formamide, 5X SSC, 5X Denhardt's solution, 50 mM sodium phosphate, 0.1% SDS, and 250 µg/ml sheared denatured calf thymus DNA or yeast RNA). Two membranes were sealed in a Ziplock (Dow Brands Canada, Inc., Ont.) bag with 25 ml of hybridization fluid, and

incubated in a water bath at 42°C. For hybridization, 2-5 x 10^6 cpm of probe DNA was added to the bag, and it was returned to the water bath for 16 h at 42°C. Membranes were washed twice with gentle agitation for 30 min at 25°C in 2X SSC / 0.1% SDS, and twice for 30 min at 42°C in 0.2X SSC / 0.1% SDS, then wrapped in Saran Wrap (Dow Brands Canada Inc., Ont.) or sealed in fresh bags and placed in cassettes with Kodak X-Omat XAR5 X-ray film and Dupont Cronex™ intensifying screens for exposure overnight at -70°C.

Probe DNA was obtained by isolating the supercoiled fraction of the pooled mtDNA of four scallops (120 g of muscle tissue) purified by the standard protocol to the lysis stage. The lysate was divided among six tubes, brought to a density of 1.619 g/ml with CsCl and 220 µg ethidium bromide, and subjected to centrifugation at 45,000 rpm at 20°C for 24 h in a Beckman 70ti rotor. The supercoiled bands were removed by pipette and combined, and this solution was extracted with four washes of isopropanol, diluted with one volume of water, and precipitated by 2 volumes of 95% ethanol at -20°C overnight. The mtDNA pellet was dissolved in TE, and a small aliquot was digested with Eco RI and checked for purity by electrophoresis. The mtDNA was then reprecipitated and stored at 4°C in 2 volumes of 95% ethanol until needed, when it was pelleted, washed, dissolved in 80 µl of TE and kept at -20°C.

Probe mtDNA was radioactively labelled by either nick translation (Rigby *et al.* 1977) or random primer extension (Feinberg and Vogelstein 1983, 1984). For nick translation, 5-10 µl of probe DNA was denatured by heating at 100°C for 5 min and chilling on ice, then incubated at 15°C for 1 h in NT buffer (150 mM Tris-HCl, 10 mM $MgSO_4$, 0.1 mM DTT, 500 µg/ml

BSA, pH 7.2) with 1 µl each of 1 mM dCTP and TTP, 25 µCi each of [α-^{32}P]dATP and [α-^{32}P]dGTP, 0.1 ng of DNase I and 7 U of DNA Pol I in a total reaction volume of 50 µl. The reaction was stopped with 2 µl of 0.5 M EDTA and 48 µl of STE (100 mM NaCl, 10 mM Tris, 0.1 mM EDTA pH 8.0), and labelled DNA was purified by elution from a 1 ml column of Sephadex G-50 equilibrated with STE.

For random primer extension, 5-10 µl of probe DNA was linearized by restriction digestion and denatured by one of two methods. Either the digest was boiled for 5 min and held on ice, or it was denatured by incubation with 0.1 volume of 2N NaOH / 2 mM EDTA at 25°C for 5 min, precipitated by 0.1 volume of 3 M sodium acetate, pH 5.5, and 2 volumes of 95% ethanol at -70°C for 20 min, pelleted, and vacuum dried. Boiled DNA was labelled by incubation with 15 µl of BRL Random Primer Buffer Mixture (0.67 M HEPES, 0.17 M Tris-HCl, 17 mM $MgCl_2$, 33 mM ß-mercaptoethanol, 1.33 mg/ml BSA, 18 OD units/ml hexanucleotide primers, pH 6.8), 2 µl each of 1 mM dCTP and TTP, 25 µCi each of [α-^{32}P]dATP and [α-^{32}P]dGTP, and 4 U of Klenow enzyme in a reaction volume of 50 µl at 25°C for 1.5 h. Alkali-denatured DNA was labelled by addition of 5 µl of 10X BRL React™ 2 buffer, 3 µl of random hexanucleotide primer mix, 3 µl each of 1 mM dCTP and TTP, 25 µCi each of [α-^{32}P]dATP and [α-^{32}P]dGTP, 4 U of Klenow enzyme and dH_2O to a final volume of 50 µl. The reaction was incubated for 1.25 h at 37°C and stopped with 5 µl of 0.5 M EDTA. Unincorporated labelled nucleotides were not removed from random primer reactions. The specific activity of labelled fragments was determined by Cerenkov counting of TCA-precipitated aliquots.

Whether labelled by nick translation or random primer extension, probe DNA was denatured by boiling for 5 min and chilling on ice for 10 min before adding to the hybridization bag.

2.5.2 Total DNA blots

Total nucleic acids were isolated from 200-400 mg portions of gill, gonad, and muscle tissue using a PEG precipitation protocol (J.-M. Sevigny, pers. comm.). As a control, the remaining adductor muscle was processed with the usual mtDNA isolation protocol. Separate mortars and pestles were used for each tissue, and these were thoroughly washed between individuals.

The excised tissue was ground to a powder in liquid nitrogen and transferred to a test tube with 5 ml of TES (10 mM Tris-HCl, 100 mM EDTA, 10 mM NaCl, pH 8.0). To this was added 0.1 volume of 10% SDS and 25 µl of proteinase K (20 mg/ml). The tube was shaken gently to suspend the powder, and incubated overnight at 37°C. Following incubation, the viscous solution was gently extracted by slow rotation for 15 min twice with one volume of 1:1 (v/v) phenol/chloroform, and once with one volume of chloroform alone. Nucleic acids were precipitated by 0.1 volume of 3.0 M sodium acetate, pH 5.5, and 2 volumes of 95% ethanol at -20°C for 4 h. The precipitate was collected by centrifugation, washed with 80% ethanol, dried, and dissolved in 5 ml of TSE (10 mM Tris-HCl, 0.1 mM EDTA, and 250 mM NaCl, pH 8.0). Nucleic acids were precipitated by an equal volume of 20% (w/v) PEG-8000 for 1 h at 25°C, and sedimented by centrifugation. After washing with 80% ethanol and drying, the pellet was dissolved in 2 ml of TE and reprecipitated by sodium acetate and ethanol. The final pellet was dissolved in 1 ml of TE, and the DNA was stored at 4°C until needed.

The concentration of nucleic acid samples was estimated by spectrophotometric readings at A_{260} of diluted aliquots. Two aliquots of 5 µg of DNA were digested 16 h with 5 U Sal I and 5 U Sph I, and 16 h with 10 U of Hpa I. Where volumes of restriction digests exceeded well capacity, digests were ethanol precipitated and resuspended in 25 µl of water before loading.

Restriction enzyme digests were subjected to electrophoresis in 1.0% agarose maxigels at 30 V for 16 h, stained with ethidium bromide and photographed. The DNA was blotted to nitrocellulose (Schleicher & Schuell) by capillary transfer in 6X SSC following soaking of the gel for 20 min in 0.25 N HCl, twice for 20 min in 1.5 M NaCl / 0.5 M NaOH, and twice for 20 min in 1.5 M NaCl / 0.5 M Tris pH 8.0. Filters were rinsed in 4X SSC, air dried, and baked *in vacuo* 2 h at 80°C.

Hybridization was carried out in Ziplock bags as described in section 2.4.1. Alkali-denatured plasmid pEK (gift of D. Cook, Dalhousie U.), containing part of restriction fragment Eco RI-B/Stu I-D, was labelled to a specific activity of 10^8-10^9 cpm/µg by random primer extension and used as probe. Each membrane was hybridized and washed in a separate bag. After washing, membranes were sealed in fresh bags and placed in film cassettes for exposure overnight with screens and/or for 3-7 days without screens.

2.6. End-labelling and densitometry

Quantification of intra-individual variation in samples identified as heteroplasmic by Southern blot screening was done by end-labelling. Labelling was performed by the rapid fill-in method (Maniatis *et al.* 1982), where 2 µl of sample was first digested to completion with Eco RI in a

reaction volume of 20 µl and then incubated at 25°C for 20 min in the presence of 5 µCi of [α-^{32}P]dATP and 1 U of Klenow enzyme. Unincorporated labelled nucleotides were removed with two precipitations by 0.5 volume of 7.5 M ammonium acetate and 3 volumes of 95% ethanol. The final pellet was washed with 70% ethanol, dried, and resuspended in 20 µl dH_2O. Incorporation was determined by Cerenkov counting of TCA-precipitated aliquots.

Labelled samples were electrophoresed in 0.6% agarose maxigels at 30 V for 16-18 h, with 1000 cpm per lane. Hind III-digested λ-phage, labelled by the same method, was included as a size standard. Following electrophoresis, gels were dried on Whatman 3MM paper and loaded into cassettes with Kodak X-Omat XAR5 X-ray film and intensifying screens for exposure overnight at -70°C.

Each sample was labelled twice, in independent labelling reactions, and 1000 cpm aliquots from each labelling reaction were run on two different gels. Because signal intensities varied among lanes, primarily due to different levels of background, several exposures of some gels were sometimes required to obtain equivalent signals for each sample.

The relative proportions of heteroplasmic bands were determined by scanning films with a Bio-Rad Model 620 video densitometer interfaced with a Zenith computer using the Bio-Rad 1-D Analyst® software package. The film background was subtracted, and a valley-to-valley baseline was drawn for each scan. Relative proportions were calculated from the area under the peaks.

Chapter 3
Molecular Characterization

3.1 Introduction

The metazoan mitochondrial genome shows remarkable consistency across widely divergent taxa in a number of basic characteristics, including genome structure, gene content, and molecular size. Departures from the structural norm [two linear molecules rather than a covalently closed circle in *Hydra* (Warrior and Gall 1985)] and the gene content norm [lack of ATPase8 in *Ascaris suum* (Wolstenholme *et al.* 1987)] are very rare, but a number of species with unusually long mtDNA have now been described. Estimates of genome size obtained from restriction mapping now range from 14.3 kb (*A. suum* ; Wolstenholme *et al.* 1987) to 42 kb (*Placopecten magellanicus* ; LaRoche *et al.* 1990), but in the majority of species the molecule measures between 15.7 and 19.5 kb (Brown 1983).

Enlarged genomes generally have either a tandem array of noncoding sequences or a tandem duplication of coding sequence (Moritz *et al.* 1987). Although loss or deletion of coding sequences has been reported (Boursot *et al.* 1987, Holt *et al.* 1988a, 1988b, 1989, Schon *et al.* 1989, Wallace 1989), no additions to the 37 known genes have ever been documented in animal mtDNA. Until more repeat sequences are known, the question of whether noncoding tandem repeats are always derived from mitochondrial sequences will remain unanswered.

In species for which an expanded mitochondrial genome has been reported, the region of expansion has most often been linked to the control region. Tandemly repeated sequences and associated length polymorphism

have been mapped to a position within or adjacent to the control region in *Drosophila* (Solignac *et al.* 1986a), shad (Bentzen *et al.* 1988), sturgeon (Buroker *et al.* 1990), crickets (Rand and Harrison 1989), and *Cnemidophorus* lizards (Densmore *et al.* 1985, Moritz and Brown 1986). In *Cnemidophorus* , the repeat is a duplication encompassing the entire control region and the adjacent rRNA genes, and it increases the genome size by as much as 9 kb. The repeat in shad adds either 1.5 or 3 kb to the mtDNA molecule, while the small repeats in insects and sturgeon increase the genome size by less than 2 kb.

Only four cases of unusually long mtDNA have been reported that cannot be explained by a single locus of tandem repeats. *Cnemidophorus* has a length variable complex of small 64 bp tandem repeats within the D-loop but distinct from the large duplication (Densmore *et al.* 1985). In the parasitic nematode *Romanomermis culicivorax* , tandem arrays of both partial and complete copies of the same repeat are present in two widely separated regions, one array inverted with respect to the other (Hyman *et al.* 1988). In one species of *Hyla* frogs, several regions show length polymorphism (Kessler and Avise 1985). Most recently, Gjetvaj (1989) has documented length variation due to a tandem array and to discrete deletions and insertions elsewhere in the mtDNA of the pectinid *Chlamys islandica*.. In each of these species the mitochondrial genome size has been estimated to be at least 23 kb, considerably larger than the overall genome sizes reported for those species with single tandem arrays.

Through restriction mapping I have localized length variation in the mtDNA of *P. magellanicus*, and determined the arrangement of constant and variable regions in the molecule. The restriction sites were distributed in

such a way that no large segment was uncleaved, and fragments were small enough to reveal even minor size variation. This permitted m to estimate the overall genome size exclusive of length polymorphic loci.

3.2 Results

3.2.1 Isolation

Mitochondrial DNA was isolated from 280 individual scallop adductor muscles. The isolation protocol typically yielded 1 - 3 μg of mtDNA per 10 g of tissue. The yield was influenced by aliquot size at two stages in the procedure. Grinding of small aliquots (less than 2 g) resulted in less viscous and more uniform homogenates, which increased the efficiency of the protease digestion and gave higher yields of organelles in the centrifugation step. In the sucrose gradient, the organelle fraction from about 10 g of tissue was the optimal load. The yield from a 20 g muscle was increased by about 150% if the organelle fraction was divided between two gradient tubes rather than pooled in one.

The nutritional and reproductive status of the animal appeared to influence the amount of glycogen co-purifying with mtDNA. In some cases the mtDNA solution appeared milky because of high glycogen content, but this did not interfere with restriction digestions. Nuclear contamination was generally low, and did not produce significant background even in end-labelled digests.

Other scallop tissues could not be processed with this protocol. Attempts to isolate mtDNA from mantle, gill, and gonad resulted either in low yields of degraded mtDNA or in mtDNA that was resistant to restriction digestion.

3.2.2 Restriction mapping

One of the most striking observations about the *P. magellanicus* mitochondrial genome was its size. Restriction digestions consistently

produced fragments summing to more than 30 kb. It has been shown by CsCl equilibrium density centrifugation and electron microscopy that the genome is a covalently closed circular molecule (Snyder *et al.* 1987), and restriction mapping confirmed the pattern of a circular molecule, where the number of fragments equals the number of cleavage sites and enzyme patterns overlap.

In addition to its large size, the genome exhibits a high frequency of size variation. The variation occurs at three distinct loci, one of which is known to be a repeat array (Snyder *et al.* 1987). The consensus restriction profile of *P. magellanicus* reflects not only the most common cleavage patterns, but also the most common length for each variable locus.

The average length of the consensus mtDNA molecule obtained from the cleavage products of Sal I, Sph I, Stu I, Hpa I, Eco RI, Bam HI, and Pst I was 35 kb. The total number of sites in the consensus profile was 39 (Table 3.1); the numbers of sites for the smallest and largest genomes scored were 35 and 47 respectively. In most cases, digests produced at least one fragment larger than 10 kb, which lies outside the linear range of the gel. Size estimates for these fragments were confirmed by double digests with other enzymes, which also served to check that small fragments were not overlooked.

A consensus restriction map derived from these digests is shown in Figure 3.1. The map has been linearized at a Sal I site that served as a reference point for two length-variable loci. The position of Pst I-D, formerly assigned to a location between Pst I-A and Pst I-B (Snyder *et al.* 1987), has been changed. Pst I digestion of Eco RI-B, extracted from a gel, confirmed that Pst I-D lies between Pst I-A and Pst I-C. The positions of Stu

I-E and -H could be reversed; no reference site was available with which to test their order. The same holds true for Hpa I-F and -G.

Cases of expanded mitochondrial genomes have been reported in which there is a distinct asymmetry of cleavage sites, with clustering of sites in one portion (Densmore *et al.* 1985, Boyce *et al.* 1989). This has been interpreted as an indication of an A+T rich region inaccessible to most restriction endonucleases. It is clear from the *P. magellanicus* map that sites are fairly evenly distributed and the longest segment without a cleavage site is Sal I-B/Sph I-D, which is just 3.4 kb (see Figure 3.1). This region is part of a contiguous stretch of nearly 16 kb that is invariant for length. Furthermore, digestion with Dra I which has a recognition site that is entirely A+T failed to produce an abundance of small fragments in *P. magellanicus* mtDNA, although it did result in extensive cleavage of the 16 kb mtDNA of the pectinid *Argopecten irradians* (D. Cook, Dalhousie U., pers. comm.).

The total length of the genome is close to twice that of most metazoan mtDNA . This raises the question of whether it is in fact a dimer. The distribution of restriction sites indicates that this cannot be so. Each enzyme shows an asymmetrical distribution of sites, yet a consistent circular map can be derived. Considered together, the cleavage sites leave no large region intact that could be construed as a duplicated region resistant to digestion.

3.2.3 Functional mapping

A limited amount of functional mapping has been done. An end-labelled mtRNA preparation used as a probe to locate ribosomal genes hybridized to the Eco RI-A/Bam HI-D fragment (LaRoche *et al.* 1990). Subsequently, a heterologous hybridization was done using the *Paracentrotus*

lividus clone Bam 1-1 (Cantatore *et al.* 1987), a gift of M. Cantatore, containing the genes for 16S rRNA , CO I, and about one third of ND 2. This clone hybridized only to the Hpa I-F fragment, within the Eco RI-A/Bam HI-D fragment (LaRoche *et al.* 1990).

Another *P. lividus* clone kindly provided by M. Cantatore, Bam 1-6, was useful in locating the CO III gene in *Placopecten magellanicus*. This clone hybridized only to Eco RI-D/Bam HI-C. Further investigation of the location of the cytochrome oxidase subunit genes was done by using the pdyHB clone of *Drosophila yakuba* mtDNA (Clary and Wolstenholme, 1985), a gift of D. Wolstenholme. The mitochondrial fragment in the clone contains a unique Cla I site that cleaves the segment into two fragments, one containing CO II and CO III, the other containing CO I, ATPase 6 and 8, and several tRNAs. The CO II/CO III fragment hybridizes with the 1.3 kb fragment Sal I-A/ Sph I-D, and the CO I fragment hybridizes with Eco RI-D (D. Cook, Dalhousie University, pers. comm.). The positions of identified genes are indicated in Figure 3.1 as bars under the map.

Table 3.1. Restriction profile of *Placopecten magellanicus* mtDNA. Fragment sizes are given for a consensus genome, having the most common restriction patterns and most common lengths in variable regions.

Sph I	Hpa I	Pst I	Stu I
14.8	20.0	19.0	7.6
8.1	4.5	10.8	7.2
6.1	3.5	1.45	7.1
4.6	2.8	1.45	3.6
1.6	2.1	1.45	2.1
35.2	1.3	0.7	1.5
	0.7	**34.85**	1.45
	34.9		1.45
			1.45
			1.3
			34.75

Eco RI	Bam HI	Sal I
17.8	17.0	15.5
10.8	9.0	10.9
4.2	5.2	8.6
2.2	3.8	**35.0**
35.0	**35.0**	

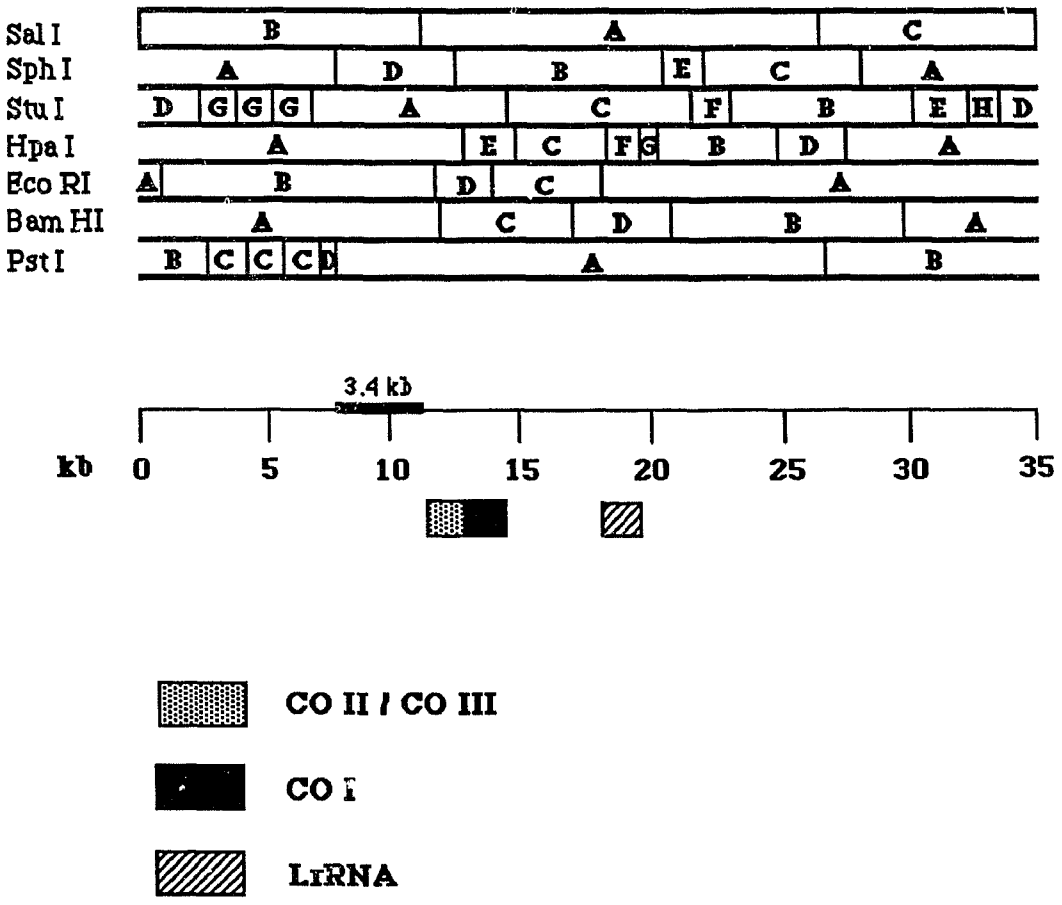


Figure 3.1 Restriction map of *P. magellanicus* mtDNA. The mtDNA genome is shown as a linear molecule starting at a Sal I site. The letters designate fragment sizes in decreasing order, with A the largest fragment for each enzyme. Repeat unit fragments are given identical designations (G for Stu I, C for Pst I). The longest uncleaved segment, 3.4 kb, is positioned on the distance line below the map. Functional regions are denoted by bars under this line.

3.3 Discussion

3.3.1 Isolation

Obtaining clean preparations of molluscan mtDNA can be difficult, as other workers have noted. Skibinski (1985) routinely used ripe mantle, full of eggs, in *Mytilus* preparations, and found the mtDNA susceptible to digestion by restriction enzymes but somewhat obscured in gels by a high nuclear DNA background. Stine (Johns Hopkins Medical Institutes, pers. comm.), in gastropod preparations, sedimented mitochondria through a sucrose cushion in order to isolate them from abundant mucopolysaccharides in the homogenate. If DNA was exposed to these mucopolysaccharides, the resultant complex would be completely resistant to restriction digestion. I found that the sucrose gradient step was essential to prevent degradation of mtDNA during lysis and extraction, but it was only effective for adductor muscle preparations.

In scallops, eggs are held in the gonad adjacent to the single adductor muscle and cannot be sampled with mantle. Furthermore, the mantle is a muscular structure with numerous tentacles and eyes, and is frequently pigmented with symbiotic algae (Naidu and South 1970). The difficulty in mechanically disrupting the cells, in conjunction with the possibility of co-purifying algal genomes, make mantle an unsuitable source of mtDNA. The same reasons apply to the hepatopancreas, or digestive gland.

Of the remaining tissues, adductor muscle, gill, and gonad, only adductor consistently yielded clean mtDNA. Although an abundant mitochondrial fraction could be recovered from gill and gonad, the purified mtDNA was often degraded. The addition of proteinase K and 100 mM EDTA during lysis prohibited degradation, but the mtDNA was then

resistant to digestion. Hybridization of mtDNA derived from adductor muscle to restricted total DNA preparations from other tissues (see Chapter 5) revealed identical mtDNA restriction patterns in the other tissues, indicating that the adductor mtDNA pattern was typical and not an artefact..

3.3.2 Restriction mapping

The length of *P. magellanicus* mtDNA is clearly exceptional. At the extreme of the intraspecific range, 42 kb, it is the largest mitochondrial genome reported in any animal species. Departures from the 15.7 to 19.5 kb range that is the metazoan norm have been described in several species, and the unusual size does not by itself make the mtDNA of *P. magellanicus* unique. Yet there are reasons to doubt that *P. magellanicus* merely occupies one end of a continuum of genome sizes.

As Moritz and co-authors (1987) discussed, cases of enlarged mitochondrial genomes generally involve single or multiple tandem duplications that typically encompass or lie within the control region. The few exceptions are genomes where repeated sequences are dispersed. A common factor, however, is that the increase in size beyond the norm can be accounted for by the repeated sequences except in the pine weevils (Boyce *et al.* 1989) where an expanded A+T rich region adjoins the repeat region that comprises different numbers and sizes of repeats.

This is not the case for *Placopecten* . When length variable regions are totally removed from the restriction fragment summation, the genome is still about 28 kb. The restriction map shows no evidence of large duplications, and there are few regions without multiple sites where duplications could remain cryptic unless they were very small and numerous. How then can this

"extra" DNA be explained? There is no evidence for the import of nuclear sequences into scallop mtDNA; the total DNA blots probed with mtDNA showed no hybridizing fragments that could not be accounted for by the mtDNA pattern. The most likely explanation must be duplications, but these must have occurred long enough ago for subsequent evolution to have obscured the pattern at the level of restriction analysis.

The overall asymmetry of the cleavage sites precludes the possibility that the genome exists as a dimer, but an ancestral duplication of the entire molecule cannot be ruled out as an explanation of the size. There are two lines of evidence suggesting that the genome may be functionally divided into two parts. First, there is a contiguous stretch of about 15.4 kb, including Sph I-B,-D, -E and Stu I-F, within which no length variation has been observed. Second, it is within this region that functions have been mapped. The cytochrome oxidase complex is situated in Sal I-A/Eco RI-B and EcoRI-D, and the LrRNA gene lies about 4.5 kb distant, in Hpa I-F.

If indeed the functional regions are packed within the invariant segment, several questions about the organization of the genome remain outstanding. For example, the position of the control region in a molecule that may have interrupted or widely separated coding sequences is of obvious importance in transcription. If transcription proceeds along the entire strand from a single initiation site, a long nascent transcript could influence elongation or the stability of the primary transcript. The vertebrate model of transcriptional attentuation to maintain higher levels of rRNAs relative to mRNAs (Dubin *et al.* 1985) depends upon the rRNA genes being the first transcribed. Jacobs *et al.* (1988) suggested that there may be multiple transcriptional units in the sea urchin *S. purpuratus* , and a similar system

may exist in the scallop that would accommodate the challenges of increased genome size or interspersion of coding and non-coding regions. Also, if replication is asynchronous as in vertebrates (Clayton 1982) and insects (Clary and Wolstenholme 1985) then the displaced strand would be exposed as single-stranded DNA twice as long as in a 16 kb molecule. This could result in an increased susceptibility to damage or to a higher frequency of intra-strand interactions contributing to replication errors.

Chapter 4

MtDNA Polymorphism

4.1 Introduction

Two kinds of polymorphism can be detected by restriction analysis. Nucleotide substitutions may lead to gain or loss of cleavage sites, resulting in novel fragments. Addition/deletion events can also cause changes in fragment mobility. In comparisons among conspecifics, size changes are generally easily distinguished from changes due to site loss or gain because the number of fragments produced by a given enzyme remains constant while the sum of fragment lengths varies. When cleavage sites are mapped, the possibility that site mutation is responsible for observed size changes can be virtually eliminated. Digestion of a variant genome with a suite of mapped enzymes will show concordant length increase or decrease in the variant region.

Intraspecific length polymorphism over a range up to about 150 bp is common in animal species (Moritz *et al.* 1987). Changes in the length of short homopolymer runs have been documented in humans (Cann and Wilson 1983), rats (Brown and DesRosiers 1983) and cows (Hauswirth *et al.* 1984), and short sequence duplications are most likely the basis for observed minor variation in many other taxa. Observed major length variation is less common. Simple tandem repeats have been characterized in insects (Solignac *et al.* 1984, Rand and Harrison 1989), and sturgeon (Buroker *et al.* 1990) but the basis of length variation in other taxa is often considerably more complex. In lizards several independently derived duplications of the D-loop and the adjacent rRNA region exist in conjunction with a variously amplified

64 bp repeat (Densmore *et al.* 1985). In the nematode *R. culicivorax* partial and complete copies of a 3.0 kb sequence are dispersed singly or in tandem arrays about the mtDNA molecule (Beck and Hyman 1988, Hyman *et al.* 1988). A 1.5 kb sequence associated with the D-loop in shad is present in two or three copies, and itself shows length variation (Bentzen *et al.* 1988), a feature it shares with the large repeat in weevils (Boyce *et al.* 1989). Restriction cleavage patterns in *Hyla* reveal dispersed polymorphic regions that have not yet been characterized (Kessler and Avise 1985), whereas length variation in *Rana esculenta* is localized but continuous rather than discrete (Monnerot *et al.* 1984). Skibinski (1985) and Edwards and Skibinski (1987) noted length variation in one region of mussel mtDNA, and Gjetvaj (1989) documented size variation in six scallop species comprising small deletions, additions, and incrementally varying regions.

Few studies of the range and distribution of length polymorphism in natural populations have been carried out, but available reports indicate that tremendous diversity exists in the amount and frequency of variation. In *Drosophila* , differences in repeat copy number appear more characteristic of species than intraspecific lineages (Solignac *et al.* 1986), while the two species of *Gryllus* described by Harrison *et al.* (1987) and Rand and Harrison (1989) show a broad overlap, with similar ranges and no significant difference in size class distribution. In the shad only two size classes were observed, with the smaller being much more frequent and the larger occurring only in heteroplasmy (Bentzen *et al.* 1988). The minimal array size is most often two in systems having tandem repeats, although one individual cricket contained a single copy of the repeat in heteroplasmy with a multiple copy array (Rand and Harrison 1989).

In *P. magellanicus* , the mitochondrial genome has three regions that may vary in length by more than 1000 bp, and at each of those loci no allele has a frequency higher than 0.9. The polymorphic loci are dispersed, and I have scored the mtDNA isolates independently for each locus. The restriction enzymes used to assay size variation were mapped and site polymorphism was noted to ensure that the two kinds of variation could not be confused. The 280 mtDNA samples were isolated from adductor muscle as described in Chapter 2, and scored in ethidium bromide-stained agarose gels. Sample numbers vary slightly among assays due to some loss during storage.

4.2 Results

4.2.1 Restriction polymorphism

The primary purpose of this study was a description of size variation, and the four enzymes used to survey scallop samples were chosen because they defined length polymorphism well. Each enzyme was informative, with at least one polymorphic site (Table 4.1).

Two restriction patterns were observed for Stu I, distinguished by a site loss between contiguous Stu I-G fragments that produced a novel 2.9 kb fragment (Figure 4.1a). Three patterns were observed for Sal I. In pattern B a site loss between Sal I-B, with 7 repeats, and Sal I-C produced a new fragment of 24.1 kb. In pattern C a site gain within Sal I-B, with 4 repeats, produced two new fragments of 6.6 kb and 4.3 kb. For Eco RI, a site loss between Eco RI-A and -B generated a new fragment of 28.6 kb in pattern B, while in pattern C a site gain in Eco RI-A resulted in novel 11.5 and 7.5 kb fragments. Sph I showed the most variation. In pattern B a site loss between Sph I-A and -D resulted in a new fragment of 19.4 kb. A site loss between Sph I-B and -D produced a novel fragment of 12.7 kb in pattern C. Both sites were lost in pattern D, resulting in an uncleaved Sph I-A, -B, and -D fragment of 28.9 kb.

Of the seven polymorphic sites, two occur within the tandem repeat and the others are dispersed about the molecule outside size variable regions (Figure 4.1b). Only one site lies within a region known to contain coding sequences; Sph I-B/D occurs in the cytochrome oxidase gene array localized by heterologous hybridization as described in Chapter 3.

When enzyme cleavage patterns are combined, there are nine composite restriction morphs (Table 4.2). The morph representing the most

common cleavage pattern for each of the four enzymes together, A, had the highest frequency, 0.79. Each of the other eight composites represented the common pattern for three enzymes and a variant for the fourth, and six of them had frequencies under 0.01. The remaining two morphs, D and E, had frequencies over 0.08. Four instances of restriction site heteroplasmy were noted for Sph I, and in scoring morphs genomes were counted rather than individuals. Thus, heteroplasmics were counted once as pattern A and again as pattern B. The clonal diversity was calculated according to Nei and Tajima (1981) who measure nucleomorph diversity as:

$$\hat{H} = N(1 - \Sigma x_i^2) / (N-1)$$

where N is the sample size and x_i is the frequency of the i^{th} nucleomorph or composite restriction morph. The resulting value is 0.365. The nucleotide diversity was also calculated according to Nei and Tajima (1981), where diversity is given by

$$\Pi = n (1-\Sigma x_i \Sigma x_j \Pi ij) / (n-1)$$

where x_i and x_j are the frequencies and Πij is the divergence estimate of the i^{th} and j^{th} nucleomorphs respectively. Πij is obtained from the number of site differences between the $i^{th}$and j^{th} morphs. Because one of the polymorphic enzymes cleaves within the repeat array, a change in the number of sites was not strictly sequence dependent in this sample. To accommodate this all genomes were normalized to two repeat copies, the minimum number seen. The array could not be excluded since it does contribute to site polymorphism. The value obtained for Π was 0.0017.

Table 4.1 *P. magellanicus* mtDNA restriction cleavage variants. The cleavage patterns are designated by letter, from most to least frequent, and fragment lengths are given in kb. Numbers in parentheses refer to actual repeat copy number in the individual displaying the pattern.

	pattern			
enzyme	**A**	**B**	**C**	**D**
Stu I	(3) 7.6	(3) 7.6		
	7.2	7.2		
	7.1	7.1		
	3.6	3.6		
	2.1	2.9		
	1.5	2.1		
	1.45	1.5		
	1.45	1.3		
	1.3			
Sal I	(4) 15.5	(7) 24.1	(4) 15.5	
	10.9	15.5	8.6	
	8.6		6.6	
			4.3	
Eco RI	(4) 17.8	(4) 28.6	(5) 12.2	
	10.8	4.2	10.0	
	4.2	2.2	7.8	
	2.2		4.2	
			2.2	
Sph I	(4) 14.8	(4) 19.4	(4) 14.8	(4) 27.5
	8.1	8.1	12.7	6.1
	6.1	6.1	6.1	1.6
	4.6	1.6	1.6	
	1.6			

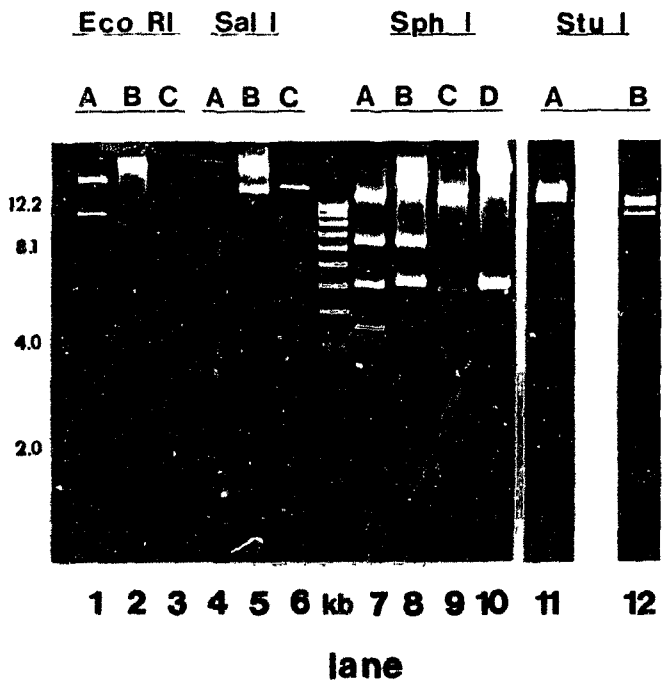


Figure 4.1a Restriction site polymorphisms. Cleavage patterns are shown for each of the informative enzymes. Numbers on the left refer to fragments in the kilobase ladder size standard (kb). The Stu I patterns in lanes 11 and 12 were obtained from different gels than the patterns shown in lanes 1 to 10.

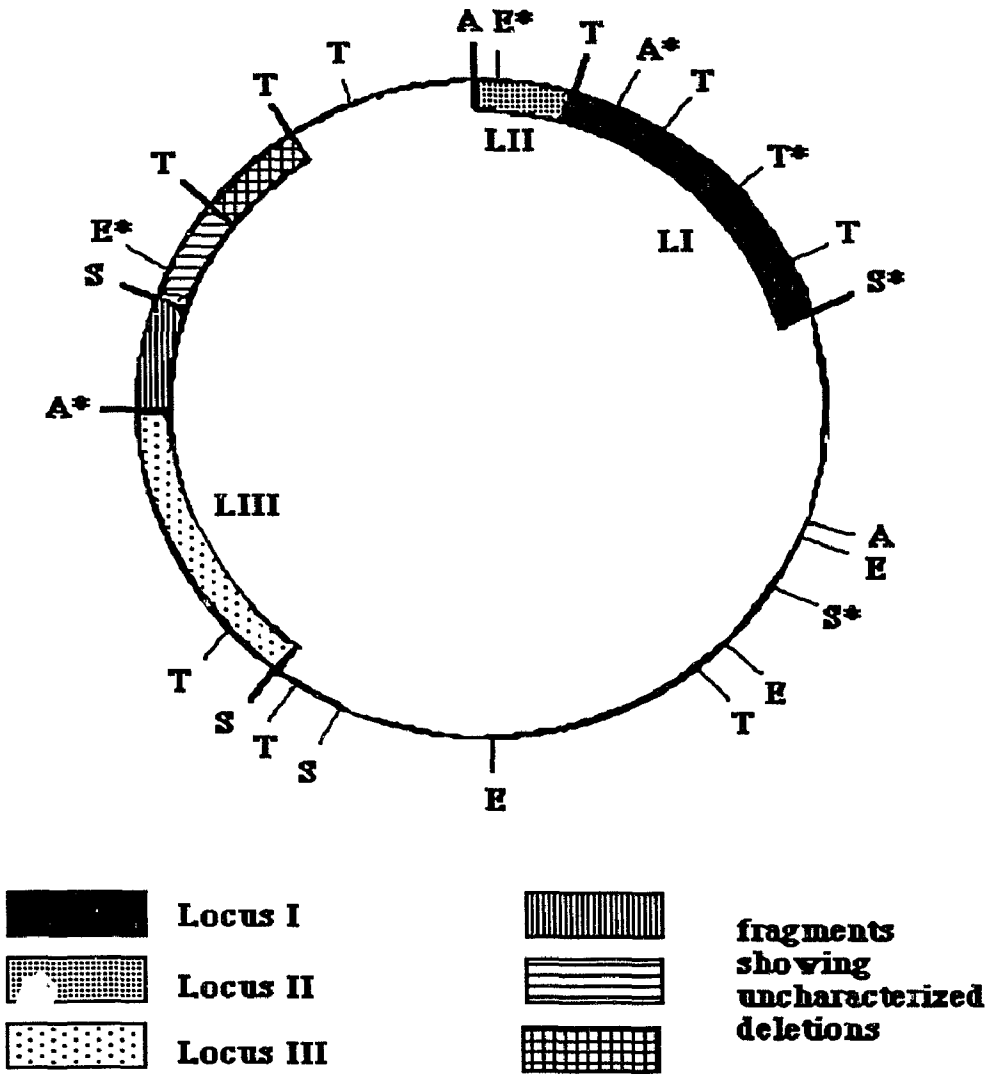


Figure 4.1b Circular restriction map showing polymorphism in *P. magellanicus* mtDNA. Polymorphic restriction sites are marked with an asterisk. A, Sal I; E, Eco RI; S, Sph I; T, Stu I. The fragments showing length polymorphism are boxed and shaded, with labels in the interior of the circle for those regions which have been characterized.

Table 4.2 Composite restriction morphs. Composition designations reflect cleavage patterns for the following enzymes, in order: Eco RI, Sph I, Sal I, Stu I. **N**, number of genomes; x_i, frequency of morph.

morph	composition	N	x_i
A	AAAA	220	.7857
B	BAAA	2	.0071
C	CAAA	1	.0036
D	ABAA	29	.1035
E	ACAA	24	.0857
F	ADAA	1	.0036
G	AABA	1	.0036
H	AACA	1	.0036
I	AAAB	1	.0036

4.2.2 Length polymorphism

4.2.2.1 Locus I

The region of the *Placopecten magellanicus* mitochondrial genome exhibiting the greatest range in size is contained within Eco RI-B. Mapping, cloning, and sequencing studies have shown that the major variation in this fragment can be attributed to differences in copy number of a 1449 bp tandemly repeated element (LaRoche *et al.* 1990).

The molecular organization of the tandem array was deduced from the sequences of one complete repeat and the 3' flanking fragment, both bounded by Pst I sites and cloned from a single individual, as well as part of the 5' flank from a second individual. Sequencing into the 3' and 5' flanks continued at least 20 bp beyond sequence similarity with the repeat. Both ends of the array are marked by the short direct repeat ACTTTAA, which was described as the beginning of the true repeat. This oligonucleotide is therefore present in n + 1 copies in an array of n repeats (Figure 4.2).

The repeated element itself is highly structured. A 10 bp inverted repeat with a 5 bp intervening sequence occurs 190 bp from the ACTTTAA start, and has the potential to form a stable hairpin. The position of this inverted repeat was mapped in an S1-nuclease experiment, suggesting that the sequence existed as an extruded cruciform and was therefore susceptible to cleavage in the ssDNA loops (LaRoche *et al.* 1990). The inverted repeat is flanked on one side by a G + C-rich region (70% over 110 bp) which could act to stabilize an extruded cruciform. The element becomes A + T-rich beyond this (80% over 550 bp) and includes periodic short homopolymer runs of A and T, a structure that has been associated with conformational changes in DNA (Marini *et al.* 1982, Wright and Dixon, 1988). The location

of the A + T-rich region of the repeat relative to the ends of the cleaved fragment was shown to influence fragment mobility in polyacrylamide gels (LaRoche *et al.* 1990).

Restriction digestion of the 280 individuals in the sample with Stu I, which cleaves once within the repeat sequence, confirmed that the repeats are in direct orientation within the array. This sample represents more than 1000 copies of the repeat, and only a single example of an incomplete repeat was noted. This exception occurred in an individual with three repeat copies, one of which was truncated to about 1300 bp. Two examples of repeat site polymorphism were observed. In one case, the internal one of three copies had lost a Stu I site; in the other, an internal one of four copies had gained a Sal I site.

Because the array comprises complete copies of the repeat, it is possible to derive the copy number from the length of Eco RI-B (Table 4.3a). Seven size classes for this fragment were observed, representing two to eight copies of the element with an absolute length difference of 8.7 kb (Figure 4.3). In relative terms, one genome may be 27% larger than another due to differences at this locus alone. The presence of two complete repeats in the smallest size class was confirmed by double digestion with Eco RI and Pst I, which produced a repeat-sized fragment and the predicted sizes of the three other fragments making up Eco RI-B. The larger size classes lie outside the linear range of the gel as Eco RI-B fragments. The correct assignment of these variant fragments was confirmed in the second assay, with Sph I and Sal I, where the marker fragment is 3.4 kb shorter and except for size class 8 can be directly compared to the kilobase ladder size standard. However, classes 7 and 8 were clearly distinguished in size standardization gels.

Heteroplasmy for locus I size class was noted in 18 individuals (6.4%). Scoring was done as for Sph I heteroplasmy, where each variant was counted once. Thus the sample size is 299 genomes rather than 280 individuals. The distribution of size classes among individuals has a mode of 4 copies (Table 4.3b), with 53.2% of scallops sharing this class. The adjacent classes 3 and 5 have frequencies of 22.1% and 16.4% respectively. Together these three classes account for 91.7% of the variation.

Diversity at this locus was estimated using the K statistic of Birky *et al.* (1989), who measure K_T, the probability that two randomly chosen molecules from a species will be different, as:

$$K_T = 1 - \sum x_i^2$$

where x_i is the frequency of the i^{th} size class. This index gives a value of 0.6388 for locus I.

It is important to note that each size class appeared in homoplasmy at least once, and therefore must be viable. The converse is not true; size class 8 did not occur in heteroplasmy. This is the rarest class, seen in only 3 individuals. A discussion of heteroplasmy will be presented in the following chapter.

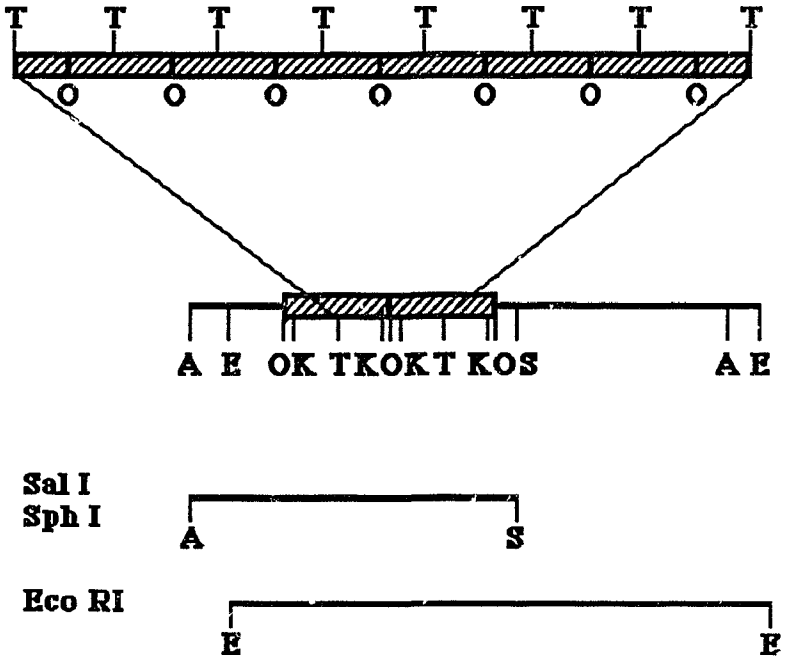


Figure 4.2 Organization of the marker fragment showing length polymorphism at locus I. The largest and smallest size classes are shown, with the 8 copy array expanded between the Stu I sites within the 2 copy array. The restriction map of the marker fragment encompasses both the Sal I/Sph I and the Eco RI fragments used to score copy number. **A**, Sal I; **E**, Eco RI; **K**, Kpn I; **T**, Stu I; **S**, Sph I; **O**, the oligonucleotide ACTTTCC that flanks each repeat copy.

Table 4.3a Restriction fragment sizes observed at locus I. Samples were scored for both Eco RI-B and Sal I-B up to Sph I-D. The copy number is inferred from fragment lengths.

fragment length		copy number
Eco RI	Sal I/ Sph I	
8.0	4.7	**2**
9.4	6.1	**3**
10.8	7.6	**4**
12.2	9.0	**5**
13.6	10.4	**6**
15.0	11.8	**7**
16.4	13.2	**8**

Table 4.3b Tabulation of repeat copy number at locus I. **N**, number of genomes; $\mathbf{x_i}$, frequency of size class.

	copy number						
	2	**3**	**4**	**5**	**6**	**7**	**8**
N	5	66	159	49	12	5	3
$\mathbf{x_i}$	.017	.221	.532	.164	.040	.017	.010

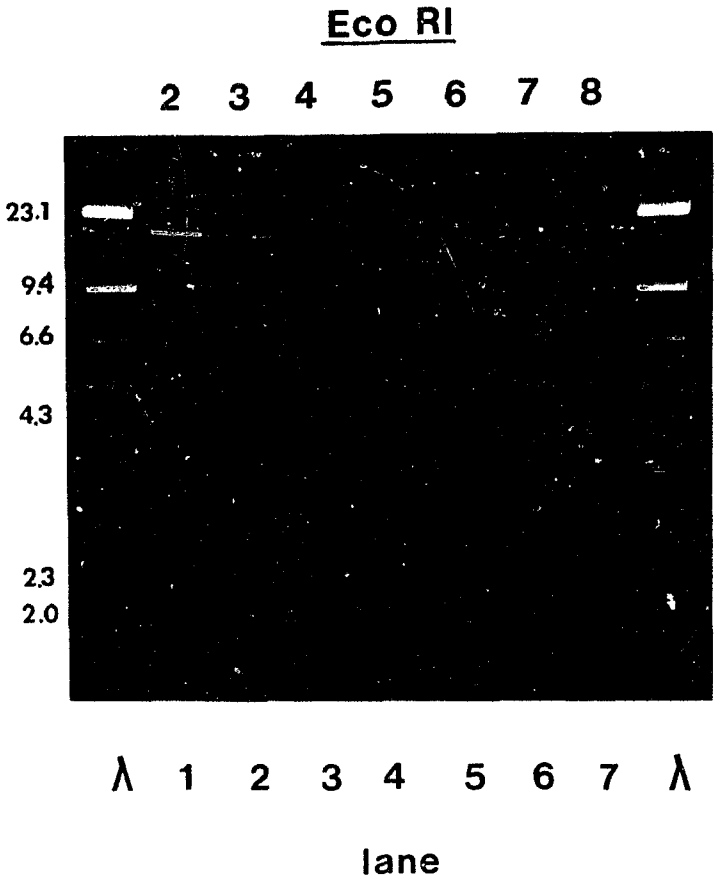


Figure 4.3 Locus I size classes. The locus I tandem array lies within the Eco RI-B fragment. Numbers on the left refer to λ Hind III fragment sizes. Numbers above the lanes denote locus I repeat copy number.

4.2.2.2 Locus II

In the survey for repeat copy number classification with Eco RI, the length of Eco RI-B was sometimes observed to vary within and among individuals by less than a full repeat increment. Because the most common length of Eco RI-B is 10.8 kb, differences in mobility were clear but not quantifiable in the gels. The possibility that the variation was due to the presence of one or more truncated or incomplete repeats was tested by digesting several variant genomes with both Eco RI and Pst I, which cleaves inside the repeat. These digests were analyzed on gels with mtDNAs known to have the common cleavage pattern to allow direct comparison of fragment mobilities.

In Eco RI/Pst I double digests, three fragments, in addition to the repeat, comprise the Eco RI-B fragment. All variation was confined to the 1.95 kb Eco RI-B/Pst I-B fragment overlapping the start of the repeat array, with the exception of the single individual mentioned above, where one internal repeat was truncated. The nature of the variation was such that digestion with Eco RI and either Pst I or Stu I produced variant bands that comigrated with other double digestion products. For this reason, individuals were scored by double digestion with Sal I and Stu I, where the Sal I-B/Stu I-D fragment spanning locus II is commonly 2.3 kb and easily separated from other bands.

The precise location of addition/deletion events at this locus is unknown, and the possibility that the 5' end of the initial repeat is involved cannot be excluded on the basis of these data. The combination of digests does show that the variation is confined to the 1.65 kb Eco RI-B/Stu I-D fragment (Figure 4.1b). The 7 bp oligonucleotide that marks the beginning

of the repeat array (LaRoche *et al.* 1990) is 1.0 kb downstream from the Eco RI site and 0.65 kb upstream from the Stu I site. Deletions larger than 0.65 kb must include some sequence outside the repeat, but other changes could be contained entirely inside or outside the repeat sequence.

Six discrete length alleles were observed in the Sal I/Stu I survey (Figures 4.4 and 4.5). The Sal I-B/Stu I-D fragment has a range of 1.55 to 2.8 kb, and the alleles have been given letter codes (Table 4.4a). Four of the variants occurred more than once, allowing fragment mobilities to be compared among individuals. Two alleles, A and E, were only seen in heteroplasmy, and two others, C and F, were only seen in homoplasmy. B and D occurred in both states. The apparently incremental differences between alleles suggests that there may be a repeated sequence of about 250 bp at locus II. This is also supported by the observation of heteroplasmy for discrete alleles, which is frequently associated with repeat arrays in animal mtDNAs (Moritz *et al.* 1987).

As with locus I, heteroplasmics were scored by counting each variant once, and the sample size is 272 genomes rather than 261 individuals. The distribution of fragment lengths reveals a much lower diversity than at locus I (Table 4.4b). Of 272 genomes scored, 242 were class D, a frequency of 0.890. Class B, 500 bp smaller, represented 0.088 of genomes. The remaining alleles were very rare, A and F being observed twice (0.0074), and C and E only once (0.0037). The locus II diversity index, using the K statistic, is 0.1899.

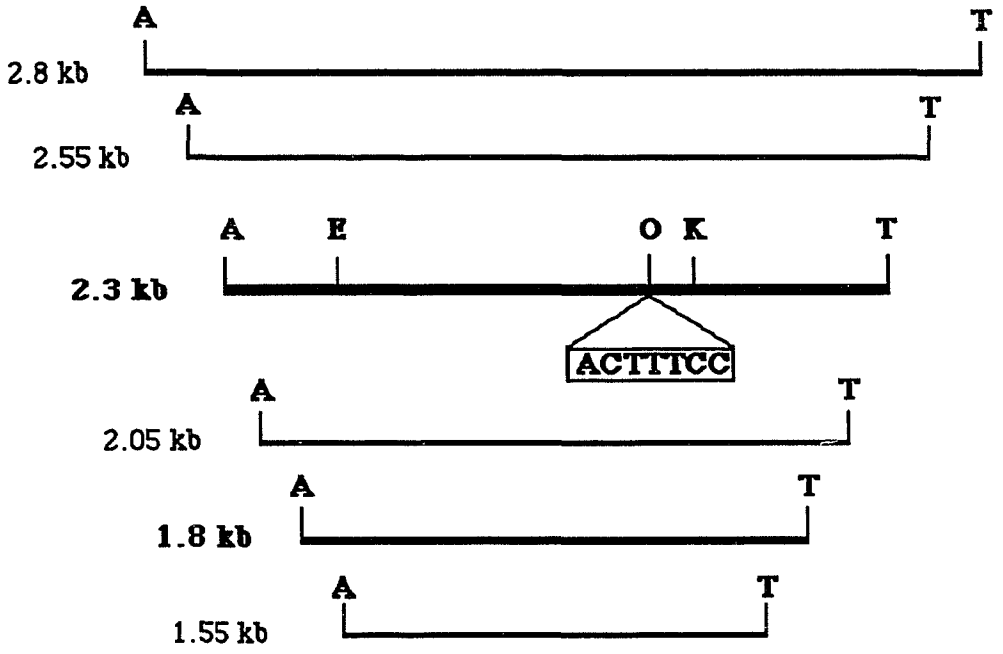


Figure 4.4 Organization of the marker fragment showing length polymorphism at locus II. The two common size alleles are shown in bold. Numbers on the left are fragment lengths for the Sal I / Stu I marker fragment. **A**, Sal I; **E**, Eco RI; **K**, Kpn I; **T**, Stu I; **O**, oligonucleotide ACTTTCC (boxed) signalling the start of the locus I repeat array.

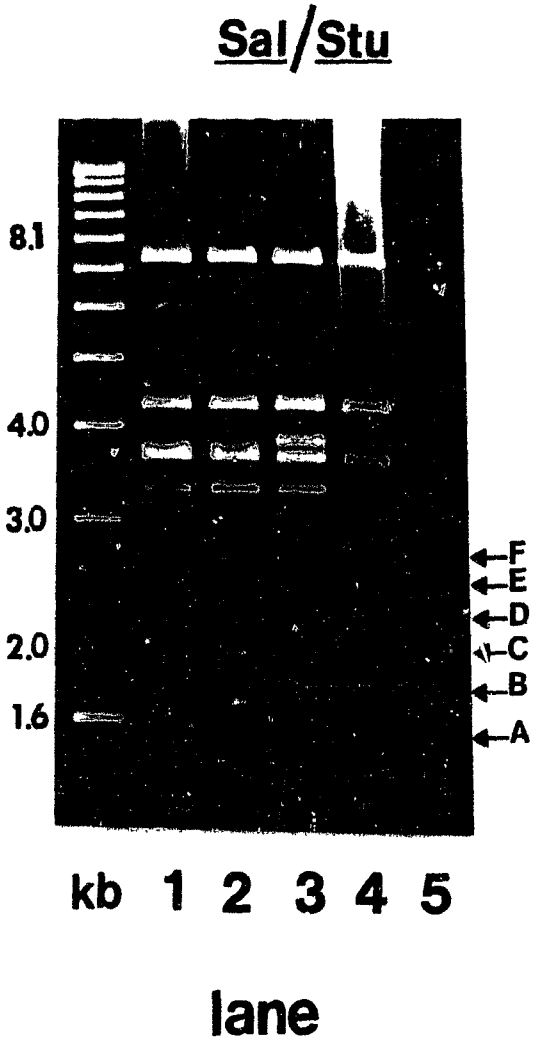


Figure 4.5 Locus II size alleles. Letters on the right designate alleles, numbers on the left refer to fragments in the kilobase ladder size standard (kb).

Table 4.4a Restriction fragment sizes at locus II.

fragment length Sal I-B / Stu I-D	allele
1.55	A
1.80	B
2.05	C
2.30	D
2.55	E
2.80	F

Table 4.4b Distribution of size alleles at locus II. **N**, number of genomes; x_i, frequency of allele.

	allele					
	A	B	C	D	E	F
N	2	24	1	242	1	2
x_i	.007	.088	.004	.890	.004	.007

4.2.2.3. Locus III

The region of the genome exhibiting the largest number of size alleles within and among individuals lies almost directly opposite the tandem array at locus I (Figure 4.1b). The lengths of fragments spanning locus III differ by as much as 1.1 kb. The variation is contained within the Hpa I-B/Stu I-C fragment, which places the boundaries about a region with a modal size of only 1.4 kb.

The frequency of heteroplasmy at locus III is exceptional. Because several variant bands were often observed in single individuals, it was necessary to choose for a survey a restriction cleavage pattern that would preclude ambiguity in the assignment of observed fragments to the locus. The total expected number of fragments was unpredictable, due to heteroplasmy, and therefore a pattern that produced no other bands close to the upper and lower limits of the variable fragment was required to obviate confusion due to the co-migration of any other fragments, as well as to permit discrimination of faint bands representing a small proportion of the variation. Although separation of bands differing in length by less than 100 bp is clearest when fragments are small (under 2 kb), the combination of Sal I and Sph I was chosen for the survey because it best met the above requirements despite having a modal fragment size at locus III of 4.2 kb (Figure 4.6). The nearest invariant bands in Sal I/Sph I double digests are 6.8 kb and 3.4 kb, outside the range of the polymorphic fragment (Figure 4.7). The only overlap of fragments occurs in genomes at the low end of the range for both locus I and II, where Sal I-B/Sph I-A can be 4.45 or 4.7 kb.

The digests were subjected to electrophoresis at 30 V for 18 h in 1.0% agarose, conditions chosen to optimize both migration and resolution. Bands

were sharp and could be discriminated, but frequently the differences in mobility were small, corresponding to length differences of about 50 bp. Patterns of bands in heteroplasmic individuals, where as many as seven bands could be seen, suggest that this is a kind of discrete size variation. Smears, which would imply continuous variation, were not observed. The relative intensity of locus III bands within individuals varied greatly (see Figure 5.3), and the possibility that some individuals included variants below the limit of detection cannot be excluded.

The large number of possible states at this locus, together with the small differences between them, made definitive identification impossible. Minor mobility differences across lanes within and among gels due to experimental error could be confused with molecular differences. For this reason no attempt was made to classify alleles incrementally. Rather, alleles were lumped into categories (Table 4.5a). Each individual was scored for the number of alleles within each category, and in multiplasmic individuals as many as four alleles of a single category could be discriminated. Examples of homoplasmic and heteroplasmic individuals showing the three categories S, M, and L, can be seen in Figures 4.7 and 5.3.

In the survey of 272 scallops, 537 locus III variant bands were scored (Table 4.5b). Each band present in an individual digest was assigned to one of the three size categories, and all bands scored were given equal weight Neither different alleles within categories nor different frequencies of alleles within individuals could be recorded. Given these conditions, comparisons with the other loci are limited in precision and scope but it was clear that locus III resembled both locus I and locus II in that the intermediate size

predominates. Category M has a frequency of 51%, S has a frequency of 26.3%, and L has a frequency of 22.7%.

Individuals were also scored for their composite locus III genotype, with each pattern of numbers and categories of alleles being assigned a letter. I propose here the term "mitotype" to designate the complete description of mitochondrial genotypes for an individual at one or more loci, wherein heteroplasmy is subsumed within the mitotype rather than given as a combination of two or more genotypes. This kind of description interposes another level of perception of individual differences and seems particularly suited to a locus where intra-individual variation is common. However, it should be noted that the grouping into categories of alleles that could be resolved with confidence within but not among individual digests limited the discrimination power of the assay, as many individuals were assigned to the same mitotype despite obvious differences in size alleles. For example, in the common mitotype describing individuals heteroplasmic for two M category variants, many individuals had different combinations of M category size alleles.

There are twenty-six mitotypes at locus III, from A (S) to Z (SSMMMML). Frequencies range from 19.5% to 0.4%, with seven mitotypes being observed only once (Table 4.6). A diversity estimate for this locus was calculated using Nei and Tajima's (1981) H statistic as in section 4.2.1, which gave a value of 0.9081.

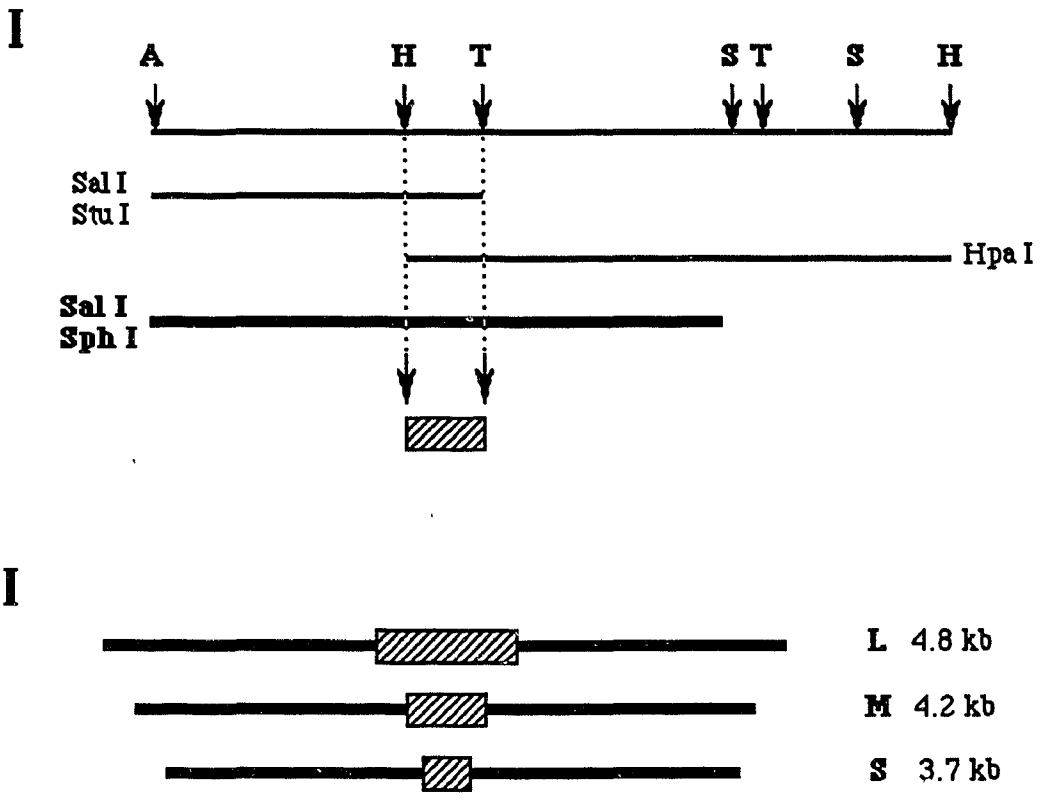


Figure 4.6 Organization of the marker fragment showing length polymorphism at locus III. In **I**, the restriction map of the region is on top and the digests that localize the variation are underneath, with the marker Sal I/Sph I fragment in bold. **A**, Sal I; **H**, Hpa I; **T**, Stu I; **S**, Sph I. In **II**, the three size categories for locus III are illustrated with a shaded box representing the variable region.

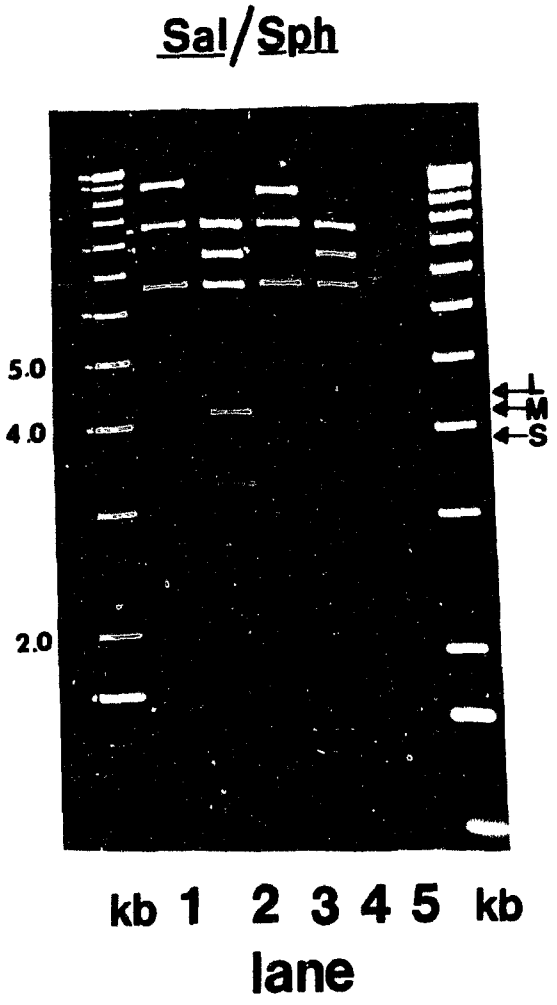


Figure 4.7 Locus III size categories. Letters on the right designate categories, numbers on the left refer to fragments in the kilobase ladder size standard (kb).

Table 4.5a Restriction fragment sizes at locus III.

fragment length Sal I-A / Sph I-C	category
3.7 - 3.95	S
4.0 - 4.3	M
4.35 - 4.8	L

Table 4.5b Distribution of size categories at locus III. **N**, number of genomes; x_i, frequency of category.

	category		
	S	M	L
N	141	274	122
x_i	.263	.510	.227

Table 4.6 Tabulation of mitotypes at locus III. Category components of mitotypes are given in parentheses. **N**, number of individuals; x_j, frequency of mitotype.

mitotype	N	x_j
A (S)	21	0.077
B (M)	53	0.195
C (L)	11	0.040
D (SS)	15	0.055
E (SM)	27	0.099
F (SL)	4	0.015
G (MM)	39	0.143
H (ML)	21	0.077
I (LL)	18	0.066
J (SSS)	1	0.004
K (SSM)	9	0.033
L (SSL)	1	0.004
M (SMM)	12	0.044
N (SML)	7	0.026
O (MMM)	4	0.015
P (MML)	8	0.029
Q (MLL)	7	0.026
R (LLL)	3	0.011
S (SSSM)	2	0.007
T (SSMM)	2	0.007
U (SMML)	1	0.004
V (MLLL)	1	0.004
W (MMLL)	2	0.007
X (SMLL)	1	0.004
Y (SSSMM)	1	0.004
Z (SSMMMML)	1	0.004

4.2.2.4 Other regions

Deletions of less than 200 bp were noted in one of three contiguous fragments within the Sal I-C fragment in eight individuals (Fig 4.1b), all homoplasmic for the deletion. A single instance of deletion within the Stu I-E fragment was observed. In four cases a deletion of about 200 bp was seen in the Sal I-C/Stu I-B fragment, but no attempt was made to ascertain the identity of the deleted sequence or its precise location. Finally, in three individuals a 150 bp deletion occurred within the Sal I-C/Sph I-C fragment. Again, its sequence and location were not determined. None of these deletions is nearer than 2 kb to one of the major polymorphic regions.

A number of cross-hybridizations have been carried out (LaRoche *et al.* 1990), with the results shown in Figure 4.8. The pP1.4 clone, containing the complete repeat sequence, has been shown to hybridize to Pst I-C, the repeat fragment, as well as to Eco RI-B/Pst I-B and to Pst I-D, which contain 936 bp and 513 bp of repeat sequence respectively. It also hybridizes to Sph I-C/Pst I-A, across the molecule from the repeat array. A deletion subclone of the pP0.7 clone derived from the Pst I-D fragment, designated pX1, was sequenced and shown to contain fewer than 10 bp of repeat sequence (J. LaRoche, pers. comm.). This clone, representing the 3' flank of the tandem array, hybridizes to Pst I-D, from which it was derived, as well as to Sph I-C/Pst I-A. It does not hybridize to the 5' flanking sequence in Eco RI-B/Pst I-B (LaRoche *et al.* 1990).

A clone of the 5' flanking fragment, extending from the Eco RI-B site to a Kpn I site 95 bp 3' to the 7 bp oligonucleotide that marks the beginning of the repeat array, was also used as a probe and designated pEK. This clone

contains 95 bp of repeat sequence and about 1,000 bp of locus II sequence, and it showed a complex hybridization pattern (D. Cook, Dalhousie University, pers. comm.). As expected, it hybridized to Pst I-C but not to Pst I-D. It also hybridized to Sph I-C/Sal I-C (overlapping Sph I-C/Pst I-A), which as noted above shares sequence similarity with the repeat. In addition, it hybridized to two other fragments within Sal I-C (Stu I-E and Sal I-C/Stu I-B, both of which have exhibited length polymorphism), and showed a strong reaction with the Sph I-C/Sal I-A locus III fragment. It also hybridized weakly to Sph I-E.

Because hybridization to pP1.4 outside the repeat array has been shown to be restricted to the 2.0 kb Sph I-C/Sal I-C fragment (LaRoche *et al.* 1990), the pEK hybridization pattern suggests that sequence similarity with some part of locus II exclusive of repeat sequence is dispersed within a 16 kb region. Within this region, five of the hybridizing fragments also exhibit size variation.

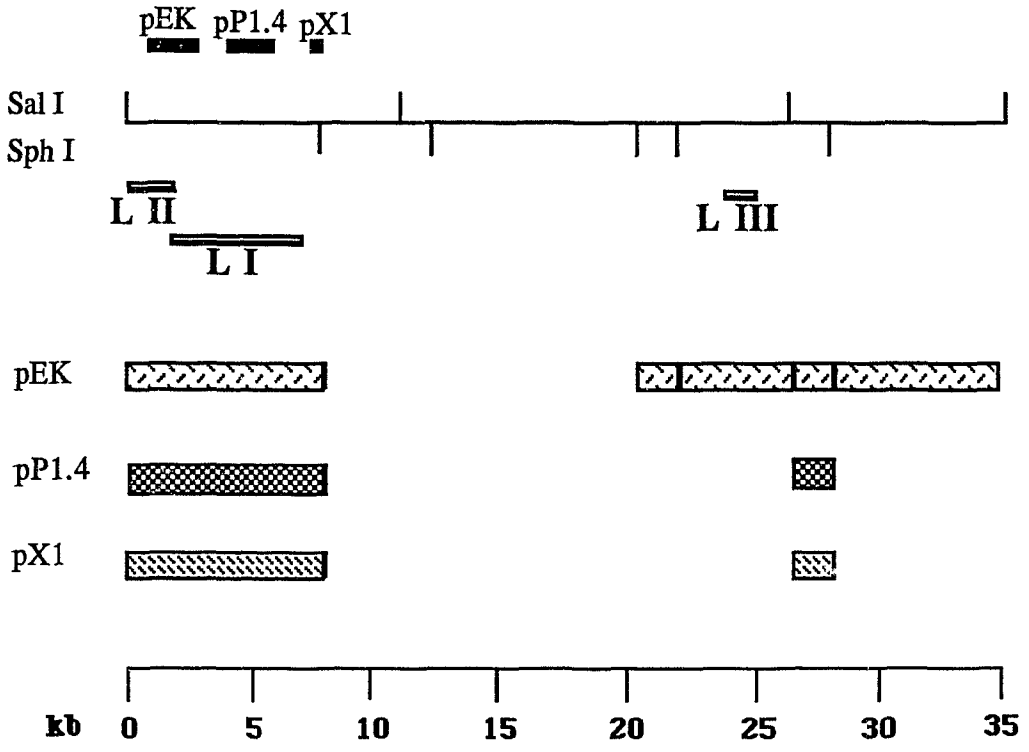


Figure 4.8 Hybridization patterns. The plasmids used as probes (see text) are indicated as black boxes above the Sal I/Sph I map. The fragments to which they hybridize are indicated below the linear map and above the distance line. The positions of the major length polymorphic loci are denoted as LI, LII, and LIII. Adapted from LaRoche *et al.* 1990.

4.3 Discussion

4.3.1 Site polymorphism

The nucleotide diversity of 0.17% in *Placopecten magellanicus* is low, but the estimate must be viewed with caution. The calculation was based on genomes normalized to two repeat copies and only four restriction enzymes were used in the survey, which results in a large variance of the estimate. The sample of 126 bp represents less than 0.4% of the genome. Since *P. magellanicus* is usually found in dense aggregations, a higher diversity would be expected based on population size. The low estimate could reflect sampling bias either due to the enzymes used or to non-random sampling of a highly subdivided population. It could also reflect a relatively recent origin of the Atlantic Canada *P. magellanicus* population, or a relatively recent change in the abundance and distribution of the species. The ice and current patterns during the last glaciation produced very different ocean temperature regimes all along the eastern coast of North America (Prest and Grant 1969, McIntyre *et al.* 1976, Cronin *et al.* 1981). The path of the Gulf Stream swung abruptly east off Cape Hatteras, resulting in only a narrow band of cold temperate waters between the northern glacial and southern warm waters that would have been appropriate habitat for *P. magellanicus* . This likely severely restricted the range of the species. According to Prest and Grant (1969), the sea did not reinvade the Bay of Fundy and Gulf of St. Lawrence until between 14,000 and 13,000 years BP. Post-glacial colonization may have begun as recently as 13,000 years BP.

Nine composite restriction morphs were found, and the clonal diversity of 0.365 resulted primarily from the high frequency of polymorphism for one enzyme, Sph I.

Two site mutations were found in the long tandem repeat. With no open reading frame of significant size this sequence was assumed to be non-coding (LaRoche *et al.* 1990), and might be expected to evolve rapidly as the non-coding control region does. In the sample of over 1000 repeat copies only two changes and one small deletion were observed, in each case in single members of an array. Sequence data from *P. magellanicus* indicated little difference between repeats from different arrays, only two single base deletions being found, but offered little information about repeats within arrays. The sequences of two complete arrays in *Gryllus* showed that one is homogeneous and the other has two site differences between two repeats (Rand and Harrison 1989). Concerted evolution in mitochondrial repeats should result in homogeneous arrays within lineages and heterogeneity among lineages. Extrapolating from restriction site data, Solignac *et al.* (1986) presented evidence for concerted evolution at least at the species level in *Drosophila* . The two and three repeat arrays in the shad are homogeneous with respect to presence or absence of a Kpn I site and may be evolving in concert (Bentzen *et al.* 1988). The repeat arrays in *P. magellanicus* appear to show remarkable homogeneity as a whole, but more data are needed on sequence divergence in various other regions of the molecule and more markers are needed to test divergence within and among arrays before the question of concerted evolution of repeats can be resolved.

4.3.2 Length polymorphism

All of the length polymorphism in *Placopecten magellanicus* has been localized and defined either as 150-200 bp deletions or as incremental variation at one or more of three loci. The length polymorphism cannot

explain the expanded size of the genome, and distinct mechanisms or events must be postulated to account for the two phenomena of enlargement and polymorphism. This contrasts with all other size-variable mitochondrial genomes with the exception of the pine weevil (Boyce et al. 1989), where a generalized non-specific expansion of the A+T-rich region mapping to the control region occurs adjacent to a length-variable repeat array. Although a similar arrangement has not been ruled out for the scallop, the patterns of restriction cleavage do not indicate that one segment of the scallop mtDNA molecule is distinctive in its base composition. The coexistence of both an expanded genome and several length-polymorphic loci implies that the evolutionary forces acting on size variation in the scallop operate in a background different from that of other animals.

The eight cases of small deletions were each homoplasmic, indicating that the sequences lost are not required for mitochondrial function. The three restriction fragments containing deletions are contiguous and positioned between two polymorphic loci. Because only one of the fragments showed a deletion in any single animal it is likely that each deleted sequence is discrete and separate. From cross hybridization it is known that one fragment, Sal I-C/Sph I-C, shows sequence similarity with part of the locus I repeat and all three show sequence similarity with some part of locus II sequence. This suggests that the deleted sequences may be dispersed duplications, but whether the loss is compensated by the presence of a functional sequence elsewhere or encompasses a non-functional sequence cannot be determined from these data.

The repeat unit of locus I is large, 1.45 kb, and highly structured (LaRoche *et al.* 1990). With the exception of a single animal showing a small

deletion in one of three repeats, repeat size was constant throughout the sample. Few tandemly repeated sequences this large have been reported in animal mtDNA, and none has shown this degree of conservation of length or arrangement. The 220 bp repeat in crickets is constant in size but organized within the array in such a way that the flanks constitute an incomplete copy, missing 26 bp (Rand and Harrison 1989).

A peculiar feature of the polymorphic tandem arrays described in animal mtDNA is the generality of a minimal array size of two. In shad and *R. culicivorax*, arrays are all either two or three copies, while in *Gryllus* and *P. magellanicus* the number of size classes is large and the smallest class of two copies is rare (0.2% and 2.0% respectively). One cricket contained a single copy array in heteroplasmy with two and three copy genomes, but no genomes with array size one were seen in the scallop. More data are needed to clarify this point, but it appears that either there is no mechanism for reducing an array to a single copy or that once generated a single copy genome is rapidly eliminated.

The two mechanisms that have been invoked to explain copy number variation, recombination (Rand and Harrison 1989) and replication slippage (Efstratiadis et al. 1980, Levinson and Gutman 1987), have been incorporated into models to study the persistence of tandem arrays by Walsh (1987), who showed that in the absence of selection or gene amplification all arrays reach size one in finite time. Intermolecular recombination redistributes repeats and unequal crossing over may create new array sizes, but not new repeat copies, while intramolecular recombination produces only deletions, removing copies from the population. Amplification by slippage or other processes creates new variation, although slippage can

result in both duplication and deletion with the single restriction that arrays of size one may not undergo slippage. The equilibrium distribution of array sizes depends upon the relative rates of slippage, duplication and deletion, and inter- and intramolecular recombination, but there is no restriction on attaining size one.

The locus I tandem array in *P. magellanicus* contains from 2-8 copies and shows nearly equal frequencies of arrays larger and smaller than the mode of 4, although there are twice as many size classes larger than smaller. The frequency of size class 4 is 53.2%. Size class 3 has a frequency of 22.1%, with a sharp drop to size class 2 at 2%. The drop in frequency beyond class 5 was less pronounced. The maximum array size of 8 represents the greatest amplification reported in any mitochondrial tandem array, both in copy number and in absolute length. It is the rarest class, at 1%, and is the only class not seen at least once in heteroplasmy. This contrasts with the cricket system, where size class VL (6 copies) also has a frequency of 1.2% but 5 of 6 VL genomes occurred in heteroplasmy, while the largest class X (7 copies) was noted only twice, and both times in triplasmic individuals. *P. magellanicus* shares a modal array size of four with *Gryllus*, as well as a high diversity index (0.64 for *P. magellanicus* , 0.48 for *Gryllus*).

At present no function can be assigned to the locus I repeat, nor can it be localized to any functional region. The cytochrome oxidase genes lie within 3.5 kb of one end of the array, and the large rRNA gene is 10 kb distant. If this array is similar to others in being associated with the control region, the gene order in *Placopecten magellanicus* must be different from that in other species. Although certain features of the repeat sequence

resemble elements typical of control regions, such as the cruciform and the A+T rich region, the short direct repeat that imposes a higher order structure on the array is unique to *Placopecten* and a similar motif has not been described in any animal control region. Further, the presence of a large portion of the repeated sequence in another region of the molecule 9 kb distant, as suggested by cross hybridization, argues against a control region hypothesis. Duplicated and widely separated origins of replication could seriously impair replication efficiency.

Locus II exhibited six size classes, differing by increments of 250 bp. The observation of discrete variants, alone and in heteroplasmy, strongly supports the inference of a tandem array. This can only be tested with sequence data or high resolution restriction mapping.

The restriction fragments used to score variation at locus II included some overlap with locus I, and therefore the possibility exists that at least some locus II size classes result from changes occurring in the terminal repeat portion of locus I. However, there is nothing in the sequence of the locus I repeat to suggest any mechanism for precise deletions of 250 and 500 bp, and the smallest locus II size class would have to include some flanking DNA as well as locus I repeat sequence in the deletion. In addition, the dispersed sequence similarity with locus II revealed by cross hybridization suggests that it contains an element associated with length variation which is distinct from the repeat sequence and present in multiple copies in the genome.

Without sequence data the putative locus II repeat size and copy number cannot be deduced, but the relative array sizes show certain similarities with locus I and other systems. The most common size is three

increments larger than the smallest observed, but the frequency distribution is skewed toward a higher frequency of genomes smaller than the mode rather than larger. The smallest class was observed twice, but both times in heteroplasmy; it would be of interest to know whether this represents an array of size one. However, despite the large number of size classes the diversity at locus II is low, only 0.19. Almost 90% of individuals share class D, and four classes occur in fewer than 1% of individuals. These observations point to a low rate of generation of variation, and emphasize even more strongly the question of why most individuals possess an amplified array rather than simple duplication. Where fixed differences in copy number distinguish closely related species, as in the *D. melanogaster* group (Solignac *et al.* 1986), it can be assumed that changes in array size are very rare, but when intraspecific polymorphism and heteroplasmy occur it is good evidence for relatively frequent mutation.

Of the three major length polymorphic regions in *P. magellanicus* , locus III is by far the most variable. Situated across the molecule from the locus I array, it can be localized to a small Hpa I/Stu I fragment with a consensus size of 1.4 kb. The range of variation is about 1.1 kb and the number of size classes is probably greater than 15. The method used to score variation limited resolution to at best 50 bp differences, and allelic comparisons between individuals were not possible, but the appearance of bands as sharp and discrete suggests that this type of variation may also be due to a tandem array of varying copy number as was described for the 64 bp repeat in *Cnemidophorus* lizards (Densmore *et al.* 1985) and the 82 bp repeat in the sturgeon (Buroker *et al.* 1990). Heteroplasmy by itself does not indicate discrete variation, but heteroplasmic systems with continuous

variation typically show "fuzzy" bands in gels, as has been described in *R. esculenta.* (Monnerot *et al.* 1984). In *P. magellanicus* locus III fragments always produced sharp bands even when three or four variants were present in a single individual.

The small size of the increment of variation presented a number of difficulties. With no restriction site in the putative repeat its true size could not be determined nor could a densitometric scan to estimate copy number be obtained. I chose to lump variation into three categories with the boundaries designated according to the sizes of marker fragments; 4 kb as the lower limit of category M corresponds to the 4 kb fragment of the BRL kilobase ladder and 4.3 kb as the upper limit of category M corresponds to the 4.3 kb fragment of a lambda Hind III digest. The assignment of mitotypes according to allele number and category made possible a conservative estimate of diversity, which despite its conservatism was strikingly high.

As with the other two loci, variants in the middle of the range were most frequent. Despite containing less than 30% of the observed range, category M had a frequency of 51%. This pattern is very similar to length polymorphic loci in other species, as well as to locus I and II in the scallop. Moreover, the frequency of heteroplasmy and the size and apparent conservatism of the increment of variation closely resemble descriptions of control region tandem arrays in lizards (Densmore *et al.* 1985), crickets (Rand and Harrison 1989), and sturgeon (Buroker *et al.* 1990).

Chapter 5

Heteroplasmy

5.1 Introduction

Organelle genomes that show uniparental inheritance also tend to show uniformity within individuals. In animal systems, maternal inheritance of mtDNA is assumed to be universal and strict based on the evidence from natural experiments of hybridogenetic species (Spolsky and Uzzell 1986, Avise and Vrijenhoek 1987) and from laboratory experiments involving repeated backcrosses between different lines (Lansman *et al.* 1983). Early studies of animal mtDNA polymorphism were primarily comparisons of restriction cleavage patterns in mammalian species, and although extensive variation among female lineages was found, individuals were homoplasmic for mitotype (Avise and Lansman 1983). Restriction site polymorphism is a consequence of nucleotide substitution and intra-individual variation must occur in the period between mutation and fixation or loss of the variant, but detection of this is unlikely for two reasons. First, point mutations are infrequent so that the waiting time to fixation or loss appears transient relative to the time between mutations. Second, the investigator must sample the correct sequence to uncover the polymorphism, and very few studies have sampled even 10% of the genome.

The discovery of intra-specific size polymorphism altered the view of individual homogeneity. As more species were characterized, particularly invertebrates and poikilothermic vertebrates, examples of length variation were found to be relatively common (Solignac *et al.* 1983, Monnerot *et al.* 1984, Densmore *et al.* 1985, Harrison *et al.* 1985, Kessler and Avise 1985, Bermingham *et al.* 1986, Hale and Singh 1986, Moritz and Brown 1986,

Powers *et al.* 1986, Wallis 1987, Bentzen *et al.* 1988, Avise *et al.* 1989, Boyce *et al.* 1989, Gach and Reimchen 1989, Buroker *et al.* 1990, LaRoche *et al.* 1990). The frequency of heteroplasmy in these systems varies widely, from 0.1% in newts (Wallis 1987) to 100% in *Rana esculenta* (Monnerot *et al.* 1984) and in pine weevils (Boyce *et al.* 1989). In general, the higher rate of length mutation makes detection of length heteroplasmy much more likely than site heteroplasmy. Only four cases of site heteroplasmy have been described, in cows (Hauswirth and Laipis 1985), shad (Bentzen *et al.* 1988), *Drosophila* (Hale and Singh 1986), and Icelandic scallops (Gjetvaj 1989).

Length heteroplasmy may result from either deletion or addition. Deletions have been described in humans (Holt *et al.* 1988a,1988b,1989, Schon *et al.* 1989, Wallace 1989) and mice (Boursot *et al.* 1987), where loss of coding sequences results in a non-viable genome that must be complemented by undeleted mtDNA, making heteroplasmy obligate. Additions are generally either tandem duplications or tandem arrays of repeats, although some examples of dispersed and continuous size variation are known, and heteroplasmy may be common or rare.

The theory of heteroplasmy, developed by Clark (1988) and Birky *et al.* (1989), predicts that the frequency of heteroplasmy is a function of mutation rate, drift, selection, and paternal contribution. With paternal leakage assumed to be absent and scant evidence for selection, mutation and drift remain as the major factors affecting levels of heteroplasmy. The small mtDNA populations within organelles, typically 2 - 10 molecules (Bogenhagen and Clayton 1974), are likely rapidly homogenized by random differential replication and/or random partitioning at mitochondrial fission and the resultant mixed population of homogenous mitochondria is sorted

during mitochondrial turnover and cell division (Birky 1983). Two empirical studies of transmission in heteroplasmic female lines, in *Gryllus* (Rand and Harrison 1986) and *Drosophila* (Solignac *et al.* 1987), have shown that heteroplasmy may persist for as long as 500 generations, while a study of heteroplasmy in a bovine lineage (Hauswirth and Laipis 1985) suggests that both generation and segregation of variation can be very rapid, on the order of a few generations.

By definition, heteroplasmy simply means the coexistence of more than one mitotype in a single cell, although it is generally used to describe a whole organism. Unless heteroplasmy exists in the germ line however, it is evolutionarily trivial. Few studies have demonstrated explicitly that heteroplasmy is in the germ line, primarily due to sampling constraints. Eggs are not always available, nor are offspring. When somatic tissues are used singly or in combination as a source for mtDNA, the homoplasmic or heteroplasmic state of those tissues does not necessarily reflect the status of other cell lineages in the individual. The individual may be a mosaic of homoplasmic and heteroplasmic tissues or even of homoplasmic and heteroplasmic cells within tissues.

In *P. magellanicus* mtDNA length variation is dispersed and discrete, occurring at three major loci. One of these is known to be a tandem array of repeats, and the other two show several characteristics of a tandem array. The increments of variation and putative repeat size, and the diversity of variation are different for each locus, and this affords the unique opportunity to compare frequencies and patterns of heteroplasmy for three markers with the same developmental history. Genomes of each animal in this study were isolated from a single tissue, adductor muscle, and would therefore have been

subject to the same forces affecting variance, whether stochastic or deterministic.

A separate experiment was performed to confirm that heteroplasmy does exist in the *P. magellanicus* germ line. Although gonad data were not available for the large sample, a second sample was scored for heteroplasmy at two loci in three different tissues including gonad. These data allow me to assess the fate of variation under different segregation regimes within a single organism.

5.2 Results

5.2.1 Adductor muscle

5.2.1.1 Locus I

Eighteen scallops were identified as heteroplasmic at locus I, a frequency of 6.4%. Seventeen individuals were biplasmic and only one triplasmic (Table 5.1a). The preliminary screening of the 280 scallops in ethidium bromide-stained gels revealed seventeen biplasmic individuals, and a second screening by Southern blotting was undertaken to search for low frequency bands that might have escaped detection in stained gels. A test of resolution using mixed proportions of Hind III- and Eco RI-digested λ DNA showed that as little as 1% of one pattern could be detected through another with radio-labelling methods, but allowing for background a resolution limit of about 5% was more realistic. Only two new bands were detected by Southern blotting, one in an individual first classed as homoplasmic and another in a biplasmic individual. Both new bands had intra-individual frequencies at or under 10%.

Seven patterns of heteroplasmy, involving six of the seven size classes, were found (Figure 5.1). Ten individuals were heteroplasmic for the most common size classes, 3 and 4, and one was triplasmic for 2, 3, and 4. Among biplasmics, in three instances the genomes differed by two repeat increments, and in two instances they differed by three repeats (Table 5.1b). An examination of the distribution of heteroplasmic size alleles shows that the rare classes 2, 6, and 7 make a relatively small contribution (10.8%) to the pool of heteroplasmic genomes (Figure 5.2a), but in the context of the total

pool a greater proportion of these classes is represented in heteroplasmy than is the case for common alleles (Figure 5.2b).

The intra-individual frequencies of size classes were estimated by end-labelling and densitometry (Table 5.2). Frequencies were taken as the mean of values obtained from two separate labelling experiments. The mean standard error of the means was 0.025, with a standard deviation of 0.016. For one individual, background was too high to allow a reliable scan, and genomes were scored as simply more (+) and less (-) frequent. With the limit of resolution set at 5%, there is no pattern of clustering in the tails of the frequency distribution. The frequency ranges of the most common size classes are 0.102-0.893 (n=12) for 3, 0.107-0.898 (n=14) for 4, and 0.070-0.426 (n=5) for 5.

Excluding the triplasmic, the remaining individuals can be viewed as pairs of small and large fragments irrespective of size class, each fragment with an associated frequency of more or less than 50%. The small members of pairs tended to be also the less frequent members, but the trend was not significant (Fisher's Exact Test, p=0.072). Although the distribution of actual size classes between higher and lower frequency groups was heterogeneous (G=11.74, df=5, p=0.038), when rare size classes were lumped the p-value rose to 0.062.

The diversity within individuals was estimated as K_b, using the K statistics as in Chapter 4, letting x_i be the frequency of the i^{th} size class within the individual. K_b ranged from 0.1832 to 0.5538. The average $\overline{K}_b$, the probability of diversity within a randomly chosen individual in the population, is 0.0241.

Table 5.1a Frequency of heteroplasmy at locus I. **N**, number of individuals; x_i, frequency of plasmy level.

plasmy level	N	x_i
homoplasmic	262	.9357
biplasmic	17	.0607
triplasmic	1	.0036

Table 5.1b Patterns of heteroplasmy at locus I. Beside the number of individuals for each pattern is the pattern in parentheses; patterns are arranged as transition steps of copy number difference.

transition steps			
1	2	3	4
1 (2/3/4)		1 (2/5)	
10 (3/4)			
	2 (3/5)		
2 (4/5)			
	1 (4/6)		
		1 (4/7)	
13	3	2	0

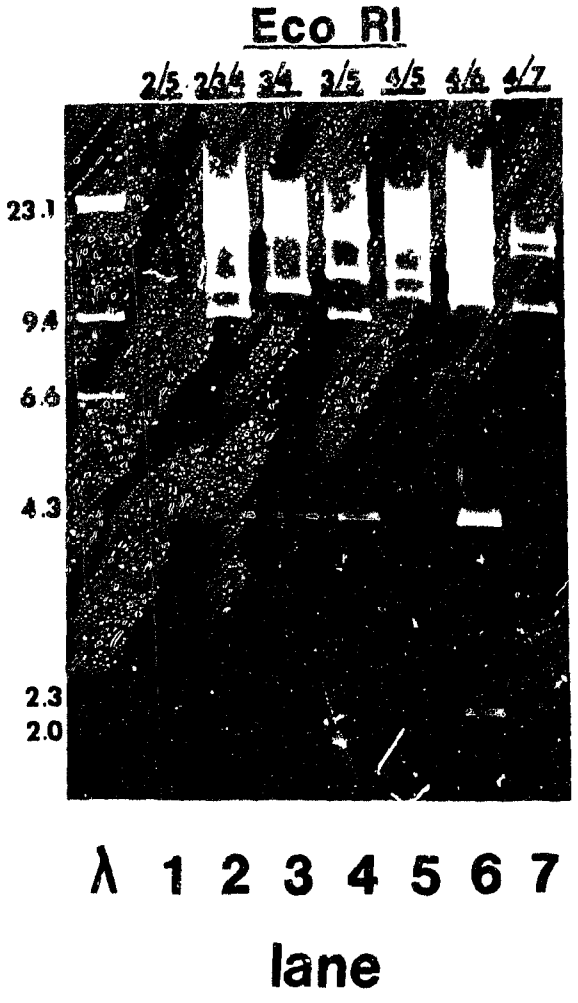
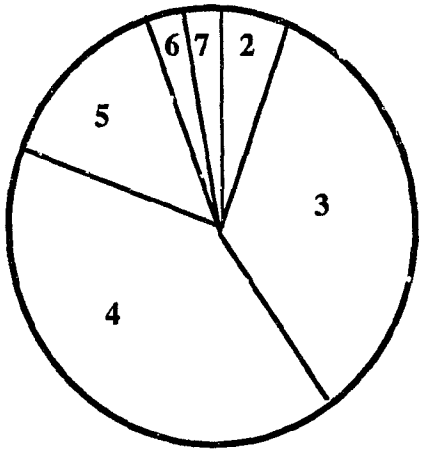


Figure 5.1 Locus I patterns of heteroplasmy. Numbers on the left refer to λ Hind III fragment sizes, numbers above the lanes denote locus I repeat copy numbers present in heteroplasmic individuals.



2	5.4%
3	35.1%
4	40.5%
5	13.5%
6	2.7%
7	2.7%

Figure 5.2a Distribution of heteroplasmic size alleles. The chart shows the relative proportions of the size classes in the pool of heteroplasmic fragments (N=37).

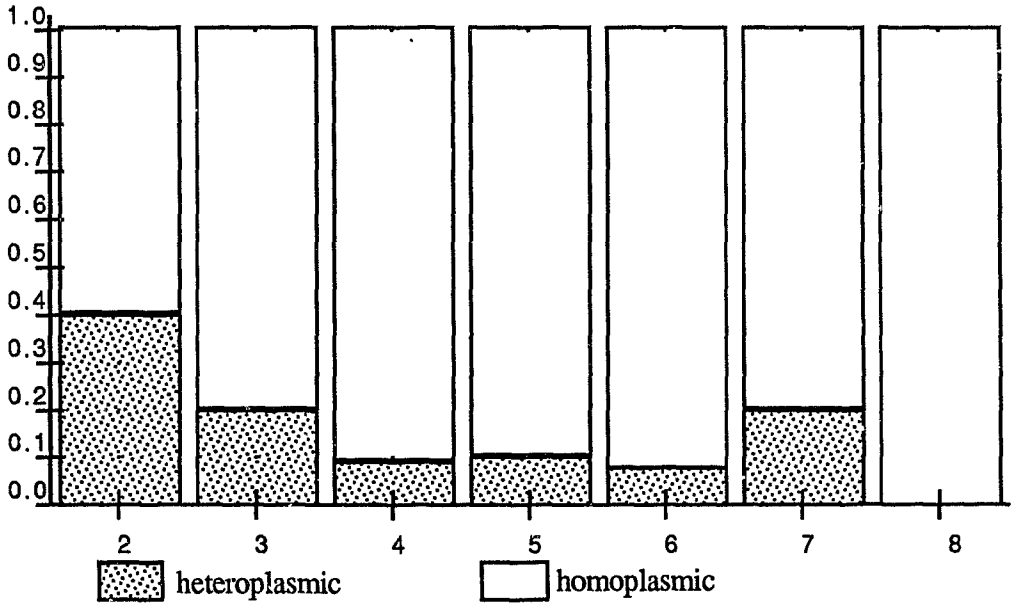


Figure 5.2b Relative distributions of size classes between homoplasmy and heteroplasmy.

Table 5.2 Frequency distributions of locus I size classes within heteroplasmic individuals; s_x, standard error of estimate; K_b, size class diversity within individuals.

animal	size class 2	3	4	5	6	7	s_x	K_b
C1	.443			.557			.017	.4936
C32			.856			.144	.023	.2466
C34		.297		.703			.035	.4176
H9		.246	.754				.010	.3710
H11		.235	.765				.022	.3596
H26[a]		.493		.507			.024	.5000
I24			.359	.641			.048	.4602
I32[b]		(+)	(–)				–	–
I33[c]		.230	.770				–	.3542
I48	.062	.434	.504				.009	.5538
I50		.102	.898				.056	.1832
N47			.243		.757		.017	.3680
G10[c]		.546	.454				–	.4958
G26[a]		.479	.521				.015	.4992
S7		.832	.168				.046	.2796
S8		.893	.107				.039	.1907
S23			.574	.426			.006	.4890
S31		.398	.602				.029	.4792

[a] sign could change

[b] background too high for estimate

[c] single estimate

5.2.1.2 Locus II

Only nine individuals were identified as heteroplasmic at locus II, and all were biplasmic. Three patterns were found, involving four size alleles (Table 5.3). Examples of heteroplasmy can be seen in Figure 4.5, Chapter 4. In contrast to locus I, all genomes heteroplasmic for this region differed by more than one size increment. The smallest allele, A, was observed only in heteroplasmy, once with a fragment of 2.3 kb and once with a fragment of 2.55 kb. The most common pattern, B/D, comprised fragments of 1.8 kb and 2.3 kb, which are also the most common alleles.

The occurrence of alleles in heteroplasmy relative to their frequency in the population is markedly different from locus I as well. Only the two common alleles, B and D, were seen in both heteroplasmy and homoplasmy; two others (A and E) were found only in heteroplasmy, and the remaining two (C and F) were seen only in homoplasmy. This distribution does not suggest any obvious mechanism for forward and back mutation among size alleles, but the rarity of four alleles indicates that a much larger sample is required if putative intermediates (homoplasmic A and E, heteroplasmic C and F) are to be found.

Table 5.3 Patterns of heteroplasmy at locus II. **N**, number of individuals.

type	N
A / E	1
A / D	1
B / D	7

5.2.1.3 Locus III

A large majority of scallops, 68.8%, were heteroplasmic at locus III. Only 85 individuals were homoplasmic; of the remaining 187, 63 were multiplasmic (Table 5.4). Genomes were scored as size categories, each of which contains several size alleles, so that a number of heteroplasmic combinations within and among categories were possible (Figure 5.3). A tabulation of categorical associations in heteroplasmic individuals revealed that within- and among-category variation was nearly equally frequent, with 42.8% of individuals showing heteroplasmy for size alleles contained within one of the S, M, or L categories and 57.2% showing heteroplasmy for size alleles from at least two categories (Table 5.5a). Categories can be broken down to show how frequently they are represented by single, double, triple, or quadruple alleles within individuals as a way of examining the distribution of heteroplasmy among categories (Table 5.5b). This distribution was homogeneous (G=4.209, p=.387). In about 53% of all individuals any category present was represented by only a single allele, and in the remaining 47% at least one category exhibited multiple alleles.

Because categories are artificial and in reality represent a continuum of size alleles, absolute differences within a category may exceed the difference between categories. For this reason, mitotypes at locus III are not directly comparable with either locus I or II, where mitotypes tend to differ by one and by more than one discrete size increment respectively. Perhaps the best index of comparison lies in the frequency with which fragments occur as single representatives of a category. Of 452 heteroplasmic fragments, 62.4% occur multiply within categories, suggesting that more

size mutations occur within than between categories. This is close to the value for single-increment heteroplasmy at locus I (70.6%).

In the 272 scallops scored, 23 heteroplasmic mitotypes were observed (Table 5.6). The frequency range for biplasmic types is 0.015-0.143, for triplasmic types 0.0037-0.044, and for mitotypes with four or more alleles 0.0037-0.0074. The latter are rare in the population; only 11 were observed. All but one biplasmic mitotype were more common than multiplasmics. Among bi- and triplasmics, the rarest mitotypes are F, J, L, O, and R. Both F and L describe the co-incidence of S and L categories, which are less frequent in the population generally, without the intermediate M. J, O, and R describe individuals with tri-allelic mitotypes solely within categories S, M, and L respectively. For 44 of 52 triplasmics, more than a single category was represented in the mitotype.

Table 5.4 Frequency of heteroplasmy at locus III. **N**, number of individuals; **p**, frequency of plasmy level.

plasmy level	N	p
homoplasmic	85	.312
biplasmic	124	.456
triplasmic	52	.191
tetraplasmic	9	.033
pentaplasmic	1	.004
septaplasmic	1	.004

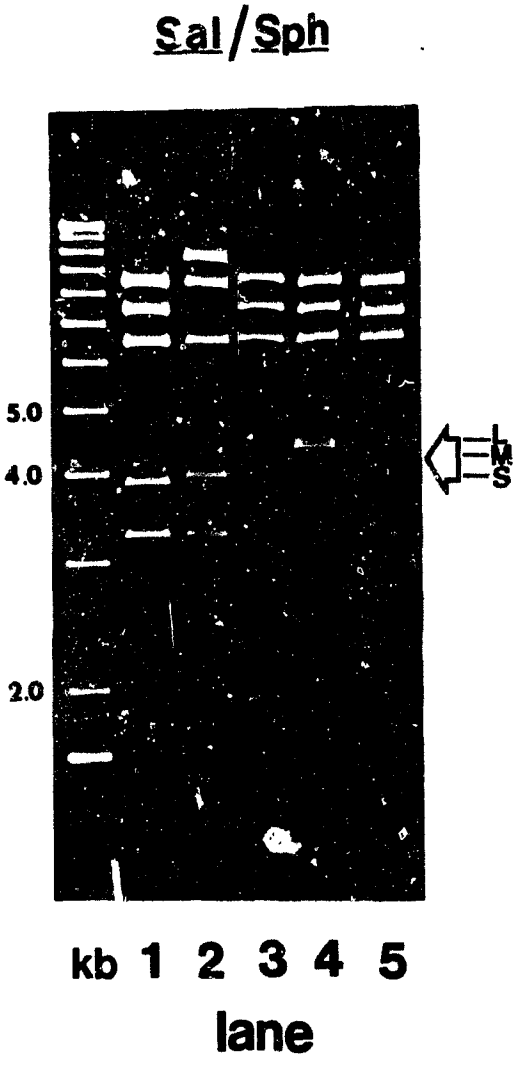


Figure 5.3 Locus III heteroplasmy. Within-category heteroplasmy can be seen in lanes 1 and 4, and between-category heteroplasmy in lanes 2, 3, and 5. Letters on the right designate categories, numbers on the left refer to fragments in the kilobase ladder size standard (kb).

Table 5.5a Distribution of size categories within locus III heteroplasmic individuals. N, number of individuals.

	size categories present	N
single category heteroplasmy	S+S	16
	M+M	43
	L+L	21
multiple category heteroplasmy	S+M	53
	S+L	5
	M+L	39
	S+M+L	10

Table 5.5b Distribution of size alleles within categories. The counts of single, double, triple, and quadruple alleles of each category are given for pooled heteroplasmic mitotypes.

number of alleles per category	size category S	M	L
1	73	128	54
2	56	130	56
3	12	12	12
4	0	4	0

Table 5.6 Patterns of heteroplasmy at locus III. Category components of mitotypes are given in parentheses. N, number of individuals; x_i, frequency of mitotype in the population; x_H, frequency of mitotype among heteroplasmic individuals.

mitotype	N	x_i	x_H
D (SS)	15	0.055	0.080
E (SM)	27	0.099	0.144
F (SL)	4	0.015	0.021
G (MM)	39	0.143	0.209
H (ML)	21	0.077	0.112
I (LL)	18	0.066	0.096
J (SSS)	1	0.004	0.005
K (SSM)	9	0.033	0.048
L (SSL)	1	0.004	0.005
M (SMM)	12	0.044	0.064
N (SML)	7	0.026	0.037
O (MMM)	4	0.015	0.021
P (MML)	8	0.029	0.043
Q (MLL)	7	0.026	0.037
R (LLL)	3	0.011	0.016
S (SSSM)	2	0.007	0.011
T (SSMM)	2	0.007	0.011
U (SMML)	1	0.004	0.005
V (MLLL)	1	0.004	0.005
W (MMLL)	2	0.007	0.011
X (SMLL)	1	0.004	0.005
Y (SSSMM)	1	0.004	0.005
Z (SSMMMML)	1	0.004	0.005

5.2.1.4 Restriction site heteroplasmy

Four individuals, 1.4%, exhibited restriction site heteroplasmy In each case the mtDNA population in the individual comprised both patterns A and B for Sph I. These patterns have frequencies of 0.807 and 0.104 respectively in the scallop population.

This type of heteroplasmy is the most difficult to score because genuine site loss in part of the mitochondrial population must be distinguished from incomplete digestion. Normally, a range of partial products will be evident following incomplete digestion, but cleavage rates are known to vary among sites due to the proximity of other sites or to the local sequence environment (Thomas and Davis 1975) so that complete cleavage can occur at some sites and only partial cleavage at others within a single digest. This is an unlikely explanation here, however. In 276 scallops, digested at least twice each with Sph I, this site alone resisted digestion in only four individuals, and reproducibly in those four. Further, the site itself is polymorphic in the population, providing as an alternate explanation the true absence of the recognition sequence.

This polymorphic site flanks the long tandem array, being 230 bp 3' of the terminal oligonucleotide ACTTTCC. The length of Sph I-A, which spans the array, is increased by 4.6 kb by the loss of this site. This is the equivalent of slightly more than three repeat increments, and consequently site heteroplasmics can appear to be locus I heteroplasmics and *vice versa.* The pattern becomes even more complex when individuals are heteroplasmic for both site and size, as is the case with three of the four (Table 5.7).

Figure 5.4 shows how locus I heteroplasmy, evident in the Sal I digests, is distributed in site heteroplasmy. Lanes 1 to 4 are Sal I-digested mtDNAs showing locus I length variation in the second fragment, Sal I-B. In lane 1, homoplasmy for size class 4 is indicated by the cleavage pattern, where only three bands are present and Sal I-B encompassing locus I is 10.9 kb. In lanes 2 and 3 heteroplasmy for size classes 3 and 4 is indicated by the pattern of four bands, two of them substoichiometric and corresponding to Sal I-B 9.5 kb and 10.9 kb. The individual in lane 4 is heteroplasmic for size classes 4 and 7, represented by Sal I-B fragments of 10.4 kb and 15 kb. The larger Sal I-B co-migrates with Sal I-A, and the smaller Sal I-B is shorter than the other examples of size class 4 because this animal is also heteroplasmic for classes B and D at locus II. Thus size class 4 exists as two populations, one class B at locus II, and the other, faintly visible, class D at locus II. Size class 7 is associated with class D at locus II.

Sal I/Sph I double digests of the individuals in lanes 1-4 are shown in lanes 5-8. When the mtDNAs are digested with both Sal I and Sph I, cleavage at the Sph I site flanking the locus I array shortens Sal I-B by 3.4 kb. The animal in lane 5, homoplasmic for size class 4, shows both an intact Sal I-B fragment at 10.9 kb and a cleaved fragment at the predicted length of 7.5 kb. In lane 6, the class 4 fragment is again intact, at 10.9 kb, while the class 3 fragment is cleaved, at 6.1 kb. The pattern is reversed in lane 7, where the class 3 fragment is intact at 9.5 kb and the class 4 fragment is cleaved at 7.5 kb. Finally, in lane 8 site heteroplasmy again occurs within a single size class. The largest band in this digest is size class 7 cleaved, at 12.6 kb. The second band is intact size class 4, just 10.4 kb with the 500 bp deletion at locus II. A very small proportion of fragments can be seen at 7.5 kb representing

the cleaved locus II-D size class 4 population, above the majority of size class 4 cleaved at 7.1 kb. A diagramatic representation of these heteroplasmic mitotypes is shown in Figure 5.5

In evaluating the significance of Sph I site heteroplasmy, the association with length heteroplasmy cannot be ignored. If the two markers (presence/absence of the site, and repeat copy number) were independent, then the predicted frequency of pattern B size class 3 would be (.104)(.221)=0.023, about six genomes in the sample, and of pattern B size class 4 (.104)(.532)=0.055, about fifteen animals. The actual distribution of pattern B among size classes is somewhat skewed for classes 3 and 4, in 4 and 20 genomes respectively.

Although mitochondrial markers cannot be independent and unlinked, if they are the result of two kinds of mutation they may have very different mutation rates which would give the appearance of linkage decay. If restriction site mutation is due to nucleotide substitution, the rate is generally held to be between 0.125 and 2 $x10^{-8}$/nt/year (Miyata *et al.* 1982, Brown 1985) for metazoan mtDNA, while the addition/deletion rate in mtDNA repeat arrays has been estimated at 10^{-2}-10^{-5}/sexual generation (Birky *et al.* 1989). Thus a site mutation should become distributed among size classes and appear to lose its association with its class of origin, although it may show a semblance of linkage within size classes. On a length mutation time scale linkage may be evident, but on a nucleotide substitution time scale linkage disappears. It is in the rare size classes that short-term linkage is most apparent. Pattern B is totally absent in size classes 2 and 7, but present in two of three class 8 animals.

Three different patterns of simultaneous heteroplasmy for the two markers were seen (4B/3A; 4A/3B; 7A/4B/4A). This raises the question of which of three hypotheses provides the best explanation. (1) The Sph I site may be a mutational hotspot, with recurrent mutation causing site loss randomly within and among size classes. (2) The waiting time to fixation or loss of the site mutation may exceed the combined waiting times to length mutation and loss of one variant. (3) Finally, double heteroplasmy may be due to paternal leakage.

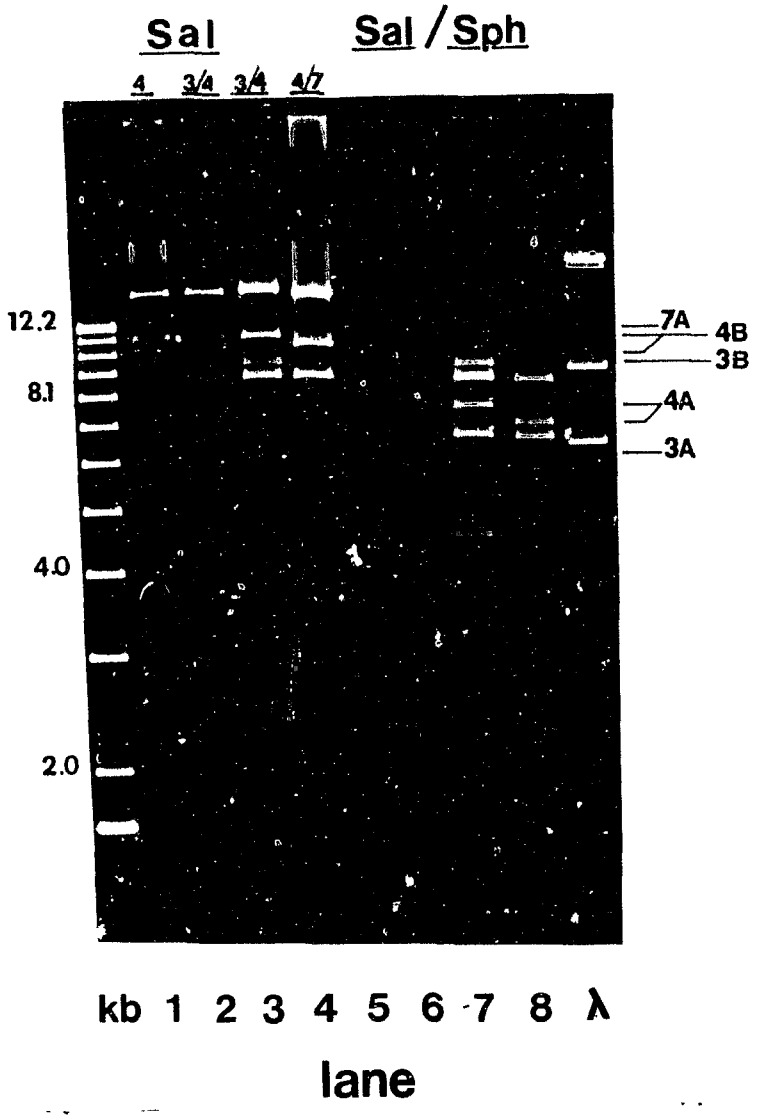


Figure 5.4 Restriction profiles of individuals heteroplasmic for Sph I cleavage patterns. In lanes 1-4 Sal I digestion shows the locus I repeat copy number, and in lanes 5-8 double digestion of the same individuals with Sal I and Sph I shows how Sph I cleavage patterns are distributed among locus I copy number variants. Numbers on the right refer to copy number, letters to Sph I pattern. Numbers on the left refer to fragments in the kilobase ladder size standard (kb).

Table 5.7 The co-occurrence of Sph I restriction site heteroplasmy and locus I copy number heteroplasmy.

	site	
	homoplasmic	**heteroplasmic**
locus I		
homoplasmic	257	1
heteroplasmic	15	3

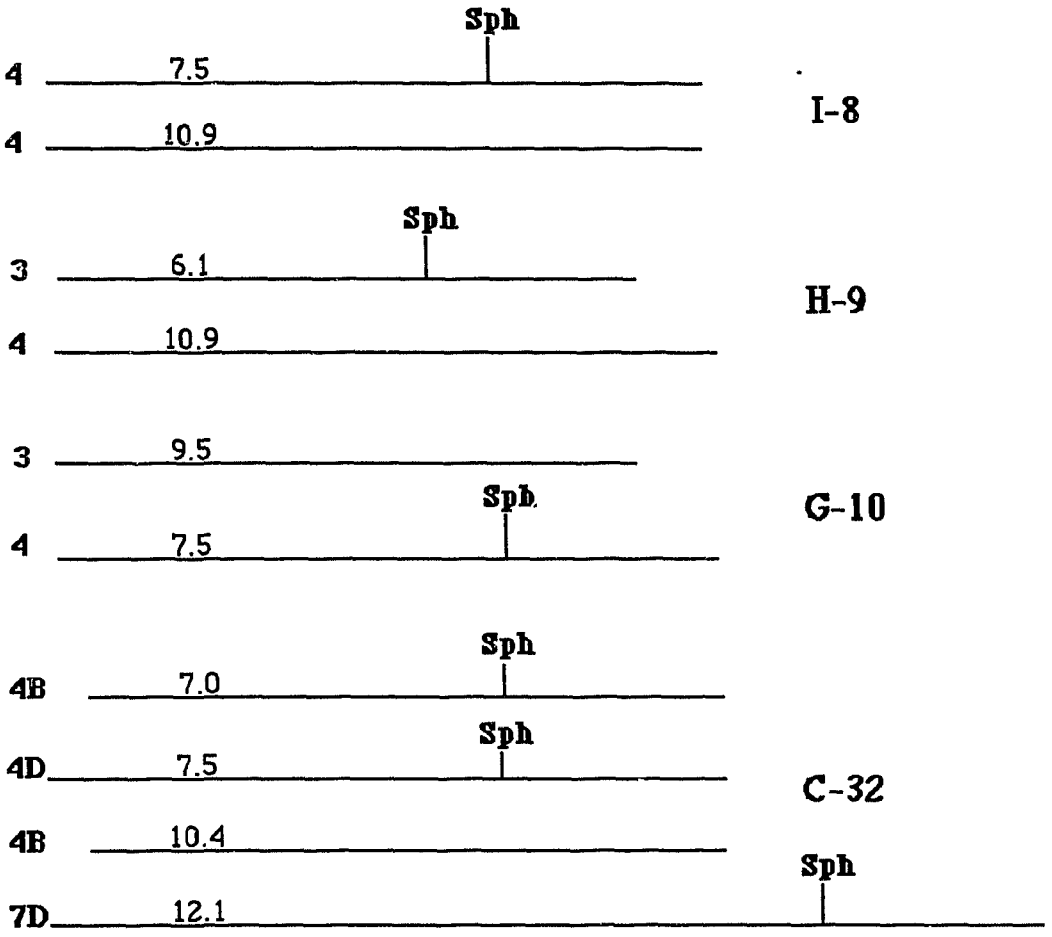


Figure 5.5 Restriction site heteroplasmy. The schematic representation of the Sal I marker fragment for locus I copy number includes the polymorphic Sph I site. Numbers to the left of the fragment sets give copy number of repeats, letters give locus II size allele. The position of the Sph I site is given, with the length of the locus I fragment above the line. Numbers to the right identify individual scallops.

5.2.1.5 Multilocus heteroplasmy

The *P. magellanicus* mitochondrial genome is rich in variation among and within individuals, with restriction site polymorphism as well as three distinct length polymorphic loci. Each kind of polymorphism has an associated frequency of heteroplasmy, ranging from 1.4% to 68.8%, suggesting different rates of mutation, dynamics of segregation, or selection. The survey of locus III showed that the 272 scallops comprised at least 537 genomes, and it is therefore of interest to ask whether heteroplasmy for other loci is a subset of locus III variation, or whether individuals homoplasmic for locus III can be heteroplasmic elsewhere.

A tabulation of locus III plasmy in individuals heteroplasmic for the other markers indicates that each type of heteroplasmy can occur when locus III is homoplasmic. Further, the individual triplasmic for locus I is only biplasmic for locus III. Among all locus I,II, and site heteroplasmics (26 individuals), heteroplasmy at locus III was distributed significantly differently than in the homoplasmic population ($G=5.696$, $df=1$, $p=0.017$). When locus III plasmy level was divided into homo-, bi-, and multiplasmic categories there was still significant heterogeneity between groups homoplasmic and heteroplasmic for other loci ($G=6.065$, $df=2$, $p=0.048$); among locus I, II, and site heteroplasmics, just 11.6% were homoplasmic for locus III, compared with 31.3% in the general population, 61.5% were biplasmic, compared with 45.6%, and 26.9% were multiplasmic, compared with 23.1%. The distribution of plasmy for the four markers together is given in Table 5.8.

In three-way tables of plasmy distribution for the three size variable loci, neither the three-factor interaction (G=0.41,df=1,p=0.52), nor any of the two-factor interactions [(IIIx II), G=0.92, df=2, p=0.633; (IIIxI), G=4.26, df=2, p=0.119; (IIxI), G=4.6, df=2, p=0.100)] were significant. However, at the next level of the model, the two-factor interactions of (IIxI) and (IIIxI) together were significant (G=8.96, df=3, p=0.030). This implies that locus II and III vary independently of one another conditioned on locus I. The result should be viewed with caution because of the small sample size, and because of the increased likelihood of making a type I error in multiple comparisons. The relationships among loci within multiple heteroplasmics supports an inference of independence.

Only one individual, C32, was heteroplasmic for all four markers, and a careful examination of the mitotype suggests that the variation is independent. The animal was heteroplasmic for locus I size classes 4 and 7, but locus II heteroplasmy was confined to the size class 4 genomes, and Sph I heteroplasmy was confined to the smaller size class 4 genome. Thus there were four distinct genomes within Eco RI-B, but only two category M alleles at locus III. The possible sequence of events leading to this mitotype is modelled in Figure 5.6.

The other instance of heteroplasmy for both locus I and II occurred between size classes, but interestingly this was the single animal triplasmic for locus I. In this case, size class 2 was associated with a deletion at locus II, but size classes 3 and 4 were not. Although size class 2 was rare in the population, it occurred twice in conjunction with heteroplasmy at locus II.

As discussed previously, restriction site heteroplasmy occurred twice within and twice between locus I size classes. One case of within-class

variation was associated with a single allele at locus III, the other with two alleles, while one case of between-class variation was associated with two alleles and the other with three alleles at locus III.

Table 5.8a. Distribution of plasmy for the four markers together. Numbers above the diagonal represent individuals heteroplasmic for rows and homoplasmic for columns, below the diagonal individuals heteroplasmic for both rows and columns.

	I	II	Sph	III
I	–	7	15	1
II	2	–	3	1
Sph	3	1	–	1
III	17	8	3	–

Table 5.8b The association of plasmy level for locus III with heteroplasmy for locus I, locus II, and Sph I.

	III		
	homoplasmic	**biplasmic**	**multiplasmic**
I			
homoplasmic	84	112	58
heteroplasmic	1	12	5
II			
homoplasmic	84	118	61
heteroplasmic	1	6	2
Sph			
homoplasmic	84	122	62
heteroplasmic	1	2	1

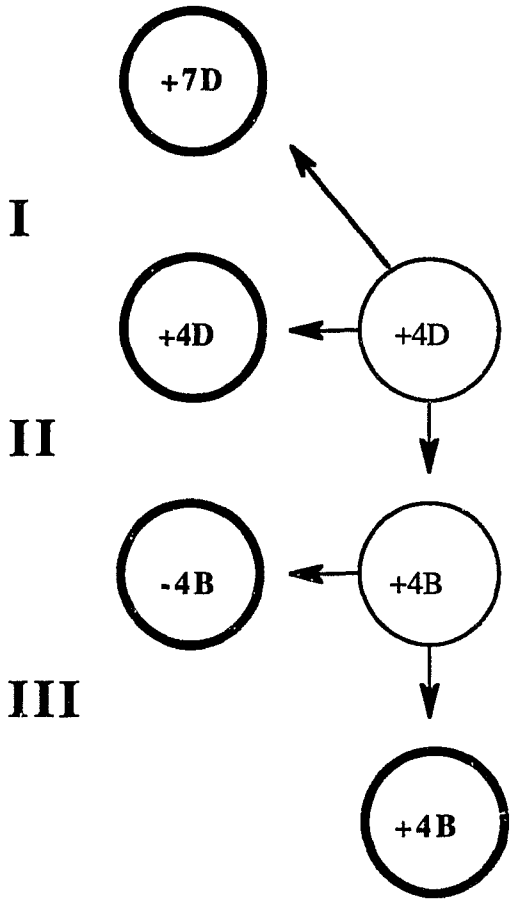


Figure 5.6 A model of the mutational steps needed to account for the complex mitotype of heteroplasmy for three markers simultaneously. The symbols within organelles are: +/- for the Sph I site that flanks the locus I array, a number designating repeat copy number for locus I, and **B/D** designating the size allele at locus II. The steps shown are: **I**, locus I mutation; **II**, locus II mutation; **III**, Sph I site mutation. The mitotypes present within individual C32 are shown in bold.

5.2.2 Tissue heterogeneity

Heteroplasmy is generally loosely defined as intra-individual variation, and under this definition variation in any sampled tissue qualifies an individual as heteroplasmic. Failure to detect variation, however, can only be interpreted as evidence for homoplasmy if the sample is truly random. With small organisms, mtDNA is often isolated from the whole body or a combination of tissues and the sample is random, but with larger animals mtDNA may be obtained from a single organ or tissue chosen because it is rich in mitochondria or simple to prepare. Such samples are not random, but represent cell lineages that diverged during the ontogeny of the organism and constitute yet another level of mitochondrial populations. The developmental history of cell lineages is crucial to understanding patterns of variation, as are population dynamic parameters such as mitochondrial population size and turnover, and the mitotic index of the cells.

The investigation of mitochondrial heterogeneity among scallop tissues was undertaken for two reasons. First, it would indicate whether sampling of adductor muscle gave a fair estimate of the amount of heteroplasmy in the population. Second, it would allow comparison of the levels of heteroplasmy in somatic and germline tissues.

In the multiple tissue sample of seventeen scallops, four animals were heteroplasmic at locus I and seven were heteroplasmic at locus III (Table 5.9). The resolution on the autoradiographs was not fine enough to permit discrimination of bands at locus III that differed by less than about 200 bp, so that seven is likely an underestimate.

At locus I, no fixed differences were observed between tissues, but three of the four heteroplasmics were heteroplasmic in only one tissue. It is

possible that heteroplasmy in the other tissues was simply below the level of resolution for genomic blots, making these quantitative rather than qualitative differences. Heteroplasmy was detected three times in gonad, once in gill, and once in muscle tissue. This represents a frequency of 5.9% for muscle, in close agreement with the value of 6.4% found in the large survey. The frequency of intra-individual variation is four times higher than this, at 23.5%, and the probability of detection by sampling muscle alone is 0.25. Gonad appears to be the best single-tissue indicator, with a probability of detection of 0.75. In all cases, the extra size class was larger than the one common to each tissue. For locus III, muscle was the best indicator, as each heteroplasmic individual was heteroplasmic in muscle. For gill tissue, the probability of detection was 0.571, and for gonad tissue 0.714. Where tissues differed, the extra fragment was twice smaller and once larger than the shared fragment.

The distributions of locus I size classes among genomes in the adductor muscle survey (n=299) and the tissue survey when each genome was counted once per individual (n=21) were not significantly different (G=1.439, df=2, p=0.487), indicating that the two samples may be considered as coming from the same population. The relative frequencies of heteroplasmy in somatic and germline tissues suggest that the large survey actually underestimated the amount of variation segregating in the population.

Table 5.9 Distribution of size variation among tissues. Size classes for locus I and size mitotypes for locus III are given for each tissue.

individual	Locus I			Locus III		
	gill	gonad	muscle	gill	gonad	muscle
1	3	3/4	3/4	M	M	M/L
2	4	4	4	M	M	M
3	4	4	4	M	S/M	S/M
4	5/4	4	4	M	M	M
5	3	3/4	3	–	–	–
6	4	4	4	M/M	M/M	M/M
7	4	4	4	L	L	L
8	5	5	5	M	M	S/M
9	4	4	4	M	M	M
10	3	3	3	M	M	M
11	3	3	3	S/M	S/M	S/M
12	4	4	4	S	S	S
13	3	3	3	S	S	S
14	5	5	5	M	M	M
15	3	3/4	3	M/M	M/M	M/M
16	4	4	4	M	M	M
17	3	3	3	S/L	S/L	S/L
probability of detection	0.25	0.75	0.25	.571	.714	1.0

5.3 Discussion

The abundance of heteroplasmy in *P. magellanicus* offers the opportunity to examine many of the questions surrounding the generation and fate of intra-individual variation in animal mitochondria. No other species with multilocus incremental polymorphism and heteroplasmy has been described, and single locus heteroplasmy has rarely been studied in discrete tissues. Gjetvaj (1989) noted three instances of multiple heteroplasmy, for a deletion and for an incrementally varying region, in the scallop *Chlamys islandica* . In shad, the two markers that exhibited heteroplasmy, a tandem repeat and a restriction site, were not observed to be both heteroplasmic within a single individual. Typically the ova or a combination of organs are used as a source of mtDNA in large animals and all or part of the body in small animals so that no comparisons among tissues are possible. In this study I have recorded heteroplasmy for four markers in samples obtained from a single tissue type, adductor muscle, and I have examined the distribution of variation among tissues within individuals.

5.3.1 Adductor Muscle

The frequencies and patterns of heteroplasmy for each marker singly and in combination suggested that regions of the mtDNA molecule experience different rates of change. Heteroplasmy can arise by mutation, the creation of new variation, or by reassortment of existing variation through recombination or paternal leakage. Its decay in a lineage occurs by drift or selection. There is little evidence for either paternal leakage or selection in animals and these factors are normally accorded little importance. Although there is no direct evidence for recombination, the

growing number of polymorphic tandem arrays that have been described has led several workers to suggest that it does occur in at least some species (Moritz *et al.* 1987, Rand and Harrison 1989).

Birky (1983) suggests that the homogenization of the small populations of mtDNA molecules within individual organelles is rapid, and presents a model of heteroplasmy as a mixed population of mitochondria homogeneous for different mtDNAs. The decay of heteroplasmy is a function of drift during replication of organelles and their segregation at cell division, and within a single tissue type mitochondrial populations can be assumed to undergo the same rate of turnover and cell division. When the organelle population sampled shares the same developmental history and segregation dynamics, differences in the frequency of heteroplasmy among markers should reflect differences in the generation, not decay, of variation. The within-generation drift can be considered standardized by sampling a single tissue, and between-generation drift should likewise be standardized among markers, all experiencing the same germline mitochondrial dynamics.

The occurrence of heteroplasmy indicates that variation is being generated but it is difficult to distinguish mutation from recombination as the mechanism. Mitochondrial recombination occurs in fungi and plants, and certain sequence elements have been implicated as preferred substrates. The GC clusters in yeast mtDNA may be recombinogenic (deZamaroczy and Bernardi 1986), and dispersed short direct repeats are implicated as sites of recombination in wheat mtDNA (Joyce *et al.* 1988) and as deletion hotspots in *E. coli* (Albertini *et al.* 1982) and human mtDNA (Schon *et al.* 1989). The long tandem repeat at locus I in *P. magellanicus* has both these features. The heterogeneity among repeats in *Gryllus*, where the homogeneous long array

incorporates both differences between repeats in the short array, appears to suggest recombination but there are no comparable data for *P. magellanicus* . However, the Sph I site heteroplasmy that occurs in conjunction with locus I heteroplasmy may be indicative of recombination. Unequal crossing over between size class 4 molecules differing at the Sph I site would produce array length heteroplasmy with an equal expectation of the site loss being associated with the long or short array. In two of four instances site heteroplasmy occurred between size classes, which would be the pattern expected if recombination is equally likely within or between site variants. The major drawback to a recombination hypothesis is the lack of evidence for mitochondrial fusion in animals except in formation of the sperm midpiece. If individual organelles are rapidly purified the opportunity for recombination between restriction morphs is limited by drift.

Paternal leakage could also explain this double heteroplasmy, as the site loss variant has a frequency in the population of 10%. Locus I has a high diversity index with a probability of 64% that randomly chosen gametes differ in size class. With site and size variation independent, matings between site variants would be likely to include size differences. There is only a single report of possible paternal leakage in an animal species; restriction cleavage patterns in *D. mauritiana* suggest that intromission occurred through male transmission (Satta *et al.* 1988). There are few data on invertebrate species where the size difference in male and females gametes is not as exaggerated as in vertebrates. In *P. magellanicus* , sperm length and egg diameter average about 40 μm and 80 μm, respectively (Naidu 1970). Avise and Lansman (1983) suggested that either there is an exclusion mechanism whereby entry of the sperm midpiece into the zygote is

prevented, or the small number of sperm mtDNA molecules is overwhelmed by a highly amplified egg mtDNA population. Alternatively, sperm mitochondria may be altered in some way that prevents transmission, as has been described in plants (Vaughn *et al.* 1980). If paternal leakage makes a significant contribution to the generation of variation it will most likely be in species where the small egg volume precludes vast overrepresentation of maternal mtDNA in the zygote. Sperm mitochondria have been observed in the zygote of the bivalve *Mytilus edulis* (Longo and Anderson 1970) but no data are available for *P. magellanicus* . Controlled matings between different restriction morphs might provide direct evidence for leakage if it is relatively frequent, but the long generation time of the scallop makes repeated backcrossing to amplify any paternal contribution impractical.

The simplest explanation for the generation of length variation is mutation, as mutation is a necessary condition for the origin of a tandem array. The mutation rate can be related to the frequency of heteroplasmy if the drift parameters of effective mitochondrial population size and number of germ cell generations are known (Clark 1988). Lacking this information, the assumption can be made that mutation is frequent enough to allow detection of heteroplasmy in some proportion of individuals despite drift, and it can be predicted that as the mutation rate increases so does the frequency of multiplasmy, where new mutations arise before segregation is complete. This is the pattern in *P. magellanicus.* At locus III the frequency of heteroplasmy is 69% and the frequency of multiplasmy is 23%; at locus I the frequency of heteroplasmy is 6.4% and of multiplasmy 0.3%; at locus II the frequency of heteroplasmy is just 3.4% and no multiplasmy was observed. These data were obtained from a single tissue type.

The patterns of intra-individual variation as well as the mutation rates are different among loci. At locus I the genomes differed most often by a single repeat, while at locus II genomes always differed by at least two increments. At locus III most variation was within-category, but frequently variants differed by as much as 200 bp within and among categories. This could be a function of the size of the unit of variation. Replication slippage at locus I would entail looping out of a long sequence, and it is reasonable to assume that extrusion of a single repeat would be most likely. At the other loci, the total range of variation is less than the length of a single locus I repeat and constraints on slippage may be relaxed. Although size may partly explain patterns of variation it cannot account for differences in mutation rates however, for the intermediate locus I has by far the largest unit of variation.

The mutation rates are likely dependent on the internal structure of the repeating units, the sequence environment, and the location of the array relative to the control region. Hauswirth and Laipis (1984) discussed the possibility that secondary structure in the displaced mtDNA strand causes the replication machinery to pause, providing an opportunity for slippage to occur. The inverted repeat in the locus I element may provide such a signal, but replication slippage is unlikely to underlie the gain and loss of perfect repeats of nearly 1.5 kb. The presence of structural elements at the other two loci, where slippage may be more feasible, can only be ascertained by sequencing. The sequence similarity between locus II and III suggests that some common sequence may be involved in events generating variation, but this too can only be resolved with sequence data.

The asymmetrical replication of animal mtDNA requires the exposure of some of the genome as single stranded DNA through much of the replication period (Brown 1983). Sequences preceding the uncovering of the light strand origin are exposed only briefly, but those exposed following the initiation of both light and heavy strand replication remain single stranded for the longest period. While single stranded, elements such as the 10 bp inverted repeat of locus I could anneal between repeats and result in a heterogeneous daughter molecule. If intra-strand interactions are involved in mutation, the location of such events must influence their frequency. The preponderance of control region-associated length polymorphism in animal systems is likely related to replication dynamics. Clayton (1982) has documented a high turnover of D-loop strands, and it may be that multiple initiations produce a multi-strand neighbourhood where local amplification occurs during the strand breakage and ligation that resolve catenated daughter molecules.

Buroker *et al.* (1990) recently proposed a model of competitive equilibrium between the D-loop strand and the H-strand to account for varying numbers of perfect direct 82 bp repeats in the control region of sturgeon mtDNA. The repeating unit is capable of stable intra-strand folding, singly or as a multimer, that can result in exact loss or gain of one or more repeats during replication provided that the D-loop strand does serve directly as the primer for elongation of the daughter H-strand. The special characteristics of this system make it unlikely that the mechanism is generalizable, especially for long repeats such as that found at locus I in *P. magellanicus*.

5.3.2 Tissues

The observation of different frequencies of heteroplasmy for single loci among tissues holds implications for both the generation and decay of variation. In contrast with the previous experiment, where single tissue segregation dynamics were assumed to be constant to allow a comparison of mutation rates among loci, here mutation rates were assumed to be constant at each locus.

In the scallop, gill tissue likely has the highest complement of mitochondria per cell and may also have the highest turnover rate due to mitochondrial decay in this actively respiring organ. The gonad is inactive through much of the year, but primary oogonia undergo many mitoses in preparation for oogenesis as the organ begins to ripen (de Jong-Brink *et al.* 1983). A female may shed $1x10^7$ ova in a single yearly spawn, and may experience ten or more of these cycles in her lifetime. Thus two levels of amplification are important in determining her lifetime output of mitochondrial variation. There are bursts of rapid cell division, resulting in many germ cell generations per animal generation, and there are bursts of mitochondrial replication resulting in the amplification of the mitochondrial complement in the egg. One or both of these factors may be implicated in the higher frequency of germ cell variation suggested by Bentzen *et al.* (1988). Muscle cells probably represent an intermediate, with a steady turnover among a smaller complement of mitochondria than gill.

Two examples of inter-tissue variation in heteroplasmic frequency have been described. Studies of human myopathies have shown that mitochondrial deletions may be associated with disease and that the severity

of the symptoms is often correlated with the ratio of deleted to normal genomes. Holt *et al.* (1988a) documented several cases in which rapidly dividing white blood cells were fixed for normal mtDNA while non-dividing muscle cells retained both normal and deleted genomes, and advanced a selection argument to account for this patte. . This requires that the cell be the unit of selection, a reasonable assumption in a situation where cell function depends on the proportion of deleted genomes although it might be argued that aberrant mitochondria that are unable to respire would have a lower turnover and different replication rate than normal mitochondria.

In the other example, Hauswirth and Laipis (1985) reported different ratios of variant genomes in brain, heart, and liver within two offspring of a heteroplasmic cow. These tissues were selected to represent cell lines that diverged early in ontogeny and the results suggested that developmental drift is important in the segregation of zygotic variation. Inter-tissue variation has not been explored in insects, where syncytial development and larval metamorphosis impose new levels of complexity.

For selection to act in cell populations, there must be variation among cells attributable to their mitotype. For selection to act intracellularly on the mitochondrial population, one mitotype must have some advantage relative to the other. Rand and Harrison (1986) have suggested that this advantage may be genome size, with smaller molecules having a faster rate of replication. In *P. magellanicus* this selection model cannot easily be tested by transmission studies but the patterns of variation among tissues can provide some insight, as change in size class frequencies should show a bias toward smaller genomes. Intra-tissue frequencies were not measured, but in all four individuals heteroplasmic at locus I at least one tissue appeared to

have been fixed for a single variant. In every case this was the smaller of the two genomes. At locus III however, in two of three cases the variant fixed was the larger one, indicating not merely a different intensity but a different direction of selection. The distribution of variation within the population as a whole shows a pattern typical of stabilizing selection. The largest and smallest classes are rare and the intermediate classes are most frequent at all loci.

No consistent pattern of variation emerged within or among tissues. Individual tissue types showed marked differences in the probability of detection of heteroplasmy between the two loci. This makes a developmental explanation for the partitioning of variation appealing. As a mollusc, *P. magellanicus* undergoes spiral cleavage and early determination of embryonic tissues. In the ontogeny of bivalve molluscs separation of the cell lineages leading to gill, and to muscle and gonad, occurs following the fourth cleavage (Galtsoff 1964, Raven 1966). Random drift at this stage of development could have a large impact on inter-tissue variation in adult animals.

Chapter 6

Population Genetics

6.1 Introduction

Genetic data from temporally or spatially discrete samples provide information about the evolution and history of populations. The advent of starch gel electrophoresis and allozyme analysis in the 1960's (Harris 1966, Hubby and Lewontin 1966, Lewontin and Hubby 1966) ushered in a period of empiricism in population genetics, which in turn sparked new theoretical developments. One of the most significant of these was Kimura's (1968) neutral theory of evolution, in which he suggested that most mutations are neutral in effect and consequently genetic drift is a much more prevalent force in genetic change than selection. The assumption of neutrality has become a cornerstone in molecular analyses of population subdivision and evolutionary history.

For nuclear loci, recombination and migration can complicate the signals of mutation and drift, and population differentiation is typically evidenced by differences in allele frequencies rather than by differential fixation (Wilson *et al.* 1985). In contrast, the animal mitochondrial genome exhibits uniparental inheritance which maintains discrete lineages unchanged by recombination. In population studies, mitochondrial markers are usually restriction sites and the nucleotide substitutions resulting in site loss or gain are assumed to be selectively neutral. The combined actions of mutation and random lineage extinction tend to produce low clonal diversity within populations and high diversity among populations (Chapman *et al.* 1982, Avise *et al.* 1984).

MtDNA polymorphisms that are due to length mutations have been described in a number of animal species. Studies of mtDNA transmission in length-heteroplasmic female lines of *Drosophila mauritiana* (Solignac *et al.* 1987) and *Gryllus* crickets (Rand and Harrison 1986) suggested that selection acts on genome length within mitochondria among mtDNA molecules. However, the observation that most variation is among and not within individuals in populations of these flies and crickets implies that random drift is a powerful force at some hierarchical level within individuals. Testing a hypothesis of selection at higher levels is exceedingly complex because of the number of interactions, the difficulty of obtaining large samples, and the outstanding population dynamic questions remaining in mitochondrial biology. These include cytonuclear interactions and the replication and segregation dynamics in organismal reproduction. Population cage experiments designed to study the evolution of mtDNA characters in *Drosophila melanogaster* resulted in some dramatic shifts in the frequency of morphs but no consistent pattern or direction of change (Clark and Lyckegaard 1988).

Several studies of mtDNA length polymorphism in natural populations have been carried out (Monnerot *et al.* 1984, Hale and Singh 1986, Solignac *et al.* 1986, Bentzen *et al.* 1988, Rand and Harrison 1989, Gjetvaj and Zouros, in press) and in many cases the polymorphism appears due to a kind of incremental variation. A repeat sequence has been identified or inferred for most of these (Solignac *et al.* 1986, Bentzen *et al.* 1988, Gjetvaj 1989, Rand and Harrison 1989). Length variation has typically shown a higher within-population than among-population variance component (Rand and Harrison 1989, Gjetvaj and Zouros, in press), the reverse of sequence

variation. This has been attributed to the bounded nature of length polymorphism, where a limited number of variants is possible and there is backward and forward mutation among them (Rand and Harrison 1989, Birky *et al.* 1989). Clark (1988) has developed a model that shows heteroplasmy can be stably maintained in populations by a balance between drift and mutation, where bidirectional mutation is necessary and sufficient for polymorphism.

The population biology of *P. magellanicus* is poorly understood, particularly larval dispersal and juvenile recruitment. The pelagic larval phase may last up to eight weeks, depending on water temperature, and some individual control over the timing of metamorphosis may be exercised in choosing suitable substrates (M. Dadswell, Acadia University, pers. comm.). Juveniles are good swimmers, but adults are basically sedentary. What gene flow occurs between beds is most likely a consequence of passive larval dispersal, with current and water temperature patterns being major factors in settlement and viability. Spat may be unrelated to the adults in the area in which they settle.

In this study I have sampled scallop populations from six discrete geographical areas (Figure 6-1), four of which have scallop beds of sufficient density to support a fishery. Five of the samples were collected by scallop dragger from depths ranging from 25 m to as much as 140 m, and the sixth was collected by divers from a depth of less than 12 m (Table 6-1).

Whether the geographic structure has contributed to the genetic structure of these populations by isolation and local differentiation is unclear, for the extent and patterns of gene flow are unknown. *P. magellanicus* is a cold water species confined to the Northwest Atlantic, from the north shore

of the Gulf of St. Lawrence to Cape Hatteras, North Carolina (MacDonald and Thompson 1985a). The sampling locations in this study include much of the range of *P. magellanicus* in Atlantic Canada. Previous allozyme studies have sampled a smaller part of the range, primarily Georges Bank, Browns Bank, and the Digby area of the Bay of Fundy (Foltz and Zouros 1984, Gartner-Kepkay and Zouros 1985), and have found little differentiation and small genetic distances between populations.

I have examined the amount and distribution of variation among populations for six nuclear loci, three length polymorphic mtDNA loci, and seven polymorphic mtDNA restriction sites and obtained genetic distance, heterogeneity, and diversity statistics. Further, I have tested correlations between these markers and other individual characters.

6.2 Results

6.2.1 Allozymes

Genotypes at six nuclear loci (Table 6.2) were scored by Mr. Kamran Ahmad (unpubl. honors thesis) by starch gel electrophoresis of adductor muscle subsamples obtained before the tissue was processed for mtDNA isolation. Enzymes scored were phosphoglucomutase (*Pgm*, EC 2.7.5.1), octopine dehydrogenase (*Odh*, EC 1.5.1.11), glucosephosphate isomerase (*Pgi*, EC 5.3.1.9), mannosephosphate isomerase (*Mpi*, EC 5.3.1.8), aspartate aminotransferase (*Aat*, EC 2.6.1.1), and 6-phosphogluconate dehydrogenase (*Pgd*, EC 1.1.1.44).

Of the six, *Pgi*, *Mpi*, *Aat*, and *Pgd* showed little differentiation among geographic populations. *Pgm* showed significant heterogeneity of allele frequencies (rare alleles lumped, G=18.588, df=10, p=.046); heterogeneity was more pronounced for *Odh* (G=31.581, df=5, p<.001). These two loci also show the greatest range of values for effective number of alleles (n_e) among populations. The Ship Harbour sample (S) had the lowest allelic diversity for four of the six loci, and was the only population fixed at any locus (*Pgi*).

Genetic distances between populations were calculated by the method of Nei (1972), where the identity index I_{xy} is given by

$$I_{xy} = \Sigma x_i y_i / (\Sigma x_i^2 \Sigma y_i^2)^{1/2}$$

averaged over all loci, and the distance index D_{xy} is given by

$$D_{xy} = -\ln(I_{xy}).$$

An identity / distance matrix is given in Table 6.3a. Pairwise G-tests for heterogeneity with all loci combined and rare alleles lumped showed

significant differences between CH (G=25.187, df=12, p=.014), HS (G=25.651, df=12, p=.012), and IS (G=33.272, df=12, p=.001).

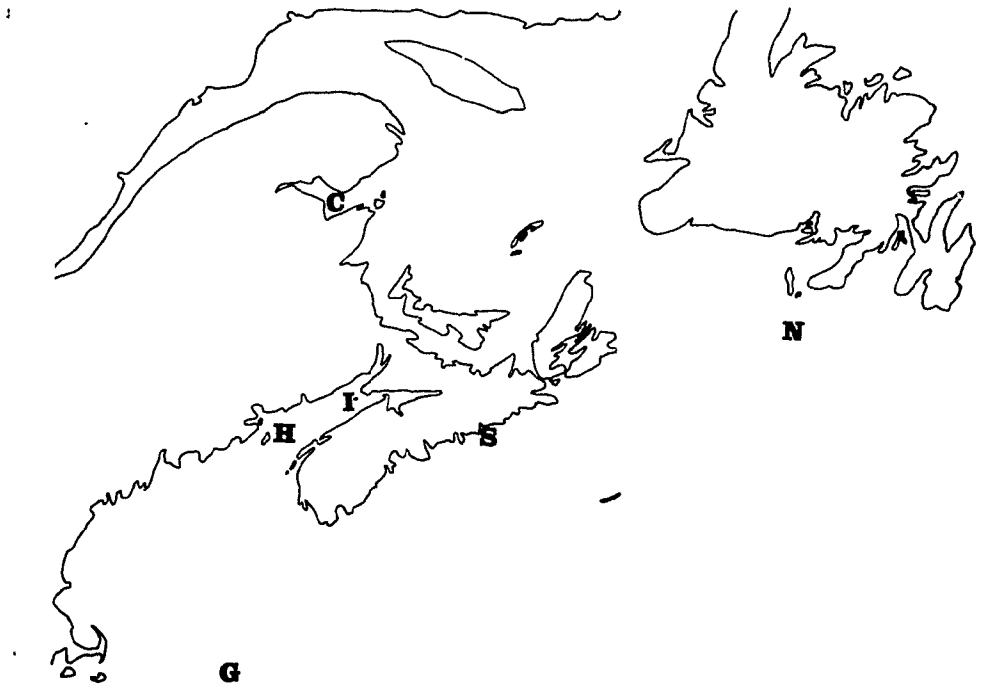


Figure 6.1 Map showing sampling locations with the following letter designations for each of the six populations: **C**, Baie des Chaleurs; **H**, Owen Basin; **I**, Isle Haute; **N**, St. Pierre Bank; **G**, Georges Bank; **S**, Ship Harbour.

Table 6.1 Collection data for scallop samples.

sample	locale	location	depth	collection
C	Baie des Chaleurs	48°N 64°35'W	25 m	dragger
H	Owen Basin	44°57'N 66°44'W	145 m	dragger
I	Isle Haute	45°12'N 65°W	40-55 m	dragger
N	St. Pierre Bank	45°36'N 56°W	40-50 m	dragger
G	Georges Bank	42°N 66°W	40-92 m	dragger
S	Ship Harbour	44°49'N 63°15'W	12 m	diver

Table 6.2 Allele frequencies at six nuclear loci. Allelic diversity for each locus is given as n_e.

locus		population C	H	I	N	G	S
Pgm	96	0.011	0.000	0.000	0.009	0.000	0.000
	98	0.120	0.115	0.064	0.139	0.107	0.224
	100	0.619	0.744	0.745	0.593	0.726	0.579
	102	0.250	0.141	0.191	0.259	0.167	0.197
	n_e	2.173	1.705	1.679	2.272	1.766	2.357
Odh	98	0.000	0.000	0.009	0.019	0.000	0.000
	100	0.652	0.346	0.400	0.519	0.512	0.697
	105	0.348	0.654	0.591	0.462	0.488	0.303
	n_e	1.831	1.827	1.963	2.070	1.999	1.731
Pgi	87	0.000	0.000	0.009	0.000	0.012	0.000
	100	0.957	0.936	0.936	0.954	0.929	1.000
	115	0.043	0.051	0.055	0.046	0.059	0.000
	140	0.000	0.013	0.000	0.000	0.000	0.000
	n_e	1.090	1.138	1.138	1.096	1.159	1.000
Mpi	75	0.000	0.000	0.000	0.009	0.012	0.000
	87	0.250	0.231	0.282	0.278	0.286	0.197
	100	0.750	0.769	0.718	0.704	0.702	0.803
	108	0.000	0.000	0.000	0.009	0.000	0.000
	n_e	1.600	1.551	1.681	1.745	1.710	1.463
Aat	50	0.022	0.000	0.018	0.009	0.024	0.000
	70	0.000	0.026	0.009	0.000	0.012	0.013
	90	0.533	0.423	0.491	0.463	0.369	0.382
	100	0.435	0.551	0.482	0.528	0.583	0.605
	105	0.000	0.000	0.000	0.000	0.012	0.000
	115	0.010	0.000	0.000	0.000	0.000	0.000
	n_e	2.110	2.069	2.111	2.027	2.096	1.953
Pgd	80	0.000	0.000	0.000	0.009	0.000	0.000
	100	0.837	0.769	0.782	0.833	0.893	0.776
	130	0.011	0.000	0.000	0.000	0.000	0.000
	150	0.152	0.231	0.218	0.158	0.107	0.224
	n_e	1.382	1.551	1.517	1.391	1.236	1.533

Table 6.3a Nei's genetic distances from allozyme allele frequencies. Numbers above the diagonal are identities, those below the diagonal are distances. Asterisks denote significant distances.

	C	H	I	N	G	S
C		.9664	.9786	.9935	.9844	.9883
H	.0342*		.9963	.9729	.9869	.9605
I	.0216	.0037		.9899	.9898	.9641
N	.0065	.0275	.0102		.9938	.9865
G	.0157	.0132	.0103	.0062		.9794
S	.0118	.0403*	.0366*	.0136	.0208	

Table 6.3b Nei's genetic distances from locus I copy number frequencies. Numbers above the diagonal are identities, those below are distances. Asterisks denote significant distances.

	C	H	I	N	G	S
C		.9539	.9609	.9943	.9277	.8146
H	.0472		.9943	.9612	.9880	.9352
I	.0398	.0057		.9707	.9929	.9380
N	.0057	.0396	.0298		.9438	.8425
G	.0751	.0121	.0071	.0578		.9670
S	.2051*	.0670	.0641	.1713*	.0336	

6.2.2 MtDNA Polymorphism

6.2.2.1 Locus I

The frequencies of locus I size classes among geographic populations are given in Table 6.4. No population included all seven size classes, and only size classes 3, 4, 5, and 6 were ubiquitous among populations. The rare classes 2 and 7 occurred in only three populations, and class 8 in two.

Size class 4 predominated in each population, with a frequency range of 0.438-0.595. The adjacent size classes 3 and 5 had ranges of 0.082-0.396 and 0.083-0.225 respectively, and together size classes 3, 4, and 5 accounted for 87.8 - 95.3% of the variation within populations.

Locus I showed significant heterogeneity of size class frequencies among populations (G=43.85, df=30, p=0.049). Many differences occurred in the rare classes, which were either absent or present only once or twice in a sample, but when rare classes were lumped in the analysis the distribution remained heterogeneous (G=27.45, df=15, p=0.025). It was apparent from the distributions that most of the differences lay in the relative frequencies of size classes 3 and 5 (Figure 6.2). The extremes were represented in populations C and N, where size class 3 had frequencies of 0.082 and 0.109 respectively, and S, where its frequency was 0.396. The converse held for size class 5. In C and N, class 5 frequencies were 0.225 and 0.218 respectively, while in S it was 0.083.

Considering size classes as alleles, genetic distances were calculated for locus I according to Nei's index (Table 6.3b). In pairwise tests, only two significant differences were found: CS (G=18.335, df=5, p=.003), and NS (G=15.552, df=5, p=.008).

As noted in the previous chapter, 18 individuals were heteroplasmic for locus I. Each population included at least one instance of heteroplasmy, but in no case did this represent a unique observation of any size class. Whenever a rare class occurred in heteroplasmy, it also occurred at least once in homoplasmy in that population.

The diversity statistics of Birky *et al.* (1989) and Rand and Harrison (1989) were used to examine the distribution of variation within and among several hierarchical levels. For each population $\overline{K_b}$ was calculated as an average within-individual diversity using intra-individual size class frequencies, and K_c was calculated using population size class frequencies. These values were used to determine G_{IP}, the proportion of the population variation accounted for by the differences among individuals, given by

$$G_{IP} = (K_c - \overline{K_b}) / K_c.$$

An examination of Table 6.4 shows that for every population greater than 94% of the variation was among, not within, individuals. Further, an average population diversity, $\overline{K_c}$=0.6147, was calculated and used to apportion diversity at three levels using the C statistics of Rand and Harrison (1989):

within individuals

$$C_I = \overline{K_b} / K_t$$

among individuals within populations

$$C_{IP} = (\overline{K_c} - \overline{K_b}) / K_t$$

among populations within species

$$C_{PS} = (K_t - \overline{K_c}) / K_t$$

where K_t is the probability that two molecules drawn from two different populations differ in size class, given in Chapter 4 as 0.6388. Just 3.8% of the total diversity can be attributed to variation among populations (C_{PS}) and

another 3.8% to variation within individuals (C_I); variation among individuals within populations (C_{IP}) accounts for the remaining 92.4%.

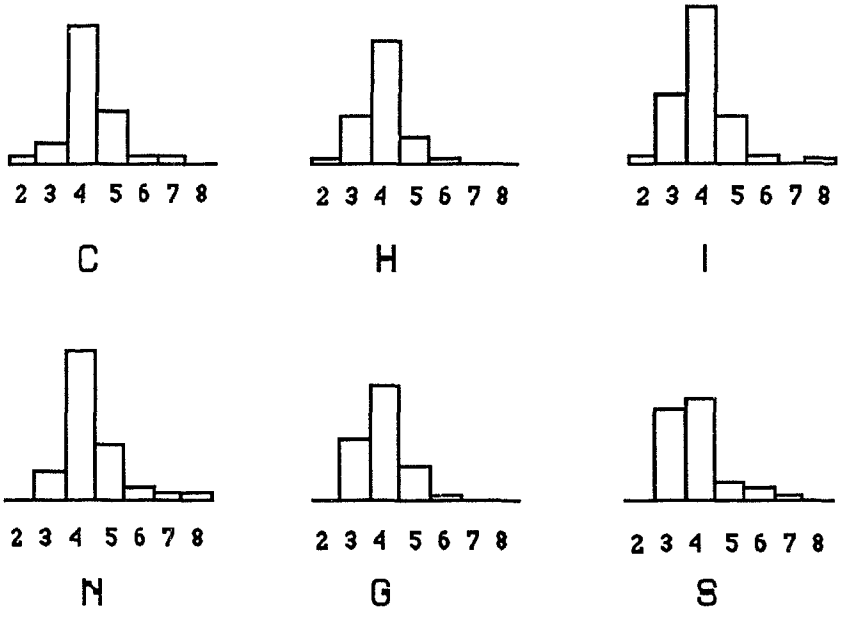


Figure 6.2 Histograms of locus I repeat copy number distributions within populations.

Table 6.4 Size class frequencies at locus I. **N**, number of individuals; **n**, number of genomes; $\overline{K}_b$, average intra-individual diversity; K_C, population diversity; G_{IP}, proportion of population variation accounted for by differences among individuals.

	population					
size class	C	H	I	N	G	S
2	0.0408	0.0238	0.0328	0	0	0
3	0.0816	0.2381	0.2295	0.1091	0.2955	0.3959
4	0.5715	0.5952	0.5246	0.5455	0.5227	0.4375
5	0.2245	0.1191	0.1639	0.2181	0.1591	0.0833
6	0.0408	0.0238	0.0328	0.0545	0.0227	0.0625
7	0.0408	0	0	0.0364	0	0.0208
8	0	0	0.0164	0.0364	0	0
N	46	39	55	54	42	44
n	49	42	61	55	44	48
$\overline{K}_b$	0.0252	0.0316	0.0315	0.0068	0.0237	0.0327
K_C	0.612	0.574	0.611	0.637	0.614	0.64
G_{IP}	0.9589	0.9450	0.9484	0.9893	0.9614	0.9489

6.2.2.2 Locus II

At locus II, four of the six size alleles were rare, and only in population C were all six present. Three were unique to this population. The frequency distribution in Table 6.5 shows a range of 0.019-0.159 for class B, and 0.800-0.981 for class D.

A G-test for heterogeneity showed that the distribution of size alleles among populations was heterogeneous (rare alleles lumped, G=12.349, df=5, p=.03). In pairwise comparisons by Yates corrected G-tests significant differences were found between N and C (G=6.541,df=2,p=.011), N and I (G=4.491, df=2, p=.034), and N and G (G=4.653, df=2, p=.031). Within populations the distributions resulted in a wide range of diversity estimates expressed as K_C values, from 0.0372-0.350. This range, encompassing an order of magnitude, is far greater than at other loci.

Not all populations included individuals heteroplasmic at locus II; both H and N were entirely homoplasmic. Each of C and I included three heteroplasmics, G had two, and S had one. When populations are ranked for proportion of heteroplasmic individuals and diversity, the association is not significant (n=6, Kendall's tau=0.6, p>0.1).

Table 6.5 Size class frequencies at locus II. **N**, number of individuals; **n**, number of genomes; K_c, population diversity.

	population					
size class	**C**	**H**	**I**	**N**	**G**	**S**
A	0.025	0	0.019	0	0	0
B	0.075	0.086	0.132	0.019	0.159	0.067
C	0.025	0	0	0	0	0
D	0.800	0.914	0.849	0.981	0.841	0.933
E	0.025	0	0	0	0	0
F	0.050	0	0	0	0	0
N	38	35	50	54	42	44
n	41	35	53	54	44	45
K_c	0.350	0.157	0.261	0.037	0.267	0.125

6.2.2.3 Locus III

The highly variable locus III is characterized by three size categories and six levels of plasmy, which together define 26 mitotypes in the total sample. Because the categories are artificial groupings of actual size alleles and no identity can be assumed between any alleles within categories among individuals, mitotypes represent an underestimate of variation.

No population contained all 26 mitotypes; the range was 11 (H) to 21 (S). Only eight were common to all populations, including two of the three homoplasmic mitotypes (Table 6.6). No single mitotype predominated in all six populations. In four, mitotype B (category M) was most common, with a frequency range of 0.167 - 0.325. In population G the most frequent mitotype was G (MM, 0.190), and in S it was H (ML, 0.136). Seven of the mitotypes were unique and another occurred twice, but within a single population.

Mitotype frequencies and the diversity statistic H are given in Table 6.6. Within population diversity was estimated as in Chapter 4, using Nei's diversity statistic $\hat{H}$, given by

$$\hat{H} = N (1 - \Sigma x_i^2) / (N-1).$$

This statistic expresses the probability of drawing at random from a population two individuals having different mitotypes. Diversity is very high, ranging from 0.8201-0.9525.

Variation at the mitotype level can be examined by grouping mitotypes to reflect specific character states. In this way various measures of similarity can be used to look for trends, such as a tendency for populations to have mitotypes that include a specified size category or that comprise a certain number of size alleles. Mitotypes grouped according to presence or absence

of one size category with the presence/absence groups further divided into homo- and heteroplasmic subgroups, represent the distribution that category within individuals among populations (Table 6.7a). This kind of classification gives qualitative information about the association between size category and intra-individual variation in populations without weighting for size frequencies within either individuals or populations. The distribution was heterogeneous for each category (S: G=30.801, df=15, p=0.009; M: G=26.741, df=15, p=0.031; L: G=26.143, df=15, p=0.037), indicating that there were significant differences in the patterns of occurrence and plasmy level association for categories among populations. Mitotype groups of size category presence/absence alone did not show significant heterogeneity (S, p=0.061; M, p=0.113; L, p=0.220), but mitotypes grouped according to plasmy level alone, as homo-, bi-, and multiplasmic, did have a heterogeneous distribution (G=30.808, df=10, p=.001) with significant differences between S and each of the five other populations (Table 6-7b).

Variation can also be examined at the genome level, where all size alleles are pooled within populations. In contrast to the previous analysis, where mitotypes A (S) and D (SS) were given equal weight in scoring for presence of S, here these individuals contributed one and two category S alleles respectively to the population pool. A tabulation of size category frequencies is given in Table 6.8. The distribution of categories among populations was heterogeneous (G=21.972, df=10, p=.015), with significant differences between H and each of the five other populations, and between G and N, and G and S.

Table 6.6 Locus III mitotype frequencies within populations, and overall (T). N_i, number of individuals; M_i, number of mitotypes; $\hat{H}_i$, mitotype diversity. Numbers in parentheses on the left give plasmy level of mitotype groups.

		population						
mitotype		C	H	I	N	G	S	T
	A	.071	.027	.093	.132	.071	.045	.077
(1)	B	.238	.325	.167	.245	.143	.068	.195
	C	.071	.054	.018	.075	.024	0	.040
	D	.048	.027	.093	.075	.024	.023	.051
	E	.143	.108	.093	.113	.095	.068	.103
(2)	F	0	0	.018	.019	.024	.023	.015
	G	.119	.270	.167	.057	.190	.091	.143
	H	.024	.054	.057	.075	.119	.136	.077
	I	.071	.027	.111	.038	.095	.045	.066
	J	0	0	0	0	0	.023	.004
	K	0	0	.018	.057	0	.114	.033
	L	0	0	0	0	0	.023	.004
	M	.024	.027	.037	.038	.071	.068	.044
(3)	N	.024	0	.037	.019	.024	.045	.026
	O	0	0	.018	.019	0	.045	.015
	P	.071	.054	.018	0	0	.023	.026
	Q	.024	0	.037	.019	.048	.045	.029
	R	.024	0	0	.019	.024	0	.011
	S	0	.027	0	0	0	.023	.007
	T	.024	0	0	0	0	.023	.007
(4)	U	0	0	0	0	0	.023	.004
	V	0	0	0	0	0	.023	.004
	W	0	0	0	0	.048	0	.007
	X	.024	0	0	0	0	0	.004
(5)	Y	0	0	.018	0	0	0	.004
(7)	Z	0	0	0	0	0	.023	.004
N_i		42	37	54	53	42	44	272
M_i		15	11	16	15	14	21	26
$\hat{H}_i$		.903	.820	.914	.898	.915	.953	.908

Table 6.7a Distribution of locus III size categories within individuals among populations. Numbers represent designated homoplasmic (heteroplasmic) mitotypes.

mitotype	C	H	I	N	G	S
includes **S**	3(12)	1(7)	5(17)	7(17)	3(10)	2(21)
excludes **S**	13(14)	14(15)	10(22)	17(12)	7(22)	3(18)
includes **M**	10(20)	12(20)	9(27)	13(21)	6(25)	3(33)
excludes **M**	6(6)	3(2)	6(12)	11(8)	4(7)	2(6)
includes **L**	3(11)	2(5)	1(15)	4(10)	1(16)	0(18)
excludes **L**	13(15)	13(17)	14(24)	20(19)	9(16)	5(21)

Table 6.7b Distribution of plasmy level among populations.

plasmy level	C	H	I	N	G	S
homoplasmic	16	15	15	24	10	5
biplasmic	17	18	29	20	23	17
multiplasmic	9	4	10	9	9	22

Table 6.8 Distribution of locus III size alleles among categories and populations. The total number of locus III restriction fragments has been pooled in each population for each category.

	population					
size category	**C**	**H**	**I**	**N**	**G**	**S**
S	18	11	30	31	14	36
M	40	45	51	41	44	53
L	21	8	24	19	27	24

6.2.2.4 Restriction site polymorphism

Individuals were scored for composite restriction cleavage pattern for four enzymes, Eco RI, Sph I, Sal I, and Stu I, as described in Chapter 4. Nine composite restriction morphs were found, each representing a site loss or gain for one enzyme in conjunction with the most common pattern for the other three emzymes. Of the nine morphs five were unique, occurring once only in one population, and a sixth occurred once in each of two populations (Table 6.9). Population N had three unique patterns, and C and I each had one. The composites A and D alone were present in all populations, although pattern D occurred only in an A/D heteroplasmic individual in population I. The restriction morph diversity index $\hat{H}$ shows almost a four-fold difference in range, from 0.1401 - 0.5632.

Only Sph I exhibited a high degree of polymorphism. Two of its four cleavage patterns were present in all six populations, and a third pattern occurred in five. It was the only enzyme involved in restriction site heteroplasmy. The distribution of Sph I patterns among populations, excluding the unique instance of pattern D and all heteroplasmics, was highly heterogeneous ($G=78.196$, $df=10$, $p<.001$) and this was clearly due to the high frequencies of patterns B and C in populations N and C respectively. In pairwise comparisons each of these populations is significantly different from all others. In addition, I differs from G.

Table 6.9 Restriction morph frequencies within populations. The composite letter designations are for Eco RI, Sph I, Sal I, and Stu I respectively.

	population					
morph	**C**	**H**	**I**	**N**	**G**	**S**
A (AAAA)	0.546	0.875	0.928	0.666	0.814	0.886
B (BAAA)	0	0.025	0.018	0	0	0
C (CAAA)	0	0	0	0.019	0	0
D (ABAA)	0.045	0.075	0.018	0.277	0.140	0.045
E (ACAA)	0.386	0.025	0.018	0	0.046	0.069
F (ADAA)	0.023	0	0	0	0	0
G (AABA)	0	0	0	0.019	0	0
H (AACA)	0	0	0	0.019	0	0
I (AAAB)	0	0	0.018	0	0	0
$\hat{H}$	0.5632	0.2334	0.1401	0.4875	0.3232	0.2130

6.2.2.5 Composite mitotypes

The term mitotype has been used in preceding sections to designate a complete description of character state at a locus. It may also be used to describe states at all loci combined, or a subset of loci.

The size mitotype (M_S) represents a combination of the three size variable loci. Locus I has 14 character states taking heteroplasmy into account, locus II has 7, and locus III has 26. In the total sample of 262 scallops, 95 composite size mitotypes were found (Table 6.10a). Of these, 57 (60%) were restricted to single populations with 55 being unique within those populations. Only 2 of the 95 mitotypes occurred in all populations, and the remaining 36 were distributed randomly so that comparisons between populations suggested that no pair was more or less similar than the others (Table 6.10b). Size mitotype diversities were very high, ranging from 0.9469 in population H to 0.9893 in population S, with an overall value of 0.9757.

The total mitotype (M_t) includes restriction site character state with size states. With site heteroplasmy there were 10 restriction site states, and these brought the number of mitotypes to 134. Only 31 mitotypes occurred in more than one population; only 1 was present in all six populations. The remaining 103 (77%) were confined to single populations and 98 of them were unique. The number of mitotypes shared by any two populations was reduced, most noticeably in pairs including C or N because of the amount of restriction site polymorphism in those populations. Total mitotype diversities were increased to a range of 0.9655 (H) to 0.9919 (G), with an overall diversity of 0.9851.

Table 6.10a Combined mitotype diversity among populations. Here mitotype is a complete description of character state at the three size variable loci together (M_S) and of size and restriction site loci together (M_t). N=number of individuals, $\hat{H}_S$=diversity M_S, $\hat{H}_t$=diversity M_t. Numbers in brackets refer to number of mitotypes unique to the population.

	population						
	C	H	I	N	G	S	T
N	38	35	50	53	42	44	262
M_S	28(11)	21(6)	35(13)	29(8)	32(7)	35(7)	95
$\hat{H}_S$	0.9712	0.9469	0.9800	0.9747	0.9859	0.9893	0.9757
M_t	33(23)	23(8)	36(17)	36(22)	36(15)	36(18)	134
$\hat{H}_t$	0.9891	0.9655	0.9812	0.9761	0.9919	0.9904	0.9851

Table 6.10b Number of shared mitotypes between populations. Numbers above the diagonal represent size mitotypes (M_S) shared, below the diagonal total mitotypes (M_t) shared.

	C	H	I	N	G	S
C	-	10	10	10	10	9
H	5	-	11	7	11	8
I	5	10	-	12	16	15
N	5	7	9	-	11	12
G	6	11	14	7	-	14
S	6	8	13	8	10	-

6.2.3 Associations among characters

There are two basic hypotheses concerning the association of mitochondrial and organismal characteristics that can be tested with these data. The first is that mitotype is correlated with the age of the animal, either due to selective mortality or to within-generation evolution. The second is that mitotype is correlated with growth.

Animals were aged by counting growth rings on the resilium and the valves (Merrill *et al.* 1966). A one-way ANOVA indicated that the age structures of populations were different and the samples should not be pooled as a single population ($F_{5,275}$=53.780, $p<0.001$). For correlation analyses animals were put into three groups: 1) under 4 years, 2) 4-8 years, 3) 9 years and older. This classification was adopted partly to reduce error, especially in older scallops where annual marks can be difficult to read accurately, and partly as a reflection of the age structure resulting from the fishery. Heavily fished areas support few older animals. In order to eliminate differences among sample sites partial correlations among variables were obtained, representing the correlation of the residuals of the dependent variables after each had been regressed on the independent variable.

This method indicated that locus I copy number (Table 6.11) was negatively correlated with age (all heteroplasmics excluded, n=244, r=-0.175, $p<0.01$), and that locus III plasmy level (Table 6.12) was positively correlated with age (n=256, r=0.173, $p<0.01$). Locus I heteroplasmy did not show a significant correlation with age (r=0.083, NS). A higher mortality rate for individuals with a higher locus I copy number is the

most likely explanation for the negative correlation of copy number with age, rather than intra-individual mutation to and selection for a lower copy number, but within-generation evolution may be significant at locus III.

Growth was recorded as shell height. Because growth rates are known to vary with the local environment depending on such factors as sea temperature and quality and quantity of food available (MacDonald and Thompson 1985a, Shumway *et al.* 1987), the method of partial correlations was used to examine associations between mitochondrial characters and growth rate as well. None of the correlations of shell height with locus I copy number ($r=-0.036$, NS), with plasmy level at locus I ($r=-0.059$, NS) or with plasmy level at locus III ($r=0.062$, NS), after regression on both age and population, were significant.

Table 6.11 Tabulation of locus I copy number by population and age group, excluding all heteroplasmic individuals.

population	age group	copy number 2	3	4	5	6	7	8
	1	0	1	2	4	0	0	0
C	**2**	1	1	16	6	1	1	0
	3	0	0	5	0	0	0	0
	1	0	0	1	0	0	0	0
H	**2**	1	2	7	1	0	0	0
	3	0	5	15	3	1	0	0
	1	0	1	1	1	0	0	0
I	**2**	0	3	19	2	0	0	1
	3	1	6	7	6	2	0	0
	1	0	4	18	7	1	2	2
N	**2**	0	0	2	4	0	0	0
	3	0	0	0	0	0	0	0
	1	0	5	13	6	0	0	0
G	**2**	0	6	8	0	1	0	0
	3	0	0	0	0	0	0	0
	1	0	0	1	0	0	1	0
S	**2**	0	8	4	2	2	0	0
	3	0	8	13	1	1	0	0

Table 6.12 Tabulation of locus III plasmy level by population and age group.

population	age group	plasmy level 1	2	3	4	5	7
	1	2	5	0	0	0	0
C	2	10	11	5	1	0	0
	3	3	1	1	1	0	0
	1	1	0	0	0	0	0
H	2	3	7	1	0	0	0
	3	11	11	2	1	0	0
	1	1	2	1	0	0	0
I	2	9	15	3	0	1	0
	3	5	12	5	0	0	0
	1	13	14	6	0	0	0
N	2	2	3	2	0	0	0
	3	0	0	0	0	0	0
	1	7	14	4	0	0	0
G	2	3	9	2	2	0	0
	3	0	0	0	0	0	0
	1	0	1	0	0	0	0
S	2	2	8	7	1	0	0
	3	3	8	10	3	0	1

6.3 Discussion

6.3.1 Overview

As has been discussed in previous chapters, certain features distinguish each of the three incrementally varying loci in *P. magellanicus* mtDNA. Locus I, with an array of long perfect direct repeats, is notable for its high diversity (0.639) but relatively low frequency of heteroplasmy (6.4%). Locus II varies by a smaller increment and has but two major alleles, with a low incidence of heteroplasmy (3.2%) and two of the four rare alleles occurring only in heteroplasmy. Locus III has the smallest unit of variation, many alleles, and a very high frequency of heteroplasmy (68.8%). These characteristic differences suggest that each locus may provide a different kind of information about population subdivision or mitochondrial evolutionary dynamics. A brief summary of observed variation at the three loci is presented here prior to a discussion of the geographic structure of the scallop population and the population genetics of scallop mtDNA.

Comparing locus I with length polymorphic systems in other animal mitochondrial genomes, two features stand out. The first of these is the occurrence of all size classes in homoplasmy, and the observation that even the rare size classes are more frequently homoplasmic than heteroplasmic. In a survey of natural populations of the American shad, Bentzen *et al.* (1988) found two size classes with an increment of change comparable to that of the scallop at 1.5 kb, and the larger variant was only seen in heteroplasmic individuals. In populations of *Gryllus* crickets Rand and Harrison (1989) described seven size classes of which only four (S,M,L,VL) occurred in homoplasmy and two of those (L,VL) were homoplasmic in single individuals. When samples were grouped as pure *G. pennsylvanicus* or *G.*

firmus, and hybrid clonal lineages, four of the five groups held homoplasmic individuals only for classes S and M. The worldwide survey of isofemale lines of *Drosophila melanogaster* conducted by Hale and Singh (1986) revealed thirteen size classes with a complex pattern of variation, and twelve of these occurred at least once in homoplasmy although several were rare and more frequently heteroplasmic. Size polymorphism reported in the pine weevils (Boyce *et al.* 1989) was extensive but not a single homoplasmic individual was found for any size class. *P. magellanicus* appears to be exceptional in the distribution of homoplasmic size classes within and among populations.

As a corollary to this, the scallop is also exceptional in the number of size classes present in geographic populations. Of the seven classes, three populations had six, two had five, and one had four classes, and as noted each class present was homoplasmic at least once in that population. In cricket populations, three size classes were typical and much of the variation was within rather than among individuals. The mean intra-individual diversity measure K_b range was an order of magnitude lower in scallop populations than in crickets, while the population diversity K_c range was about double that of crickets. Clearly these observations demonstrate that each locus I size class in the scallop is viable and suggest that either purification of heteroplasmic lines is rapid or that size states are relatively stable with a much lower mutation rate than in *Gryllus* for example (μ= 1-2 x 10^{-4} per animal generation, Rand and Harrison 1989). Geographic populations of *D. melanogaster* showed a maximum of six of the thirteen size classes, and the shad river populations revealed a latitudinal cline, where southern rivers included both size classes and northern rivers, with one exception, did not.

Ignoring migration, new homoplasmic lines appear in a population through loss of intra-individual variation. Transmission data from crickets (Rand and Harrison 1986) indicate a bias toward the smaller genome in offspring of heteroplasmic females, and the frequency distributions from natural populations show that although mutation is bidirectional only classes S and M are fixed in individuals, with two exceptions. In contrast, the frequency distributions in scallop populations show that fixation is frequent in both larger and smaller classes.

Locus II also shows two peculiarities when compared to other length polymorphic systems; the two common alleles are non-adjacent size classes, and the range of diversity among populations is large. The increment of variation, 250 bp, is similar to that seen in some insects (Hale and Beckenbach 1985, Solignac *et al.* 1986, Rand and Harrison 1989) but the pattern of variation is quite different. As in *Gryllus* , only four size alleles were seen in homoplasmy but these did not represent what might be seen as the intermediates in a forward and backward shuffling among size classes. In fact, the size extremes at locus II were more frequent than two of the intermediate classes and the largest allele, F, was only observed in homoplasmic individuals. Of nine heteroplasmic individuals, not one contained adjacent alleles. The large range of diversity is partially due to one population, N, being nearly fixed for one allele (D) and in this it resembles the shad where the smaller variant was fixed in northern populations (Bentzen *et al.* 1988). However, even eliminating this population the range remains almost threefold, not only because of frequency differences of alleles B and D, but also because of the occurrence of other alleles in two populations. Indeed, the observation of all recorded alleles in a single

population, as is the case for population C at locus II, has not been reported for any system with more than two alleles.

Polymorphism at locus III is very complex. The small increment of variation most closely resembles the short tandem repeats described in vertebrate mtDNA control regions (lizards, Densmore*et al.* 1985; stickleback, Gach and Reimchen 1989; sturgeon, Buroker *et al.* 1990), as do the range of variation, encompassing many alleles, and the frequency of heteroplasmy. Under the conditions of the present study individual allleles could not be characterized and variants were grouped into S, M, and L size categories making certain comparisons impossible. However, within-category heteroplasmy was more common than between-category heteroplasmy, suggesting that mutation to adjacent classes may be the most frequent kind of event. Homoplasmic individuals for each category were noted in all populations with the single exception of category L in population S.

Each locus showed spatial heterogeneity in the distribution of genome size among populations with significant differences between several, but not the same, pairs of populations. The question that arises is whether this kind of variation might be useful in population discrimination or whether it might better describe current characteristics of populations regardless of their genetic affinities.

6.3.2 Population subdivision

In order to understand the relative contributions of historical zoogeography and of isolation or restricted gene flow to the genetic structure of scallop populations, more data are needed on the ecology and reproductive

biology of the species. The lifestyle of *P. magellanicus* , with large aggregations of relatively sessile and long-lived iteroparous adults and pelagic larvae, suggests that larval dispersal is the main conduit of gene flow. Despite large population sizes breeding adults occupy neighbourhoods defined by the dispersal limits of viable gametes released by simultaneously spawning individuals, analogous to pollen dispersal in plants (Van Dijk 1987). The mobile juvenile phase following metamorphosis perhaps adds a second level of gene flow in promoting migration among neighbourhoods. Demographic data concerning movements of larvae and juveniles would help to clarify the independence or interdependence of major scallop beds in recruitment to the breeding population, but in the absence of these data genetic analysis may permit some inferences regarding the relative isolation of geographic populations. Ideally, local differentiation would lead to diagnostic markers that clearly identify the genetic origin of an individual or sample and could be used to determine the boundaries or pathways of gene flow.

Little information is yet available on spawning cycles in *P. magellanicus*. Naidu (1970) discussed the likelihood of sea temperature change being the stimulus for gamete discharge, noting that in Newfoundland scallops the gonads appeared ripe for an extended period beginning in late June but the major spawn did not occur until late September or October. He speculated that acclimation to warming temperatures was rapid enough to preclude a major spawn but that autumn cooling provided a temperature shock sufficient to result in the discharge of gametes. Different temperature regimes in various parts of the species range might then result not only in different spawning times but also in temporal isolation of populations if

subsequent spat settlement is influenced by temperature. Gene flow between early and late spawning populations could be impeded by temperature restrictions on metamorphosis. An important correlate of any population discrimination based on mtDNA size should be allozyme and restriction site polymorphism genetic distances or differentiation which, although perhaps less sensitive over the relatively short post-glacial period, are more likely to be strictly neutral.

Many studies of genetic structure of populations have used the standardized variance in alleles frequencies among populations to assess levels of gene flow (Wright 1978, and see Slatkin 1985). Under the neutral hypothesis local diffentiation will result from random drift and is counterbalanced by gene flow (Kimura 1983). However, strong local selection at a locus or subset of loci can mimic the effect of drift and lead to the inference that gene flow is very small or absent (Slatkin 1987). In the scallop, the spatial distributions of allele frequencies show little differentiation with the exception of the *Odh* locus where different alleles occur at high frequencies in different populations. This pattern suggests that selection may be acting at this locus or one closely linked to it, and that despite significant heterogeneity of allele frequencies among populations with all loci combined the neutral hypothesis of no differentiation should not be rejected. This does not mean that there is a single large panmictic population nor that the migration rate is necessarily high, as a rate as small as the inverse of local population size is sufficient to overcome drift (Wright 1931). The pattern could equally well be explained as incipient differentiation following the relatively recent range expansion at the end of the last glaciation.

The mitochondrial genome is considered less sensitive to the effects of migration due to maternal inheritance and transmission as a clonal lineage (Wilson *et al.* 1985, Birky *et al.* 1989), and for this reason it is a more sensitive indicator of local differentiation. In the majority of population surveys where mtDNA is the genome sampled, restriction site polymorphism provides the basis for inferences about population structure (see Avise *et al.* 1987, Moritz *et al.* 1987, Harrison 1989). In this study, individuals were scored for only four restriction enzymes and of these only one, Sph I, was highly polymorphic.

The limited restriction site data from the six scallop populations suggests two things. First, excluding the major Sph I variation, most of the site variants are unique and confined to single populations; only the Eco RI site loss (Eco RI pattern B) was seen in individuals from two populations, which are geographically contiguous. If most of the variation predates the repopulation of the region, it has subsequently been lost. If the mutations occurred after repopulation, they have either not spread among populations or not become established in most populations since the colonization. It is conceivable that colonization occurred as a series of founder events rather than overlapping waves of immigrants and that the initial frequencies of site variants differed among founder populations, but unlikely that the subsequent patterns of loss of variation would be the same.

Second, the pattern of Sph I site variation casts a different light on population differentiation than any other marker. Just two populations (C,N) show substantial differentiation, both in the frequencies of Sph I cleavage patterns yet for different patterns. All populations had the Sph I A pattern in highest frequency, and four (H,I,G,S) included both the variant patterns Sph

I B and Sph I C in low frequencies. In two of those four populations, H and I, the variant Sph I C was observed in only a single individual and in I the variant Sph I B was seen only in an individual heteroplasmic for patterns A and B. In population N, not a single variant Sph I C was observed while Sph I B had a frequency of 28%. In population C, Sph I B was represented by one homoplasmic individual and one individual heteroplasmic for patterns A and B while Sph I C had a frequency of 40%. Whether Sph I polymorphism was present at colonization or not, the current distribution suggests that gene flow has been very limited at least between populations C and N despite the observation that these two populations closely resemble one another in other ways. Both of these populations occupy limits of the species range, in the case of C a geographic limit as the Baie des Chaleurs is a pocket of the Gulf of St. Lawrence and in the case of N a biological limit at the cold extreme of the temperature range for the species, and this may have contributed significantly to their genetic isolation.

At locus I the six populations fall into three groups based on the relative frequencies of classes 3 and 5. Two populations (C, N) have low frequencies of 3 and high frequencies of 5, one (S) has a high frequency of 3 and low frequency of 5, and the remaining populations (H,I,G) have intermediate frequencies of both classes. If these differences are historical, resulting from founder effect in the postglacial repopulation of Atlantic Canada by scallops, or if they are a result of drift in isolated subpopulations, then at least some concordance with allozyme or restriction site differentiation should exist. On the other hand, if gene flow is high and differentiation is deterministic then locus I distributions may be a

consequence of environmental and life history parameters and show no agreement with other genetic markers.

Although the genetic distances calculated from allozyme data were small, in three pairs of populations they were significant (CH, SH, SI). These pairs fell into the three groups discussed above for locus I, but did not include either of the pairs (CS, NS) showing significant differences in locus I size class distribution. If indeed nuclear differentiation is primarily due to selection at the *Odh* locus, the important role that this enzyme plays in scallop energetics (Volkaert 1988) makes it a good candidate for cytonuclear interactions. More extensive sampling would be required to detect an association between *Odh* genotype and mitochondrial genome size, but the possibility of such an association should be kept in mind when considering the spatial distributions of the two markers.

The patterns of differentiation suggested by restriction site polymorphism do not accord with the locus I groups but place populations C and N each in isolation and separate I and G. This reflects the higher site diversity in C, N, and I as well as the distribution of morphs, and contrasts with the pattern for locus II where significant differences resulted from the near homogeneity of population N. Although C was notable for the number of size alleles present at locus II, including three private alleles, it differed only from N where a single exception to the common size class was observed. N differed also from I and G. How can this pattern in population N of relatively high diversities for nuclear loci, mtDNA locus I, and restriction morphs in conjunction with a very low relative diversity at locus II be explained? The near fixation of size allele D may characterize this population but it does not indicate any current or historical patterns of gene

flow beyond buttressing the conclusion that populations C and N have not exchanged mitochondrial genomes frequently. Some possible molecular explanations will be considered in the following section.

The extraordinary diversity of mitotypes and frequency of heteroplasmy at locus III suggest that that the mutation rate at this locus is too high to allow any differential segregation to be discerned. Nevertheless, two aspects of population level differentiation were recorded: distribution of homo-, bi-, and multiplasmic individuals, and distribution of genome size category. It is not clear how plasmy level itself could be an evolving character unless the generation of variation is under genetic control, in which case it would resemble genetic control of mutation rates or recombination rates in natural populations. A simpler explanation for heterogeneity of plasmy level would be ecological or demographic, where environmental conditions might influence mutation rates and demographic parameters such as age structure and age-specific fecundity might contribute to differentiation.

The distribution of size categories resembles locus I in that some populations appear to show a different direction of mutation or segregation, with differential accumulation of variants smaller or larger than the intermediate and most frequent category, but at locus III the population with a high frequency of larger variants was G in contrast to C and N at locus I and those with a high frequency of smaller variants were N and S as opposed to G and S at locus I. Clearly the two loci are not behaving in concert. The observation of differentiation indicates that mutation is not swamping other processes at either locus, but the problem of what forces are responsible for differentiation cannot be easily addressed. If individual alleles could be

identified and patterns of allelic heteroplasmy, including relative frequencies of variants, recorded then it is possible that categorical differentiation would prove illusory but equally possible that the more specific patterns of variation could provide insight into the mechanisms of variation at both individual and population levels.

The amount of mtDNA variability catalogued by multilocus mitotype indicates that divergence between mitochondrial populations defined as individual organisms is extensive. The apparent independence of each locus in its patterns of variation has resulted in a set of markers that can almost completely distinguish individuals but does not reveal any natural groupings at the population level. The availability of well-characterized, completely linked but distinctive markers holds great promise for investigations of the population dynamics of mitochondrial evolution but would appear to be of little use as markers of scallop population structure.

The two major conclusions emerging from the examination of geographical structuring of the *P. magellanicus* population in Atlantic Canada are: 1) there is abundant evidence of local differentiation but the patterns for different markers are not concordant and the bases are likely a combination of stochastic and deterministic forces; and 2) mtDNA size polymorphism is not a useful marker for population structure. An extension of the restriction site polymorphism survey to include more enzymes, particularly those lacking cleavage sites within the long tandem repeat of locus I, would likely be the best method of estimating divergence among local populations. Wider sampling of scallop beds with known environmental parameters, such as water depth, temperature range, and degree days, would be useful in evaluating the role external factors might play in the evolution of

genome size. As well, estimates of variance at spatfall and of different year classes in sample locations could indicate the source or direction of local differentiation.

6.3.3 mtDNA evolution

The negative correlation between age and locus I copy number suggests that there may be selection against longer mtDNA molecules, but does not indicate whether selection occurs at the level of mitochondrial populations or individual populations. The mtDNA transmission data from crickets led Rand and Harrison (1986) to propose that selection against longer tandem arrays occurred within mtDNA populations as the result of a "replication race" between molecules of different lengths. The locus I repeat unit in *Placopecten* is seven times as large as the cricket repeat whereas the overall genome length is only twice as large, and a simple replication race model would therefore predict an even more pronounced effect in the scallop. Although no transmission data are available from *P. magellanicus* , other observations suggest that the loss of intra-individual variation is both more rapid and less directional than in the cricket.

It is conceivable that syncytial development in insects retards drift in embryonic development by removing one level of segregation, the partitioning of organelles during cytokinesis. A further explanation for more rapid purification in scallops than in insects could relate to differences in mitochondrial dynamics during oogenesis. Scallop eggs are much smaller in relation to maternal biomass and in absolute size than insect or vertebrate eggs and probably experience less mitochondrial amplification. The effect of drift in these smaller mitochondrial populations would be more pronounced.

It is difficult to distinguish rapid loss of variation from infrequent generation of variation without transmission data, but although the rates of these processes might vary among populations only the loss of variation should be subject to differences in direction. The observations of homoplasmy for even the largest size class in conjunction with the equal frequencies in the population of size classes larger and smaller than the mode of four repeat copies imply that selection against longer genomes is more likely occurring at the individual level than at the mitochondrial level in *Placopecten.*

In their study of mtDNA inheritance in heteroplasmic isofemale lines of *Drosophila* , Solignac and co-workers (1987) demonstrated an age effect in the change in variance and genome length. With increasing maternal age both variance and average mitochondrial genome length in offspring increased, implying that there was a bias toward longer molecules in late eggs. This observation again suggests that differences in replication rate among mtDNA molecules underlie the segregation of intra-individual variation, but moves the level of selection from among molecules within mitochondria to among oocytes within ovarioles or ovaries. The race is effectively an "amplification race", with oocytes that have higher frequencies of long genomes being slower to mature, and the evolution of genome size is thereby influenced by age-specific fecundity.

In the scallop, mature eggs are released in a single spawn each year and an amplification race could only be detected by comparing stem cells before oogenesis with mature ova, or by examining the change in variance and mean genome length in the eggs of consecutive yearly spawns. Leaving aside the mechanism contributing to the correlation of lower copy number with increasing age, a tendency for older scallops to have smaller mitochondrial

genomes could be an influential factor in the evolution of this character because age-specific fecundity is an important parameter in scallop demography (MacDonald and Thompson 1985). Fecundity increases with body size, and therefore with age, so that older females contribute proportionately more gametes to the population pool.

Is there evidence for an age effect? The significant difference in locus I copy number distribution between the St. Pierre Bank (N) and Ship Harbour (S) samples could be related to the age structure of these populations. St. Pierre Bank (N) supports a commercial fishery and the larger animals are harvested, leaving predominantly younger scallops as breeding stock. Few individuals in the N sample were older than four years. The Ship Harbour sample (S) came from a small local population and included many animals older than eight years. The greater fecundity of older females may have contributed to the high frequency of smaller genomes in Ship Harbour, and the removal of older females from St. Pierre Bank may have allowed larger genomes to increase in frequency. Georges Bank (G) also supports a large commercial fishery, and that sample, like N, comprised mostly young animals yet it did not differ from S in copy number distribution. The Baie des Chaleurs sample (C) included scallops from all age classes but did differ from S in copy number distribution. What do these apparent contradictions mean? Only N and C have useful restriction site markers, and these indicate that the two populations are genetically isolated from one another and that perhaps each has a low emigration rate, but they do not give information about immigration rates. A regular or sporadic influx of larvae from other populations could mask the evolutionary effects of age-specific fecundity. Alternatively, there may be no age effect.

Sampling of cohorts and of larvae and spat in small local populations could indicate whether copy number distribution differs among year classes and if so, whether such differences correlate with recruitment patterns.

The observation of different mitotypes in different tissues of the same individual, as described in Chapter 5, can be explained in two ways. Inherited variation may be lost in some tissues during embryogenesis, or through mitochondrial turnover and cell division during the life of the animal. Alternatively, *de novo* mutation and rapid increase in the frequency of the mutant genome may create variation in some tissues. Both processes can result in offspring with different mitotypes from year to year. If mitochondrial populations within females are evolving rapidly then perhaps each spawn should be considered as a separate lineage. A measure of within-generation evolution might be obtained by sequential sampling of tissues and gametes from individuals for a period of years. This information is important not only because of its bearing on mtDNA transmission but also because any selection argument at the individual level must take into account the somatic component of mtDNA variation.

The positive correlation between age and plasmy level for locus III could be due to *de novo* mutation, an individual level process, or to selective mortality, a population level process. If homoplasmic individuals experience selective mortality there should be a pattern of increasing heteroplasmy with age, but it is difficult to see why heteroplasmic animals would be more viable. At locus III, with its small increment of variation and possible location in the control region of the molecule, plasmy level could increase simply as a function of time, through the accumulation of errors that would have a minor effect on overall genome size.

Many of the outstanding questions about locus I can be addressed by more extensive spatial and temporal sampling of populations and individuals: why is diversity so high yet the frequency of heteroplasmy so low? is paternal leakage a significant factor? are there environmental or demographic correlates that might indicate whether selection plays a role in the evolution of array length? can *de novo* mutation be demonstrated?

More molecular data are needed for locus III. Sequencing of the region would determine whether it comprises a tandem array and could provide the means for a more definitive evaluation of intra- and inter-individual variation if it revealed a tighter bracketing of the locus by restriction enzymes. If the variable region could be assayed in a smaller fragment, size alleles might be differentiable among individuals. The better gel resolution of small fragments would make densitometric estimation of intra-individual frequencies practicable, and this would be especially useful for individuals that are multiply heteroplasmic as it would reveal any association between variants at different loci. Finally, the question of whether locus III lies within or adjacent to the control region might be resolved.

Chapter 7
General Conclusions

In the course of this study I have recorded the kinds and amounts of mtDNA length polymorphism present in natural populations of the scallop *Placopecten magellanicus*. The range, the number of kinds, and the diversity of variation within and among individuals are each exceptional, and together they constitute a system of greater complexity than any other characterized animal mitochondrial genome.

Length polymorphism that presents as a series of size classes differing by discrete and predictable amounts is usually interpreted as a tandem array of repeats, and is often characterized by a high frequency of backward and forward mutation among a limited number of array sizes (Moritz *et al.* 1987, Birky *et al.* 1989). The two factors of high mutation rate (relative to point mutations) and finite number of states are the basis for the frequent observation of heteroplasmy in such systems. Because *P. magellanicus* has three distinct incrementally varying loci, each with its own mutation rate and frequency of heteroplasmy, simultaneous heteroplasmy for two and three loci was observed in several individuals. In species with a high frequency of heteroplasmy, such as *Gryllus* crickets (Rand and Harrison 1989) and pine weevils (Boyce *et al.* 1989), and for locus III in *P. magellanicus* itself, multiplasmy (simultaneous occurrence of more than two variants for one locus) is often observed, but multiple heteroplasmy (simultaneous occurrence of heteroplasmy for more than one locus) has been reported only once, in three individuals of the scallop species *Chlamys islandica* (Gjetvaj 1989).

Multiple heteroplasmy in *P. magellanicus* extends beyond length polymorphic loci. Four instances of heteroplasmy for a restriction site polymorphism were observed, and three of them were associated with length heteroplasmy. Site heteroplasmy has been rarely documented, and only the Sal I site heteroplasmy in shad (Bentzen *et al.* 1988) occurred in several individuals from widely scattered locations as is the case for *P. magellanicus*. However, this is the first report of simultaneous site and length heteroplasmy, and it is noteworthy also for the fact that site heteroplasmy was distributed both within and between locus I variants. Such a pattern suggests that paternal leakage may be a factor in the maintenance of variation.

The amount of variation in scallop mtDNA provides a unique opportunity to study segregation patterns both within and between generations. I found that there were sometimes differences in mitotype among tissues from single individuals, demonstrating that within-generation evolution had occurred but not indicating the underlying mechanism. In general, the high diversity and low incidence of heteroplasmy for locus I imply that both mutation and segregation are rapid; segregation processes might best be monitored at locus III, especially if size alleles could be defined and associated with their respective locus I markers in double heteroplasmics. Sequential sampling of a cohort from larval through adult stages might indicate the relative importance of drift in oogenesis, embryogenesis, and mitochondrial turnover as a contributing factor to mitotypic change.

These kinds of data could also answer the important questions of whether the germ line mitotype is constant throughout a female's

reproductive life, and whether selection is involved in any mitotypic changes within individuals. Until these are resolved, the basis for mitochondrial differentiation among populations will remain unclear. Significant differences between populations in the distribution of variation at each locus were found, but there was no consistent pattern (Table 7.1). With the largest proportion of the variance being among and within individuals rather than among populations, length polymorphism is not a good marker for studies of population structure but it is an excellent marker for studies of the population genetics of mtDNA lineages.

Evolutionary processes within the population of molecules in a mtDNA lineage and within the population of individuals together determine the tempo and mode of mitochondrial evolution in a species. The novel and unusual features of the mitochondrial genome of the scallop *Placopecten magellanicus* can contribute significantly to the understanding of mitochondrial evolution in the metazoa.

Because it represents an extreme in genome length even after excluding size variable loci, the scallop mitochondrial genome could provide insights into transcription processes. Knowledge of the distribution of coding sequences between strands and the mechanisms of transcription and transcript processing, including the positions and number of promoters and the number of mature transcripts produced, is needed to understand how such a large genome operates efficiently. The study of exceptions often leads to a better understanding of the rule, and in this case elucidating the functional aspects of scallop mtDNA could clarify some of the constraints on the standard stream-lined animal mtDNA.

Although tandem duplications and tandem arrays have been reported in or adjacent to the mtDNA control region in many species, there has been little speculation on the effect they may have on transcription initiation and elongation. Attardi (1985) reported differences in both transcription rates and transcript stabilities between H- and L-strands in HeLa cells; possibly the internal structure of primary transcripts has a large effect. In *P. magellanicus*, with its complex patterns of tandems arrays and dispersed sequence similarity, questions regarding transcription processes and metabolic efficiency might be addressed with relative ease due to the availability of individuals having large differences in mtDNA size at readily defined loci.

Despite the large amount of variation imposed by the three incrementally varying length polymorphic loci and the three regions exhibiting discrete deletions, the genome arrangement appears to be stable. Whatever mechanisms underlie length mutation, they do not seem to affect other regions although two findings suggest that rearrangements may have been involved in the origins of length polymorphism. In the case of locus I, the distant fragment Sph I-C/Pst I-A shows sequence similarity with part of the repeated element and with one of its flanking sequences (LaRoche *et al.* 1990). This resembles the arrangement described in the nematode *Romanomermis culicivorix* by Hyman and co-workers (1988), where part of the 3 kb repeat was found in inverted orientation in a region distant from the repeat array. The other suggestive observation is the sequence similarity between some part of locus II and several restriction fragments showing length variation. Sequencing could reveal the extent of similarity, and perhaps the contrast in the behaviour of the different fragments (some

incrementally varying, some having single defined deletions, and some length stable), in conjunction with knowledge of their hybridizing sequences and the flanks, would provide insight into the origins and mechanisms of length polymorphism.

Gjetvaj (1989) found that size variation and size increase were general, but not universal, features of pectinid mitochondrial genomes. He described incremental variation in several species, and in hybridization experiments could not show sequence similarity between repeat regions among species. *P. magellanicus* represents an extreme in mtDNA length and complexity within the pectinid lineage; a comparative study of scallops detailing gene arrangement and relative positions of length variable loci, as well as life history patterns, and physiological and environmental constraints, might indicate what makes the genome susceptible to size alterations and why the changes can be tolerated.

Table 7.1 Pairwise differences between populations. Each population pair was tested in a contingency table, and significant differences ($p \leq 0.05$) for each marker are denoted by a dot.

	locus I	locus II	locus III	Sph I
CH			•	•
CI				•
CN		•		•
CG				•
CS	•			•
HI			•	
HN			•	•
HG			•	
HS			•	
IN		•		•
IG				•
IS				
NG		•	•	•
NS	•			•
GS			•	

Appendix A

Identification numbers of scallops by population (C, H, I, N, G, S), with physical data and nuclear genotypes. Age is given in years, weight in grams, shell height (D/V) in millimeters. Coding for nuclear alleles is as follows:

Pgm	Odh	Pgi	Mpi	Aat	Pgd
3 96	4 98	4 87	3 75	1 50	4 80
4 98	5 100	5 100	4 87	2 70	5 100
5 100	6 105	6 115	5 100	4 90	6 130
6 102		7 140	6 108	5 100	7 150
				6 105	
				7 115	

C	sex	age	wt	D/V	Pgm	Odh	Pgi	Mpi	Aat	Pgd
1	M	.	.	.	64	66	55	55	44	55
2	F	.	.	.	54	55	55	55	55	55
3	F	.	.	.	56	55	55	44	54	55
4	M	.	.	.	56	66	55	44	54	55
5	M	.	.	.	55	55	55	44	54	55
6	F	.	.	.	54	56	55	55	55	57
7	M	4	9.2	87	56	56	55	55	55	55
8	F	7	10.4	105	55	56	55	55	57	77
9	M	5	7.0	90	54	55	55	54	44	55
10	M	6	11.6	95	55	66	55	55	41	57
11	M	5	6.9	86	55	56	55	54	55	57
12	M	4	6.1	83	56	55	55	55	55	57
13	M	7	9.0	97	54	55	55	55	44	57
14	F	7	7.4	91	54	56	55	55	44	55
15	F	7	9.4	96	55	66	55	55	44	55
16	F	10	11.7	108	55	56	55	55	54	55
17	M	7	8.9	100	66	56	56	55	54	55
18	.	6	9.0	91	54	66	55	55	54	55
19	M	5	10.1	90	44	55	55	55	54	55
20	F	5	8.9	95	66	56	55	44	55	57
21	M	6	7.6	93	55	56	55	54	55	55

C	sex	age	wt	D/V	Pgm	Odh	Pgi	Mpi	Aat	Pgd
22	F	10	25.7	125	55	56	55	54	54	55
23	M	7	11.9	98	56	55	55	54	44	57
24	F	6	16.7	114	64	56	55	55	54	55
25	M	5	7.6	90	66	56	56	54	54	55
26	F	5	7.7	92	55	56	55	55	54	55
27	F	5	9.2	91	55	56	55	54	44	55
28	M	4	7.5	86	55	55	55	54	44	55
29	F	4	6.4	81	55	55	55	54	54	55
30	F	4	7.1	86	55	55	56	55	54	55
31	F	6	15.7	110	55	55	55	44	54	55
32	F	7	9.0	97	56	55	55	55	54	57
33	F	6	18.3	114	55	55	55	55	44	55
34	F	13	12.8	116	56	55	56	55	44	57
35	M	5	10.9	96	55	55	55	54	55	57
36	F	8	8.4	99	54	56	55	55	44	55
37	M	6	7.4	90	53	55	55	54	54	55
38	F	5	6.9	96	56	55	55	55	44	55
39	F	4	7.4	81	66	56	55	54	55	56
40	M	5	9.4	91	55	66	55	55	44	55
41	F	6	14.3	111	66	55	55	55	54	55
42	M	6	11.5	103	55	56	55	55	55	55
43	F	12	12.9	124	66	56	55	55	44	55
44	F	13	23.1	135	55	56	55	55	41	77
45	F	11	23.3	131	55	56	55	54	55	55
46	M	4	7.8	84	56	55	55	55	44	55

H	sex	age	wt	D/V	Pgm	Odh	Pgi	Mpi	Aat	Pgd
2	.	10	7.6	114	55	56	55	55	44	57
3	.	10	8.9	118	55	55	55	44	54	55
4	.	7	9.6	105	55	55	55	55	54	55
6	.	9	9.3	108	55	66	56	54	55	57
7	.	10	6.2	107	55	66	55	55	55	55
8	.	6	7.1	104	55	66	56	54	44	57
9	M	10	8.7	110	55	66	55	55	55	77
10	F	10	8.6	113	54	56	56	55	54	57
11	.	11	5.5	106	55	55	55	55	54	57
14	F	10	5.5	114	55	66	55	55	54	55
15	F	11	6.9	119	56	66	55	55	54	57

H	sex	age	wt	D/V	Pgm	Odh	Pgi	Mpi	Aat	Pgd
16	F	7	7.0	103	56	66	55	54	55	55
17	M	10	7.8	100	64	56	55	55	54	57
18	F	10	3.3	107	55	66	55	55	54	55
19	M	9	1.8	103	55	66	55	54	54	57
20	F	6	8.8	112	55	66	55	55	55	57
21	F	9	11.1	112	55	56	56	54	54	55
22	M	10	4.2	109	55	56	57	55	55	57
23	F	11	6.5	107	55	56	55	55	54	55
24	M	10	5.7	107	66	66	55	55	54	55
25	M	10	6.5	104	55	56	55	54	44	55
26	F	10	8.5	106	56	56	55	55	55	55
27	F	9	7.6	109	55	55	55	54	55	55
28	M	10	21.2	126	55	66	55	55	54	55
29	F	5	6.4	108	66	56	55	54	52	55
30	F	10	5.3	103	55	55	55	55	55	57
31	F	10	5.1	110	55	66	55	54	54	55
32	F	10	6.9	100	44	56	55	54	54	57
33	M	7	7.6	106	55	56	55	54	44	57
34	M	10	7.3	106	56	66	55	55	54	55
35	M	10	4.7	107	54	56	55	54	54	55
36	M	7	6.8	103	54	56	55	54	44	55
37	F	6	4.8	103	55	66	55	54	55	55
38	M	9	6.7	103	55	56	55	55	55	55
39	F	7	6.0	98	55	66	55	55	42	55
40	F	5	7.9	105	54	66	55	55	44	57
41	M	8	6.0	103	64	56	55	55	44	77
42	M	4	7.3	103	56	56	55	54	54	55
43	M	10	6.9	97	54	56	55	54	55	55

I	sex	age	wt	D/V	Pgm	Odh	Pgi	Mpi	Aat	Pgd
1	F	8	9.1	126	55	66	55	54	44	57
2	M	10	11.5	124	55	56	55	55	44	55
3	F	7	10.9	119	56	56	56	54	42	57
4	M	6	7.9	107	55	66	55	55	54	57
5	F	7	7.6	113	55	66	55	54	54	57
6	M	10	11.8	125	55	56	55	54	54	55
7	F	5	9.1	112	.	.	.	.	.	.

I	sex	age	wt	D/V	Pgm	Odh	Pgi	Mpi	Aat	Pgd
8	F	8	11.9	119	56	66	55	54	44	57
9	F	9	10.4	121	55	55	55	54	54	55
10	F	11	10.0	112	55	46	55	54	55	55
11	F	6	11.5	117	55	55	55	55	54	57
12	M	8	11.8	122	56	56	55	55	55	55
13	M	10	6.6	120	55	55	55	55	44	55
14	F	10	10.1	120	55	56	55	55	54	57
15	M	9	10.6	123	56	55	56	55	54	55
16	M	9	11.8	117	56	66	55	44	44	57
17	F	7	9.0	116	55	56	55	55	55	55
18	M	10	5.4	116	56	55	55	54	54	77
19	F	5	8.0	113	56	66	55	54	54	57
20	F	10	14.2	119	55	55	55	54	54	57
21	M	10	6.3	130	55	66	55	55	54	55
22	F	10	10.5	118	55	66	55	55	54	57
23	F	12	12.7	120	55	56	55	55	54	57
24	M	8	12.0	115	55	66	66	54	54	55
25	M	8	10.3	116	56	66	55	54	44	55
26	M	10	7.6	115	54	66	55	54	55	55
27	M	5	17.9	110	55	56	55	54	44	55
28	F	5	17.6	118	55	66	55	55	55	55
29	F	5	15.0	108	54	56	55	55	54	55
30	F	9	14.0	118	56	66	55	44	54	55
31	F	8	13.4	110	55	56	55	55	54	55
32	M	10	13.7	115	56	56	55	55	54	57
33	M	4	13.5	110	55	56	55	54	54	55
34	F	6	10.0	115	44	56	55	54	41	57
35	F	10	11.2	113	56	55	55	55	55	55
36	F	8	19.5	119	56	56	55	54	54	57
37	F	7	12.2	111	56	66	55	54	54	55
38	M	9	9.6	104	55	66	55	55	54	55
39	F	10	19.0	128	55	56	55	55	54	55
40	M	10	17.8	108	55	55	55	54	44	55
41	M	10	19.4	120	55	66	55	55	54	77
42	M	9	14.7	111	55	66	55	55	54	57
43	M	6	10.3	106	64	55	55	44	44	55
44	F	7	10.8	104	55	66	54	54	55	57
45	M	8	9.7	102	56	66	56	55	55	55
46	M	4	8.3	101	56	56	55	55	54	57

I	sex	age	wt	D/V	Pgm	Odh	Pgi	Mpi	Aat	Pgd
47	M	8	10.3	95	66	55	56	54	54	55
48	F	7	10.5	101	56	56	55	55	44	55
49	F	9	9.1	109	55	56	55	54	54	55
50	M	6	11.2	105	55	66	55	55	55	55
51	F	8	21.2	119	54	66	55	54	55	57
52	F	3	7.3	92	56	55	55	54	41	55
53	M	4	9.3	100	55	66	55	54	44	55
54	F	6	10.3	106	64	56	55	55	55	57
55	F	5	13.2	100	55	55	55	55	55	55
56	F	7	25.2	123	55	56	55	55	54	55

N	sex	age	wt	D/V	Pgm	Odh	Pgi	Mpi	Aat	Pgd
1	.	.	9.8	84	55	55	55	44	55	55
2	.	.	9.1	82	54	56	55	54	54	57
3	.	.	7.4	78	55	56	55	55	44	55
4	.	.	8.7	88	54	56	55	55	54	57
5	.	.	10.1	85	56	55	55	55	44	77
6	.	.	14.8	92	56	66	55	55	54	57
7	.	.	9.6	84	56	55	55	44	54	55
8	.	.	9.4	80	55	56	55	55	54	55
9	.	.	8.9	77	55	56	55	55	54	55
10	.	.	7.5	78	56	55	55	54	55	57
11	.	.	8.0	75	54	55	55	55	54	55
12	.	.	8.5	80	66	66	55	56	55	55
23	.	.	6.9	82	56	66	56	55	44	57
24	.	4	7.5	80	64	56	55	54	54	55
25	.	4	7.7	80	64	55	55	55	54	55
26	.	4	7.6	83	55	56	55	54	55	55
27	.	4	8.9	84	56	56	55	55	55	54
28	.	4	7.4	81	66	56	55	55	54	77
29	.	4	10.1	85	55	66	55	55	54	55
30	F	3	7.9	82	55	56	55	55	55	55
31	.	4	7.2	75	66	55	55	55	44	55
32	.	3	7.0	80	55	66	55	54	55	55
33	.	4	7.3	80	55	55	55	55	54	55
34	.	4	8.3	83	56	66	55	54	54	55
35	.	4	8.5	83	54	66	55	55	44	57

N	sex	age	wt	D/V	Pgm	Odh	Pgi	Mpi	Aat	Pgd
36	.	5	9.1	85	54	56	66	54	54	55
37	.	4	9.1	82	55	66	55	55	55	55
38	.	6	9.2	82	56	66	55	44	55	55
39	.	3	7.2	79	56	66	55	54	55	57
40	.	4	8.5	83	55	56	55	55	55	55
41	.	4	10.1	83	55	46	55	54	54	55
42	.	4	10.5	85	54	55	55	44	54	55
43	.	4	8.6	82	56	56	55	54	54	55
44	.	4	9.5	82	55	56	55	55	54	55
45	.	4	8.2	77	54	55	55	54	44	55
46	.	4	8.2	83	55	56	55	54	44	55
47	.	5	7.0	83	55	56	55	55	54	55
48	.	4	9.1	84	64	56	55	54	54	55
49	.	5	8.1	87	56	46	56	55	54	55
50	.	5	7.4	79	54	66	55	54	54	55
51	.	4	8.8	78	54	56	55	44	44	55
52	.	4	8.9	83	56	55	55	55	44	57
53	.	4	7.3	80	55	66	55	55	54	55
54	.	5	16.0	101	55	56	55	55	55	55
55	.	4	9.0	82	56	56	55	55	55	55
56	.	4	8.7	82	55	55	56	44	54	55
57	.	4	7.7	83	56	56	55	54	55	55
58	.	4	8.7	81	54	55	55	55	54	57
59	.	4	7.2	80	43	56	55	53	54	55
60	.	4	10.6	86	66	56	55	54	44	57
61	.	4	10.4	81	55	55	55	55	54	55
62	.	5	7.3	79	66	55	55	54	44	57
63	.	4	7.6	85	56	56	55	54	41	77
64	.	4	9.2	82	54	55	55	55	55	55

G	sex	age	wt	D/V	Pgm	Odh	Pgi	Mpi	Aat	Pgd
1	F	6	12.4	102	55	55	55	54	55	55
2	F	7	12.6	94	55	66	55	54	54	55
3	M	5	15.8	102	54	55	55	54	54	57
4	F	5	11.3	101	55	56	55	44	55	55
5	.	6	19.6	110	55	56	56	54	55	55
6	.	4	12.6	95	54	66	55	55	54	55

G	sex	age	wt	D/V	Pgm	Odh	Pgi	Mpi	Aat	Pgd
7	M	4	14.3	108	55	56	55	55	44	55
8	.	5	11.9	97	56	66	55	44	44	55
9	M	4	11.4	97	56	56	55	54	55	55
10	M	3	12.9	101	56	66	55	54	54	55
11	.	4	12.0	96	56	56	55	53	54	55
12	.	5	10.6	92	55	56	56	54	55	57
13	M	4	10.5	101	55	56	55	55	42	55
14	M	6	14.2	104	56	56	55	55	55	57
15	F	5	14.1	99	54	66	55	54	54	55
16	M	4	12.7	99	55	56	55	55	54	57
17	.	5	11.2	100	55	66	54	55	54	55
18	.	7	11.2	96	44	56	55	55	54	55
19	F	3	14.1	104	56	56	55	54	54	55
20	F	3	15.4	101	55	55	55	55	51	55
21	M	4	13.0	103	56	55	55	55	64	55
22	M	5	14.2	100	55	56	55	55	54	57
23	.	4	12.7	100	55	66	55	54	54	55
24	M	4	12.7	97	55	56	56	55	54	55
25	.	4	13.7	108	55	66	55	55	54	55
26	M	5	18.2	108	56	56	55	54	55	55
27	M	5	13.7	106	55	56	56	55	54	55
28	F	4	14.1	101	54	56	56	55	54	55
29	.	3	13.3	95	56	55	55	55	55	57
30	F	4	14.7	99	55	55	55	54	55	55
31	.	3	10.3	91	55	56	55	54	51	55
32	M	4	11.6	91	55	55	55	54	54	57
33	.	3	10.6	90	55	56	55	54	54	55
34	M	5	8.8	88	56	66	55	55	54	55
35	.	4	12.9	98	54	56	55	54	55	55
36	F	3	10.7	89	54	55	55	55	54	55
37	.	4	13.4	99	56	56	55	55	44	55
38	F	4	10.2	94	56	55	55	44	55	55
39	M	4	12.2	99	56	55	55	54	54	57
40	.	.	13.7	101	56	56	55	55	54	55
41	.	5	14.3	97	55	56	55	54	55	57
42	F	4	17.5	105	54	56	55	55	54	55

S	sex	age	wt	D/V	Pgm	Odh	Pgi	Mpi	Aat	Pgd
1	M	9	23.6	119	.	.	.	.	.	.
2	.	9	17.3	127	.	.	.	.	.	.
3	F	9	19.2	120	.	.	.	.	.	.
4	M	7	18.9	123	.	.	.	.	.	.
5	M	6	8.7	85	.	.	.	.	.	.
6	F	5	9.7	88	.	.	.	.	.	.
7	F	8	16.1	110	55	56	55	55	55	55
8	F	9	18.0	111	64	55	55	55	54	57
9	F	10	29.0	128	54	55	55	44	55	77
10	M	10	26.6	121	55	55	55	55	54	57
11	F	6	10.3	90	55	55	55	55	55	57
12	F	5	6.0	75	55	56	55	54	44	77
13	F	8	25.5	120	54	56	55	54	44	55
14	F	8	22.9	124	56	56	55	54	44	55
15	M	7	24.2	122	56	56	55	55	44	55
16	F	9	23.0	122	44	56	55	55	55	57
17	F	8	12.4	102	66	56	55	55	55	55
18	M	6	9.0	88	55	55	55	55	54	57
19	M	7	31.3	127	44	55	55	55	54	55
20	F	10	22.4	130	54	55	55	55	42	57
21	M	9	18.6	118	54	55	55	55	54	55
22	F	9	22.6	127	54	55	55	55	55	57
23	F	5	6.0	75	54	56	55	55	54	55
24	M	5	5.0	70	66	56	55	55	55	55
25	F	8	19.3	124	55	55	55	54	54	55
26	F	8	20.0	113	55	56	55	55	55	55
27	F	11	13.6	127	55	55	55	55	54	57
28	M	10	26.4	122	55	56	55	54	54	55
29	F	10	13.2	119	56	66	55	55	55	55
30	M	6	12.9	98	44	55	55	55	55	55
31	F	9	20.1	124	54	66	55	44	55	57
32	F	9	6.3	121	55	56	55	54	55	55
33	M	10	18.4	125	56	55	55	54	44	55
34	F	10	16.1	127	66	55	55	55	55	55
35	M	10	29.6	130	54	56	55	55	55	57
36	M	9	25.1	136	55	55	55	54	54	57
37	M	10	35.3	132	54	55	55	55	55	55
38	F	11	14.2	132	56	56	55	55	44	55
39	F	11	18.2	129	56	55	55	44	44	55

S	sex	age	wt	D/V	Pgm	Odh	Pgi	Mpi	Aat	Pgd
40	M	10	20.3	135	56	56	55	54	55	57
41	M	10	32.7	140	56	66	55	55	54	55
42	F	12	11.0	123	55	56	55	55	54	57
43	M	4	6.1	73	54	56	55	55	54	55
44	M	6	10.0	110	55	55	55	55	54	55

Appendix B

Identification numbers of scallops by population, with the length mitotypes for locus I (LI), locus II (LII), and locus III (LIII), and the cleavage patterns for the restriction enzymes Eco RI, Sph I, Sal I, and Stu I. The restriction morphs are designated as in Chapter 4, with the addition of pattern Sph E to designate heteroplasmy for Sph A and Sph B. The mitotype designations for locus III are listed in Chapter 4. For locus I and locus II the following composite mitotype designations are used:

locus I

code	copy number
1	2
2	3
3	4
4	5
5	6
6	7
7	8
8	2/3/4
9	2/5
10	3/4
11	3/5
12	4/5
13	4/6
14	4/7

locus II

	code	alleles
1	B	
2	C	
3	D	
4	F	
5	B/D	
6	A/D	
7	A/E	

C	LI	LII	LIII	Eco	Sph	Sal	Stu
1	9	3	Q	A	A	A	A
2	3	.	.	A	.	.	.
3	5	.	.	A	A	A	A
4	2	.	.	A	.	.	.
5	3	.	.	A	.	.	.
6	3	.	A	A	A	A	.
7	4	3	H	A	A	A	A

C	LI	LII	LIII	Eco	Sph	Sal	Stu
8	3	4	P	A	A	A	A
9	3	3	E	A	B	A	A
10	3	3	I	A	C	A	A
11	5	3	M	A	C	A	A
12	3	3	D	A	A	A	A
13	3	3	T	A	A	A	A
14	4	3	C	A	C	A	A
15	3	3	D	A	A	A	A
16	3	3	I	A	A	A	A
17	3	3	B	A	.	A	A
18	3	3	R	A	C	A	A
19	4	3	E	A	C	A	A
20	3	5	G	A	C	A	A
21	2	3	C	A	C	A	A
22	3	.	P	A	A	A	.
23	3	.	B	A	A	A	.
24	3	3	E	A	C	A	A
25	4	3	B	A	D	A	A
26	4	3	B	A	C	A	A
27	1	7	E	A	A	A	A
28	4	3	I	A	C	A	A
29	4	3	G	A	C	A	A
30	4	3	B	A	A	A	A
31	4	3	G	A	A	A	A
32	14	5	E	A	E	A	A
33	6	3	A	A	C	A	A
34	11	3	X	A	A	A	A
35	4	3	A	A	C	A	A
36	3	3	P	A	C	A	A
37	3	.	E	A	A	A	.
38	3	1	B	A	A	A	A
39	3	3	B	A	C	A	A
40	3	3	B	A	A	A	A
41	3	3	N	A	A	A	A
42	3	2	G	A	C	A	A
43	3	3	B	A	A	A	A
44	3	3	B	A	A	A	A
45	3	4	C	A	A	A	A
46	2	3	G	A	C	A	A

H	LI	LII	LIII	Eco	Sph	Sai	Stu
2	3	.	B	B	A	A	A
3	3	3	B	A	A	A	A
4	3	1	B	A	A	A	A
6	3	3	H	A	A	A	A
7	3	.	C	A	A	A	A
8	4	3	I	A	A	A	A
9	10	3	M	A	E	A	A
10	3	3	G	A	C	A	A
11	10	3	B	A	A	A	A
14	4	3	B	A	A	A	A
15	3	3	G	A	A	A	A
16	3	3	G	A	A	A	A
17	2	3	B	A	A	A	A
18	3	.	.	A	B	.	.
19	3	.	.	A	A	.	.
20	3	3	P	A	A	A	A
21	3	3	G	A	A	A	A
22	5	3	A	A	A	A	A
23	2	1	E	A	A	A	A
24	4	3	G	A	A	A	A
25	3	3	H	A	A	A	A
26	11	3	P	A	A	A	A
27	3	3	B	A	A	A	A
28	3	3	S	A	A	A	A
29	2	3	G	A	A	A	A
30	2	3	E	A	A	A	A
31	4	3	B	A	A	A	A
32	2	3	C	A	A	A	A
33	2	3	B	A	A	A	A
34	3	3	E	A	A	A	A
35	2	3	G	A	A	A	A
36	3	3	B	A	A	A	A
37	1	3	E	A	A	A	A
38	3	1	B	A	A	A	A
39	3	3	G	A	B	A	A
40	3	3	D	A	A	A	A
41	3	3	G	A	A	A	A
42	3	3	B	A	A	A	A
43	3	3	G	A	A	A	A

I	LI	LII	LIII	Eco	Sph	Sal	S:
1	7	.	D	A	A	A	.
2	3	3	I	A	A	A	A
3	3	1	B	B	A	A	A
4	3	3	O	A	A	A	A
5	3	3	I	A	A	A	A
6	3	3	B	A	A	A	A
7	3	3	G	A	A	A	A
8	3	3	A	A	E	A	A
9	5	3	C	A	A	A	A
10	3	.	M	A	A	A	.
11	3	3	B	A	A	A	A
12	3	3	B	A	A	A	A
13	4	3	D	A	A	A	A
14	4	3	E	A	A	A	A
15	2	3	G	A	A	A	A
16	2	3	Q	A	A	A	A
17	2	3	A	A	A	A	B
18	3	3	G	A	A	A	A
19	4	3	G	A	A	A	A
20	3	3	I	A	A	A	A
21	5	.	.	A	.	.	.
22	2	1	I	A	A	A	A
23	4	3	G	A	A	A	A
24	12	3	G	A	A	A	A
25	2	3	D	A	A	A	A
26	2	3	H	A	A	A	A
28	3	.	D	A	A	A	A
29	4	6	A	A	A	A	A
30	3	3	G	A	A	A	A
31	3	3	A	A	A	A	A
32	10	3	N	A	A	A	A
33	10	.	F	A	A	A	.
34	3	3	E	A	A	A	A
35	3	3	D	A	A	A	A
36	3	3	P	A	A	A	A
37	3	3	H	A	A	A	A
38	4	1	N	A	A	A	A
39	2	3	B	A	A	A	A
40	4	3	B	A	A	A	A

I	LI	LII	LIII	Eco	Sph	Sal	Stu
41	4	5	E	A	A	A	A
42	2	1	B	A	C	A	A
43	3	3	A	A	A	A	A
44	3	1	I	A	A	A	A
45	3	3	Y	A	A	A	A
46	2	3	B	A	A	A	A
47	3	3	B	A	A	A	A
48	8	5	E	A	A	A	A
49	1	3	K	A	A	A	A
50	10	3	E	A	A	A	A
51	2	3	G	A	A	A	A
52	4	3	Q	A	A	A	A
53	3	3	I	A	A	A	A
54	3	3	M	A	A	A	A
55	3	3	H	A	A	A	A
56	3	3	G	A	A	A	A

N	LI	LII	LIII	Eco	Sph	Sal	Stu
1	3	3	A	A	B	A	A
2	3	3	H	A	B	A	A
3	3	3	B	A	A	A	A
4	5	3	B	A	A	A	A
5	3	3	A	A	A	A	A
6	2	3	M	A	A	A	A
7	3	3	A	A	A	A	A
8	3	3	G	A	A	A	A
9	3	3	B	A	A	A	A
10	2	3	A	A	B	A	A
11	3	3	B	A	B	A	A
12	4	3	C	A	A	A	A
23	3	3	E	A	B	A	A
24	3	3	B	A	A	A	A
25	3	3	O	A	B	A	A
26	3	3	A	A	A	A	A
27	4	3	K	A	A	A	A
28	6	3	.	A	A	B	A
29	3	3	C	A	B	A	A
30	3	3	B	A	A	A	A

N	LI	LII	LIII	Eco	Sph	Sal	Stu
31	3	3	B	A	A	A	A
32	3	3	H	A	A	A	A
33	7	3	I	A	B	A	A
34	3	3	B	A	A	A	A
35	4	3	H	C	A	A	A
36	4	3	B	A	A	A	A
37	4	3	C	A	A	A	A
38	4	3	M	A	B	A	A
39	3	3	B	A	A	C	A
40	3	3	Q	A	B	A	A
41	3	3	D	A	A	A	A
42	2	3	E	A	A	A	A
43	3	3	G	A	A	A	A
44	3	3	R	A	B	A	A
45	3	3	G	A	A	A	A
46	3	3	D	A	B	A	A
47	13	3	F	A	A	A	A
48	6	3	K	A	A	A	A
49	3	3	A	A	A	A	A
50	4	3	E	A	A	A	A
51	2	3	D	A	A	A	A
52	4	3	B	A	A	A	A
53	4	3	C	A	A	A	A
54	4	3	K	A	A	A	A
55	2	3	D	A	A	A	A
56	3	1	E	A	A	A	A
57	5	3	H	A	A	A	A
58	4	3	B	A	B	A	A
59	3	3	N	A	A	A	A
60	3	3	I	A	B	A	A
61	2	3	E	A	A	A	A
62	3	3	E	A	A	A	A
63	7	3	A	A	B	A	A
64	4	3	B	A	A	A	A

G	LI	LII	LIII	Eco	Sph	Sal	Stu
1	5	3	A	A	A	A	A
2	3	3	I	A	A	A	A

G	LI	LII	LIII	Eco	Sph	Sal	Stu
3	3	1	D	A	B	A	A
4	2	5	W	A	A	A	A
5	3	3	W	A	C	A	A
6	3	3	B	A	A	A	A
7	3	3	E	A	A	A	A
8	3	3	M	A	B	A	A
9	3	5	G	A	A	A	A
10	10	3	H	A	E	A	A
11	4	3	G	A	A	A	A
12	3	1	G	A	A	A	A
13	3	3	M	A	A	A	A
14	2	3	R	A	A	A	A
15	3	3	G	A	A	A	A
16	3	3	I	A	A	A	A
17	3	3	G	A	B	A	A
18	2	3	H	A	A	A	A
19	2	3	Q	A	A	A	A
20	3	3	A	A	A	A	A
21	2	3	H	A	B	A	A
22	2	3	B	A	A	A	A
23	3	3	C	A	A	A	A
24	2	3	B	A	A	A	A
25	4	3	B	A	A	A	A
26	10	3	E	A	A	A	A
27	2	1	G	A	A	A	A
28	4	3	I	A	A	A	A
29	3	3	A	A	A	A	A
30	2	3	G	A	C	A	A
31	3	3	H	A	A	A	A
32	3	3	Q	A	A	A	A
33	3	3	I	A	B	A	A
34	3	3	F	A	A	A	A
35	3	1	B	A	A	A	A
36	4	3	E	A	A	A	A
37	4	3	G	A	A	A	A
38	3	3	E	A	A	A	A
39	2	3	N	A	A	A	A
40	4	3	M	A	A	A	A

G	LI	LII	LIII	Eco	Sph	Sal	Stu
41	2	3	B	A	A	A	A
42	4	1	H	A	A	A	A

S	LI	LII	LIII	Eco	Sph	Sal	Stu
1	2	3	T	A	A	A	A
2	5	1	B	A	B	A	A
3	3	3	F	A	A	A	A
4	3	3	M	A	C	A	A
5	2	3	H	A	A	A	A
6	3	3	P	A	A	A	A
7	10	3	H	A	A	A	A
8	10	3	E	A	A	A	A
9	3	3	M	A	A	A	A
10	3	3	Z	A	A	A	A
11	2	3	K	A	A	A	A
12	2	3	N	A	C	A	A
13	3	3	I	A	A	A	A
14	2	3	H	A	A	A	A
15	2	3	G	A	A	A	A
16	3	3	N	A	A	A	A
17	2	3	L	A	A	A	A
18	5	3	B	A	A	A	A
19	2	3	V	A	B	A	A
20	2	3	K	A	A	A	A
21	3	3	E	A	A	A	A
22	3	3	Q	A	A	A	A
23	12	3	G	A	C	A	A
24	2	3	A	A	A	A	A
25	3	3	D	A	A	A	A
26	5	1	K	A	A	A	A
27	3	3	I	A	A	A	A
28	2	3	K	A	A	A	A
29	2	5	U	A	A	A	A
30	4	3	G	A	A	A	A
31	10	3	H	A	A	A	A
32	3	3	H	A	A	A	A
33	2	3	G	A	A	A	A
34	2	3	S	A	A	A	A

S	LI	LII	LIII	Eco	Sph	Sal	Stu
35	3	3	M	A	A	A	A
36	2	3	E	A	A	A	A
37	3	3	K	A	A	A	A
38	4	3	B	A	A	A	A
39	3	3	O	A	A	A	A
40	3	3	A	A	A	A	A
41	2	3	J	A	A	A	A
42	3	3	Q	A	A	A	A
43	4	3	H	A	A	A	A
44	6	3	O	A	A	A	A

References

Albertini, A. M., M. Hofer, M. P. Calos and J. H. Miller. 1982. On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. *Cell* 29: 319-328.

Anderson, S., A.T. Bankier, B.G. Barrell, M.H.L. de Bruijn, A.R. Coulson, J. Drouin, I.C. Eperon, D.P. Nierlich, B.A. Rae, F. Sanger, P.H. Schreier, A.J.H. Smith, R. Staden and I.G. Young. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290: 457-465.

Anderson, S., M.H.L. de Bruijn, A.R. Coulson, I.C. Eperon, F. Sanger, and I.G. Young. 1982. Complete sequence of bovine mitochondrial DNA: conserved features of the mammalian mitochondrial genome. *J. Mol. Biol.* 156: 683-717.

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

Avise, J. C., B. W. Bowen and T. Lamb. 1989. DNA fingerprints from hypervariable mitochondrial genotypes. *Mol. Biol. Evol.* 6: 258-269.

Avise, J.C., R.M. Ball and J. Arnold. 1988. Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. *Mol. Biol. Evol.* 5: 331-344.

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 mitochondrial DNA bridge between population genetics and systematics. *Ann. Rev. Ecol. Syst.* 18: 489-522.

Avise, J. C. and R. C. Vrijenhoek. 1987. Mode of inheritance and variation of mitochondrial DNA in hybridogenetic fishes of the genus *Poeciliopsis* . *Mol. Biol. Evol.* 4: 514-525.

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.

Avise, J.C. and R.A. Lansman. 1983. Polymorphism of mitochondrial DNA in populations of higher animals. pp.147-164. *In*: M. Nei and R.K. Koehn, eds. *Evolution of Genes and Proteins*. Sinauer Associates, Inc., Sunderland, Mass.

Ayala, F.J., J.R. Powell, M.L. Tracey, C.A. Mourão and S. Pérez-Salas. 1972. Enzyme variability in the *Drosophila willistoni* group. IV. Genetic variation in natural populations of *Drosophila willistoni* . *Genetics* 70: 113-139.

Beck, J. L. and B. C. Hyman. 1988. Role of sequence amplification in the generation of nematode mitochondrial DNA polymorphism. *Curr. Genet.* 14: 627-636.

Bentzen, P., W. C. Leggett and G. G. Brown. 1988. Length and restriction site heteroplasmy in the mitochondrial DNA of the American shad (*Alosa sapidissima*). *Genetics* 118: 509-518.

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

Bibb, M.J., R.A. Van Etten, C.T. Wright, M.W. Walberg and D.A. Clayton. 1981. Sequence and gene organization of mouse mitochondrial DNA. *Cell* 26: 167-180.

Birky, C. W. 1983. Relaxed cellular controls and organelle heredity. *Science* 222: 468-475.

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

Bogenhagen, D. F., S. S. Cairns and B. K. Yoza. 1985. Nucleotide sequences involved in the control of transcription and displacement loop DNA synthesis for *Xenopus laevis* mtDNA. pp.175-182. *In*: E. Quagliariello, ed. *Achievements and Perspectives of Mitochondrial Research*. Vol. II Biogenesis. Elsevier, Amsterdam.

Bogenhagen, D. and D. A. Clayton. 1974. The number of mitochondrial deoxyribonucleic acid genomes in mouse L and human HeLa cells. *J. Biol. Chem.* 249: 7991-7995.

Boursot, P., H. Yonekawa and F. Bonhomme. 1987. Heteroplasmy in mice with deletion of large coding region of mitochondrial DNA. *Mol. Biol. Evol.* 4: 46-55.

Boyce, T.M., M.E. Zwick and C.F. Aquadro. 1989. Mitochondrial DNA in the pine weevils: size, structure and heteroplasmy. *Genetics* 123: 825-836.

Brown, G.G. and L.J. DesRosiers. 1983. Rat mitochondrial DNA polymorphism: sequence analysis of a hypervariable site for insertions/deletions. *Nucleic Acids Res.* 11: 6699-6707.

Brown, W. M. 1985. The mitochondrial genome of animals. pp. 95-130. *In*: R.J. MacIntyre, ed. *Molecular Evolutionary Genetics*. Plenum, New York.

Brown, W.M. 1983. Evolution of animal mitochondrial DNA. pp. 62-88. *In*: M. Nei and R.K. Koehn, eds. *Evolution of Genes and Proteins*. Sinauer Associates, Inc., Sunderland, Mass.

Buroker, N.E., J.R. Brown, T.A. Gilbert, P.J. O'Hara, A.T. Beckenbach, W.K. Thomas and M.J. Smith. 1990. Length heteroplasmy of sturgeon mitochondrial DNA; an illegitimate elongation model. *Genetics* 124: 157-163.

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.L., M. Stoneking and A.C. Wilson. 1987. Mitochondrial DNA and human evolution. *Nature* 325: 31-36.

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

Cantatore, P., M. Roberti, G. Rainaldi, C. Saccone and M. N. Gadaleta. 1988. Clustering of tRNA genes in *Paracentrotus lividus* mitochondrial DNA. *Curr. Genet.* 13: 91-96.

Cantatore, P., M. N. Gadaleta, M. Roberti, C. Saccone and A. C. Wilson. 1987. Duplication and remoulding of tRNA genes during the evolutionary rearrangement of mitochondrial genomes. *Nature* 329: 583-585.

Chang, D.D., T.W. Wong, J.E. Hixson and D.A. Clayton. 1985. Regulatory sequences for mammalian mitochondrial transcription and replication. pp. 135-144. *In*: E. Quagliariello, ed. *Achievements and Perspectives of Mitochondrial Research.* Vol. II Biogenesis. Elsevier, Amsterdam.

Chapman, R.W., J. C. Stephens, R.A. Lansman and J.C. Avise. 1982. Models of mitochondrial DNA transmission genetics and evolution in higher eucaryotes. *Genet. Res., Camb.* 40: 41-57.

Clark, A. G. 1988. 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: 471-481.

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.

Cronin, T.M., B.J. Szabo, T.A. Ager, J.E. Hazel and J.P. Owens. 1981. Quaternary climates and sea levels of the U.S. Atlantic coastal plain. *Science* 211: 233-240.

Culliney, J.L. 1974. Larval development of the giant scallop *Placopecten magellanicus* (Gmelin). Biol. Bull. (Woods Hole, Mass.) 147: 321-332.

Dadswell, M.J. and D. Weihs. 1990. Size-related hydrodynamic characteristics of the giant scallop, *Placopecten magellanicus* (Bivalvia:Pectinidae). Can. J. Zool. 68: 778-785.

de Jong-Brink, M., H.H.Boer and J. Joosse. 1983. Mollusca. pp. 297-355. *In*: K.G. and R.G. Adiyodi, eds. *Reproductive Biology of Invertebrates.* Vol I. Oogenesis, oviposition, and oosorption. John Wiley and Sons, New York.

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.

deZamaroczy, M. and G. Bernardi. 1986. The GC clusters of the mitochondrial genome of yeast and their evolutionary origin. *Gene* 41: 1-22.

Dubin, D.T., D.L. Prince and R.M. Kotin. 1985. Post-transcriptional modification of animal mitochondrial RNA: a methylation-dependent step in ribosomal RNA processing. pp. 165-174. *In*: E. Quagliariello, ed. *Achievements and Perspectives of Mitochondrial Research.* Vol. II Biogenesis. Elsevier, Amsterdam.

Edwards, C. A. and D. O. F. Skibinski. 1987. Genetic variation of mitochondrial DNA in mussel (*Mytilus edulis* and *M. galloprovincialis*) populations from South West England and South Wales. *Mar. Biol.* 94: 547-556.

Efstratiadis, A., J.W. Posakony, T. Maniatis, R.M. Lawn, C. O'Connell, R.A. Spritz, J.K. DeRiel, B.G. Forget, S.M. Weissman, J.L. Slightom, A.E. Blechl, O. Smithies, F.E. Baralle, C.C. Shoulders and N.J. Proudfoot. 1980. The structure and evolution of the human ß-globin gene family. *Cell* 21: 653-668.

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

Feinberg, A.P. 1984. Addendum. Anal. Biochem. 137: 266-267.

Foltz, D. W. and E. Zouros. 1984. Enzyme heterozygosity in the scallop *Placopecten magellanicus* (Gmelin) in relation to age and size. *Mar. Biol. Lett.* 5: 255-263.

Gach, M. H. and T. E. Reimchen. 1989. Mitochondrial DNA patterns among endemic stickleback from the Queen Charlotte Islands: a preliminary survey. *Can. J. Zool.* 67: 1324-1328.

Galtsoff, P.S. 1964. The American oyster *Crassostrea virginica* (Gmelin). Fishery Bulletin of the Fish and Wildlife Service Vol. 64. U.S. Dept. of the Interior U.S. Govt. Printing Office, Washington.

Garey, J. R. and D. R. Wolstenholme. 1989. Platyhelminth mitochondrial DNA: evidence for early evolutionary origin of a $tRNA^{ser}AGN$ that contains a dihydrouridine arm replacement loop, and of serine-specifying AGA and AGG codons. *J. Mol. Evol. 28*: 374-387.

Gartner-Kepkay, K.E. and E. Zouros. 1985. Influence of environment and human selection on the genetic structure of some economically important marine animal species. Final Report: Department of Supplies and Services Contract (D.F.O., Halifax, Canada). File No. 08sc.FP101-3-0301.

Gjetvaj, B. 1989. Mitochondrial DNA size variability in several species of scallops (Mollusca: Bivalvia). M.Sc. thesis, Dalhousie University, Halifax, Canada.

Gjetvaj, B., D.I. Cook and E. Zouros. 1990. Repeated sequences and large scale size variation of mitochondrial DNA: a common feature among scallops. *Mol. Biol. Evol.* (in press).

Gray, M.W. 1989. Origin and evolution of mitochondrial DNA. *Ann. Rev. Cell Biol.* 5: 25-50.

Hale, L. R. and R. S. Singh. 1986. Extensive variation and heteroplasmy in size of mitochondrial DNA among geographic populations of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 83: 8813-8817.

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.

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

Harrison, R. G. 1989. Animal mitochondrial DNA as a genetic marker in population and evolutionary biology. *TREE* 4: 6-11.

Harrison, R. G., D. M. Rand and W. C. Wheeler. 1987. Mitochondrial DNA variation in field crickets across a narrow hybrid zone. *Mol. Biol. Evol.* 4: 144-158.

Harrison, R. G., D. M. Rand and W. C. Wheeler. 1985. Mitochondrial DNA size variation within individual crickets. *Science* 228: 1446-1447.

Hauswirth, W. W. and P. J. Laipis. 1985. Transmission genetics of mammalian mitochondria: a molecular model and experimental evidence. pp. 49-59. *In*: E. Quagliariello, ed. *Achievements and Perspectives of Mitochondrial Research.* Vol. II Biogenesis. Elsevier, Amsterdam.

Hauswirth, W. W., M. J. Van De Walle, P. J. Laipis and P. D. Olivo. 1984. Heterogeneous mitochondrial DNA D-loop sequences in bovine tissue. *Cell* 37: 1001-1007.

Hertlein, L.G. 1969. Family pectinidae (Rafinesque, 1815). pp. 348-373. *In*: R.C. Moore, ed. *Treatise on invertebrate paleontology*, Part N, Vol 1, Mollusca 6, Bivalvia. The Geological Society of America, Inc. and the University of Kansas.

Holt, I.J., A.E. Harding and J.A. Morgan-Hughes. 1989. Deletions of muscle mitochondrial DNA in mitochondrial myopathies: sequence analysis and possible mechanisms. *Nucleic Acids. Res.* 17: 4465-4469.

Holt, I.J., A.E. Harding and J.A. Morgan-Hughes. 1988a. Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. *Nature* 331: 717-719.

Holt, I.J., A.E. Harding and J.A. Morgan-Hughes. 1988b. Mitochondrial DNA polymorphism in mitochondrial myopathy. *Hum. Genet.* 79: 53-57.

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.

Hyman, B.C., J.L. Beck and K.C. Weiss. 1988. Sequence amplification and gene rearrangement in parasitic nematode mitochondrial DNA. *Genetics* 120: 707-712.

Jacobs, H. T., D. J. Elliott, V. B. Math and A. Farquharson. 1988. Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. *J. Mol. Biol.* 202: 185-217.

Joyce, P. B. M., D. F. Spencer and M. W. Gray. 1988. Multiple sequence rearrangements accompanying the duplication of a $tRNA^{pro}$ gene in wheat mitochondrial DNA. *Plant Mol. Biol.* 11: 833-843.

Kessler, L. G. and J. C. Avise. 1985. A comparative description of mitochondrial DNA differentiation in selected avian and other vertebrate genera. *Mol. Biol. Evol.* 2: 109-125.

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

Kimura, M. 1968. Evolutionary rate at the molecular level. *Nature* 217: 624-626.

Kotylak, Z., J. Lazowska, D. Hawthorne and P. Slonomski. 1985. Intron encoded proteins of mitochondria: key elements of gene expression and genomic evolution. pp. 1-20. *In*: E. Quagliariello, ed. *Achievements and Perspectives of Mitochondrial Research.* Vol. II Biogenesis. Elsevier, Amsterdam.

Lansman, R. A., J. C. Avise and M. D. Huettel. 1983. Critical experimental test of the possibility of "paternal leakage" of mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 80: 1969-1971.

LaRoche, J., M. Snyder, D.I. Cook, K. Fuller and E. Zouros. 1990. Molecular characterization of a repeat element causing large-scale size variation in the mitochondrial DNA of the sea scallop *Placopecten magellanicus*. *Mol. Biol. Evol.* 7: 45-64.

Latorre, A., A. Moya, and F.J. Ayala. 1986. Evolution of mitochondrial DNA in *Drosophila subobscura*. *Proc. Natl. Acad. Sci. USA* 83: 8649-8653.

Levinson, G. and G. A. Gutman. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol. Biol. Evol.* 4: 203-221.

Lewontin R.C. and J.L. Hubby . 1966. A molecular approach to the study of genic heterozygosity in natural populations II. Amount of variation and degree of heterozygosity in natural populations of *Drosophila pseudoobscura*.. *Genetics* 54: 595-609.

Longo, F. J. and E. Anderson. 1970 Cytological aspects of fertilization in the lamellibranch, *Mytilus edulis*. II. Development of the male pronucleus and the association of the maternally and paternally derived chromosomes. *J. Exp. Zool.* 172: 97-120.

MacDonald, B. A. and R. J. Thompson. 1985a. Influence of temperature and food availability on the ecological energetics of the giant scallop *Placopecten magellanicus* . I. Growth rates of shell and somatic tissue. *Mar. Ecol. Prog. Ser.* 25: 279-294.

MacDonald, B. A. and R. J. Thompson. 1985b. Influence of temperature and food availability on the ecological energetics of the giant scallop *Placopecten magellanicus* . II. Reproductive output and total production. *Mar. Ecol. Prog. Ser.* 25: 295-303.

Maniatis, T., E.F. Fritsch and J. Sambrook. 1982. *Molecular Cloning: a Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.

Marini, J.C., S.D. Levene, D.M. Crothers, and P.T. Englund. 1982. Bent helical structures in kinetoplast DNA. *Proc. Natl. Acad. Sci. USA* 79: 7664-7668.

McIntyre, A., T.C. Moore, B. Anderson, W. Balsam, A. Bé, C. Brunner, J. Cooley, T. Crowley, G. Denton, J. Gardner, K. Geitzenauer, J.D. Hays, W. Hutson, J. Imbrie, G. Irving, T. Kellogg, J. Kennett, N. Kipp, G.

Kukla, H. Kukla, J. Lozano, B. Luz, S. Mangion, R.K. Matthews, P. Mayewski, B. Molfino, D. Ninkovich, N. Opdyke, W. Prell, J. Robertson, W.F. Ruddiman, H. Sachs, T. Saito, N. Shackleton, H. Thierstein and P. Thompson. 1976. The surface of the ice-age earth. *Science* 191: 1131-1137.

Merrill, A. S., J. A. Posgay and F. E. Nichy. 1966. Annual marks on shell and ligament of sea scallop (*Placopecten magellanicus*). *Fishery Bulletin* 65: 299-311.

Miyata, T., H. Hayashida, R. Kikuno, M. Hasegawa, M. Kobayashi and K. Koike. 1982. Molecular clock of silent substitution: at least six-fold preponderance of silent changes in mitochondrial genes over those in nuclear genes. *J. Mol. Evol.* 19: 28-35.

Monnerot, M., J-C. Mounolou and M. Solignac. 1984. Intra-individual length heterogeneity of *Rana esculenta* mitochondrial DNA. *Biol. Cell* 52: 213-218.

Moritz, C. and W. M. Brown. 1987. Tandem duplications in animal mitochondrial DNAs: variation in incidence and gene content among lizards. *Proc. Natl. Acad. Sci. USA* 84: 7183-7187.

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.

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

Naidu, K. S. 1970. Reproduction and breeding cycle of the giant scallop *Placopecten magellanicus* (Gmelin) in Port au Port Bay, Newfoundland. *Can. J. Zool.* 48: 1003-1012.

Naidu, K.S. and C.A. South. 1970. Occurrence of and endozoic alga in the giant scallop Placopecten magellanicus (Gmelin). *Can. J. Zool.* 48: 183-185.

Nei, M. 1972. Genetic distance between populations. *Amer. Natur.* 106: 283-292.

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

Olivo, P.D., M.J. Van de Walle, P.J. Laipis and W.MW. Hauswirth. 1983. Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. *Nature* 306: 400-402.

Powell, J.R., A. Caccone, G.D. Amato and C. Yoon. 1986. Rates of nucleotide substitution in Drosophila mitochondrial DNA and nuclear DNA are similar. *Proc. Natl. Acad. Sci. USA* 83: 9090-9093.

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

Prest, V.K. and D.R. Grant. 1969. Retreat of the last ice sheet from the Maritime provinces - Gulf of St. Lawrence region. *Geological Survey of Canada Paper 69-33*. Queen's Printer, Ottawa.

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

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

Raven, C.P. 1966. *Morphogenesis: The Analysis of Molluscan Development.* 2nd ed. Pergamon Press, Oxford.

Rigby, P.W.J., M. Dieckmann, C. Rhodes and P. Berg. 1977. Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J. Mol. Biol. 113: 237-251.

Roe, B.A., D-P. Ma, R.K. Wilson and JF-H. Wong. 1985. The complete nucleotide sequence of the *Xenopus laevis* mitochondrial genome. *J. Biol. Chem.* 260: 9759-9774.

Runnegar, B. 1987. Rates and modes of evolution in the Mollusca. pp. 39-60. *In*: K.W.S. Campbell and M.F. Day, eds. *Rates of Evolution.* Allen and Unwin, London.

Satta, Y., N. Toyohara, C. Ohtaka, Y. Tatsuno, T.K. Watanabe, E.T. Matsuura, S.I. Chigusa and N. Takahata. 1988. Dubious maternal inheritance of mitochondrial DNA in *D. simulans* and evolution of *D. mauritiana*.. *Genet. Res. Camb.* 52: 1-6.

Schaal, B.A. and W.W. Anderson. 1974. An outline of techniques for starch gel electrophoresis of enzymes from the American oyster *Crassostrea virginica* Gmelin. *Geo. Mar. Sci. Cen., Tech. Rep. Ser.* 74: 1-18.

Schon, E.A., R. Rizzuto, C.T. Moraes, H. Nakase, M. Zeviani and S. DiMauro. 1989. A direct repeat is a hotspot for large-scale deletion of human mitochondrial DNA. *Science* 244: 346-349.

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

Selander, R.K., M.H. Smith, S.Y. Yang, W.E. Johnson and J.B. Gentry. 1971. Biochemical polymorphism and systematics in the genus *Peromyscus* . I. Variation in the old-field mouse (*Peromyscus polionotus*). *Stud. Genet.* 6: 49-90.

Shumway, S. E., R. Selvin and D. F. Schick. 1987. Food resources related to habitat in the scallop *Placopecten magellanicus* (Gmelin, 1791): a qualitative study. *J. Shellfish Res.* 6: 89-95.

Siebenaller, J.F. 1979. Octopine dehydrogenase (E.C. 1.5.1.a) polymorphism in *Mytilus edulis* . *Isozyme Bull.* 12: 67-68.

Skibinski, D. O. F. 1985. Mitochondrial DNA variation in *Mytilus edulis* L. and the Padstow mussel. *J. Exp. Mar. Biol. Ecol.* 92: 251-258.

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

Slatkin, M. 1985. Gene flow in natural populations. *Ann. Rev. Ecol. Syst.* 16: 393-430.

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. Natl. Acad. Sci. USA* 84: 7595-7599.

Solignac, M., J. Génermont, M. Monnerot and J-C. Mounolou. 1987. Drosophila mitochondrial genetics: evolution of heteroplasmy through germ line cell divisions. *Genetics* 117: 687-696.

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

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

Solignac, M., J. Génermont, 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., M. Monnerot and J-C. Mounolou. 1983. Mitochondrial DNA heteroplasmy in *Drosophila mauritiana* . *Proc. Natl. Acad. Sci. USA* 80: 6942-6946.

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

Sutcliffe, W. H. 1976. Coastal circulation and physical oceanography of the Scotian Shelf and the Gulf of Maine. *J. Fish. Res. Board Can.* 33: 98-115.

Thomas, M. and R.W. Davis. 1975. Studies on the cleavage of bacteriophage lambda DNA with EcoRI restriction endonuclease. *J. Mol. Biol.* 91: 315-328.

Thuring, R.W.J., J.P.M. Sanders and P. Borst. 1975. A freeze-squeeze method for recovering long DNA from agarose gels. *Anal. Biochem.* 66:213-220.

Van Dijk, H. 1987. A method for the estimation of gene flow parameters from a population sructure caused by restricted gene flow and genetic drift. *Theor. Appl. Genet.* 73: 724-736.

Vaughn, K.C., L.R. DeBonte and K.G. Wilson. 1980. Organelle alteration as a mechanism for maternal inheritance. *Science* 208: 196-198.

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

Wallace, D.C. 1989. Mitochondrial DNA mutations and neuromuscular disease. *TIG* 5: 9-13.

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

Walsh, J.B. 1987. Persistence of tandem arrays: implications for satellite and simple-sequence DNAs. *Genetics* 115: 553-567.

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.

Wilson, A. C., R. L. Cann, S. M. Carr, M. George, U. B. Gyllensten, K. M. Helm-Bychowski, R. G. Higuchi, S. R. Palumbi, E. M. Prager, R. D. Sage and M. Stoneking. 1985. Mitochondrial DNA and two perspectives on evolutionary genetics. *Biol. J. Linn. Soc.* 26: 375-400.

Wolstenholme, D.R., J.I. Macfarlane, R. Okimoto, D.O. Clary and J.A. Wahleithner. 1987. Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. *Proc. Natl. Acad. Sci. USA* 84: 1324-1328.

Worms, J.M. and L.A. Davidson. 1986. Some cases of hermaphroditism in the sea scallop *Placopecten magellanicus* (Gmelin) from the Southern Gulf of St. Lawrence, Canada. *Venus* 45: 116-126.

Wright, J.M. and G.H. Dixon. 1988. Induction by torsional stress of an altered DNA conformation 5' upstream of the gene for a high mobility group protein from trout and specific binding to flanking sequences by the gene product HMG-T. Biochemistry 27: 576-581.

Wright, S. 1978. *Evolution and the Genetics of Populations.* Vol. 4: Variability within and among Natural Populations. University of Chicago Press, Chicago.

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