CELLULAR, MOLECULAR AND GENETIC ASPECTS OF GENDER SPECIFIC MITOCHONDRIAL INHERITANCE IN THE MARINE MUSSELS *MYTILUS*

by

Liqin Cao

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

at

Dalhousie University
Halifax, Nova Scotia
September 2003

Your file Votre référence
ISBN: 0-612-83697-5
Our file Notre référence
ISBN: 0-612-83697-5

DALHOUSIE UNIVERSITY

DEPARTMENT OF BIOLOGY

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "Cellular, Molecular and Genetic Aspects of Gender Specific Mitochondrial Inheritance in the Marine Mussels *Mytilus*" by Liqin Cao in partial fulfillment for the degree of Doctor of Philosophy.

Dated: September 4, 2003

External Examiner:

Research Supervisor:

Examining Committee:

Departmental Representative:

DALHOUSIE UNIVERSITY

DATE: September 4, 2003

AUTHOR: Liqin Cao

TITLE: Cellular, Molecular and Genetic Aspects of Gender Specific Mitochondrial Inheritance in the Marine Mussels *Mytilus*

DEPARTMENT OR SCHOOL: Biology

DEGREE: Ph. D. CONVOCATION: October YEAR: 2003

Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions.

Signature of Author

Table of Contents

List of Tables vii

List of Figures ix

Abstract xii

Abbreviations xiii

Acknowledgments xv

Chapter 1. General Introduction

Mitochondrial origin and genome diversity 1

Genome reduction 3

Mitochondrial genome inheritance in animals 3

Maternal inheritance of mtDNA 3

Doubly uniparental inheritance of mtDNA 4

Properties of the DUI system 6

Relaxed selection on mtDNA genomes 6

Role reversals of mtDNA genomes - masculinization and feminization 7

Origin of DUI: multiple vs. single 8

Female-dependent sex ratio 9

Leakage of DUI 10

Recombination of F and M mtDNAs 10

DUI breakdown and mtDNA introgression 11

Objective and Experimental Plans 11

Chapter 2. Differential Segregation Patterns of Sperm Mitochondria in Embryos of the Blue Mussel (*Mytilus edulis*)

Introduction 14

Materials and methods 17

Mussel crosses 17

Fluorescence labeling 18

Visualization of labeled mitochondria in developing mussel Embryos 19

Sex determination 21
Statistical analyses 21
Results 23
The distribution of the two patterns in families 27
Correspondence between mitochondrial distribution pattern and gender 28
Non-involvement of the male parent 29
No effect of sperm mitochondria labelling on sex-ratio 30
Discussion 30
Chapter 3. Genetics of Mother-Dependent Sex Ratio in Marine Mussels (*Mytilus* spp.)
Introduction 50
Materials and methods 53
Results 55
Pedigreed crosses 55
The model 59
Fitting the model to sex ratios produced by wild-caught animals ... 61
Pedigreed families revisited 63
Discussion 64
Chapter 4. The Control Region of Maternally and Paternally Transmitted Mitochondrial Genomes in Marine Mussels (*Mytilus* spp.)
Introduction 76
Materials and methods 78
Sample collection and identification 78
DNA amplification, cloning and sequencing 78
Sequence analysis 79
Results 80
The three parts of the large unassigned region (LUR) 81
The LUR as the control region of the mussel mtDNA 84
Discussion 89
Chapter 5. Characterization of Maternally and Paternally Transmitted

Mitochondrial Genomes in Mussels of the Genus *Mytilus*

Introduction 104
Materials and methods 105
Sample collection and DNA extraction 105
PCR amplification, cloning and sequencing 105
Data analysis 106
Results 107
Secondary structure features 108
Base composition among taxa 109
Compositional asymmetry in mtDNA genes 110
Codon usage pattern 113
Search for distinct domains between genders 114
Discussion 115

Chapter 6. Conclusions and Future Directions 144
Appendix A 151
Appendix B 155
Appendix C 156
Appendix D 167
Appendix E 172
References 176

List of Tables

Table 2.1. Distribution of sperm mitochondria in 2-cell embryos of *M. edulis* 39

Table 2.2. Distribution of sperm mitochondria in 4-cell embryos of *M. edulis* 40

Table 2.3. Distribution of the aggregated and dispersed sperm mitochondria patterns from 2- and 4-cell embryos within families of *M. edulis* (A) and cell localization of sperm mitochondria in 2-cell embryos within pattern within family (B). 41

Table 2.4. Probability of homogeneity of families grouped according to common sire (A) and sex-bias of progeny (B) for two characters: frequency of sperm mitochondria pattern (aggregated versus dispersed) and number of mitochondria in the specific cells of 2-cell embryos (AB versus CD) 42

Table 3.1. Sex ratio of progeny in pedigreed pair matings 68

Table 3.2. The model 70

Table 3.3. Progeny numbers of females from wild populations 71

Table 4.1. Origin, type and the parts of eleven mtDNA sequences studied 95

Table 4.2. Kimura two-parameter nucleotide divergences (with standard error in parentheses) at four mtDNA regions.....96

Table 5.1. Primer sequences and combination of primers for PCR amplification 121

Table 5.2. Sizes of the genes sequenced in the present study..... 122

Table 5.3. Base composition comparison among taxa for genes 123

Table 5.4. Base composition of *M. edulis* F-type mtDNA genes..... 124

Table 5.5. Cumulative codon usage of *Mytilus* spp. F-type and M-type mtDNA from *COI* and *Cytb* genes..... 125

List of Figures

Figure 2.1. Sperm of *Mytilus edulis* 43

Figure 2.2. Zygote of *Mytilus edulis* 44

Figure 2.3. Dispersed pattern of sperm mitochondria at various cell stages in embryos of *M. edulis* 45

Figure 2.4. Aggregated pattern of sperm mitochondria at various cell stages in embryos of *M. edulis* 47

Figure 3.1. Crosses of pedigreed females (and five of their daughters) to wild males 72

Figure 3.2. Crosses of pedigreed females to pedigreed and wild males 73

Figure 4.1. Nucleotide divergence in steps of 30 bp between the F and M sequences of individual w143 after removal of gap positions. 97

Figure 4.2. Neighbor-joining trees of the twelve sequences of Appendix 4.I for each of the four segments (16S-rRNA and the three regions of LUR) 98

Figure 4.3. Graphical presentation of the corresponding domains of the human mtDNA control region and the LUR of the F and M genomes 99

Figure 4.4. Sequence matching of some well defined functional elements of the human mtDNA control region with corresponding putative elements of the LUR of the F and M genomes 100

Figure 4.5. tRNA-like structures in the HVS1 of human mtDNA control region and the VD1 of the F and M LUR 101

Figure 4.6. Stem-and-loop structure in the CD of the human control region and in the CD of the F and M LUR 102

Figure 4.7. The divergence at the CD against the divergence at the 16S-rRNA of human mtDNA and seven other primates 103

Figure 5.1. Gene map of sequenced segment of *Mytilus* spp. mtDNA 126

Figure 5.2. Secondary structure model of *M. edulis* F-type 16S mitochondrial rRNA, from Cannone et al., (Unpublished) 127

Figure 5.3. Potential secondary structure models of *M. edulis* F-type 16S mitochondrial rRNA from this study 129

Figure 5.4. Potential secondary structure model of the third domain of *M. edulis* F-type 12S mitochondrial rRNA 132

Figure 5.5. Secondary structures of 8 F-type and M-type tRNA genes of *M. edulis* 133

Figure 5.6. Simplified genome map of *M. edulis* F-type mtDNA 135

Figure 5.7. Plot of nucleotide frequency on the third position of four-fold degenerate codons vs. gene location away from replication origin of H strand 136

Figure 5.8. Correlation between the nucleotide frequency at the third position of four-fold degenerate codons and the duration of single-stranded state (DssH) for each gene in the *M. edulis* F genome 137

Figure 5.9. Plot of effective number of codons (Nc) against GC content of the third codon position. 138

Figure 5.10. Correlation between the nucleotide frequency at the third codon position and the duration of single-stranded state (DssH) for each gene in the *M. edulis* F genome 139

Figure 5.11. Nucleotide diversity among the sequences of the same gender or between the pairs of sequences from the same species 140

Figure 5.12. Diagram illustrates the proposed deamination and oxidation processes in mussel mtDNA..... 143

Abstract

In the blue mussel *Mytilus*, females carry predominantly maternal mitochondrial DNA (mtDNA) but males carry maternal mtDNA in their somatic tissues and paternal mtDNA in their gonads. This phenomenon has been termed doubly uniparental inheritance (DUI) of mtDNA. In this study, the behavior of sperm mitochondria in *M. edulis* embryos was monitored. In female embryos, sperm mitochondria are randomly dispersed among blastomeres. In male embryos, sperm mitochondria tend to aggregate and land in one blastomere at the 2- and 4-cell stages. It is postulated that this aggregate eventually ends up in the first germ cells, thus accounting for the presence of paternal mtDNA in the male gonad. To investigate the mechanism of sex-ratio determination in *Mytilus*, pedigreed crosses were made for several generations. The data confirm the previous observations that in mussel crosses there is a strong sex-ratio bias and that the bias is a characteristic of the female. Furthermore, the data suggest that this bias is controlled by the mother's nuclear rather than mitochondrial genotype. A female-dependent sex determination model is proposed and tested by the data from the crosses made with wild mothers. To understand mtDNA features of DUI species, sequences from *Mytilus* F(emale) and M(ale) lineages involving various regions (the large unassigned region (LUR), *16S*, *12S*, *tRNA*s, *COI* and *Cytb*) were compared. Secondary structure modeling of 16S rRNA, 12S rRNA and tRNAs was performed. The LUR is identified as the control region of mussel mtDNA and appears to contain gender-specific motifs. M genomes evolve faster than F genomes for all regions examined, but they resemble each other in terms of RNA secondary structures, base compositions and codon usage. Different start codons are found to be used for the *Cytb* gene by *M. trossulus* F and M molecules. The location of the light strand replication origin is speculated. The base composition in the third codon positions (P_3) and four-fold degenerate sites (P_{4FD}) was analyzed for each of 12 *M. edulis* F mtDNA genes. In both P_3 and P_{4FD} the nucleotide frequencies for each gene are correlated with the gene's duration at the single-stranded state during replication and transcription, suggesting that an asymmetrical composition bias exists and that the bias is the main source shaping the codon usage. The hydrolytic deamination and oxidation of mtDNA during replication and transcription are likely responsible for the composition asymmetry in mussels.

Abbreviations

bp	base pair
CD	conserved domain
CSB	conserved sequence block
DNA	deoxyribonucleic acid
D_{ssH}	duration of the paternal H strand genome remaining single-strand state
DUI	doubly uniparental inheritance
GC3s	GC content at the third position of synonymously degenerate codons
HSP	major heavy strand promoter
kb	kilobase
LSP	major light strand promoter
LUR	large unassigned region
MitoFM	MitoTracker Green FM
mtDNA	mitochondrial DNA
mtTF	mitochondrial transcription factor
Nc	effective number of codons
O_H	H strand replication origin
O_L	L strand replication origin
PCR	polymerase chain reaction
RNA	ribonucleic acid
rRNA	ribosomal RNA

SRB	sex ratio bias
SSTS	sex specific transmission sequences
TAS	termination associated sequence
TAS(alt)	alternative termination associated sequence
TEM	transmission electron microscope
tRNA	transfer RNA
UV	ultraviolet
VD1	variable domain 1
VD2	variable domain 2

Acknowledgements

I am very grateful to my supervisor, Dr. Ellen Kenchington, for this opportunity to study the fascinating DUI story and her guidance, support and encouragement throughout the course of my research.

I would like to express my appreciation to my internal supervisor Dr. Brian K. Hall and my committee members, Drs. Robert W. Lee and Bill Pohajdak, for their advice and support. I thank Dr. Robert W. Lee for allowing me to join his lab meetings, which was very helpful in broadening my knowledge of the molecular field. Special thanks are due to Drs. Eleftherios Zouros and George C. Rodakis for their valuable discussions and suggestions on my projects. I thank Drs. Murray N. Schnare and David F. Spencer for their advice and assistance with rRNA secondary structure modeling.

Many thanks to Mr. Barry MacDonald for his excellent job raising mussels and useful tips for handling mussel babies. Thanks also to Dr. Dan Jackson, Ms. Amy Thompson and Mr. Barry MacDonald for their assistance in producing and sexing the mussel families. Dr. Dan Jackson further assisted with microscopic procedures and imaging. My thanks also go to Ms. Caroylyn J. Bird for allowing me to use the microscope and associated equipment at the Institute for Marine Biosciences, NRC-IMB, and to Dr. Mark Johnston for helpful suggestions on statistics. I would also like to thank many friends in Halifax, who have made my life here much more delightful. I am grateful for financial support from a NSERC grant issued to Dr. Ellen Kenchington, a Dalhousie Graduate Scholarship and the Patrick Lett Fund and Fisheries and Oceans, Canada.

Finally, I would like to thank my husband, Hiroshi Saito, for his understanding, love and support. I appreciate my parents for their endless love, encouragement and confidence in me.

Chapter 1

General Introduction

Mitochondrial origin and genome diversity

Mitochondria are commonly considered to have originated from an α-Proteobacterium through endosymbiosis (Gray, 1992; Martion and Müller, 1998). Several lines of evidence derived from phylogenetic analysis of small ribosomal RNA genes and proteins as well as comparison of gene content and gene arrangement suggest that mitochondria have arisen only once in evolution (Gray et al., 1998, 1999; Lang et al., 1999). Despite this common origin, the mitochondrial genome has displayed substantial variation in terms of size, conformation and gene content. In protists, *Plasmodium falciparum* has a 6 kb linear genome encoding only five genes (Wilson and Williamson, 1997), whereas *Reclinomonas americana* has a 69 kb circular genome encoding 94 genes (Lang et al., 1999). In Fungi, mitochondrial genomes are either linear or circular, range in size from 19 to 100 kb and encode 24 to 43 genes (Lang et al., 1999; Forget et al., 2002). Land plant mitochondrial genomes exhibit the most striking size variability, ranging from 180 to 2400 kb, although the numbers of genes encoded by these genomes range from 50 to 70 (Lang et al., 1999; Kubo et al., 2000), similar to the smaller protist mitochondrial genome. Mitochondrial DNA (mtDNA) of plants also shows a complex organization, of which linear or circular

molecules, either smaller or larger than the genome size can be found (Backert et al., 1997; Oldenburg and Bendich, 2001).

In contrast to the above lineages, animal mtDNAs are relatively conserved. They are usually circular molecules ranging in size from 14 kb to 20 kb (Lewis et al., 1995; Keddie et al., 1998). With few exceptions, these mtDNAs contain the same gene set, namely 13 proteins, small and large rRNA species and a complete set of 22 tRNAs, which are sufficient to decode all mitochondrial protein codons. In addition, there is at least one main non-coding region which demonstrates elements that may control mtDNA replication and transcription. This area has been referred to as the "control region". One remarkable feature of animal mtDNAs is their extreme organization compactness. There are no intronic sequences and flanking untranslated regions. Genes are arranged apart from each other only by a few nucleotides and even overlap sometimes. Although greater genomes up to 42 kb have been reported, they are mainly due to length variation in the control region (Wolstenholme, 1992). Genome organization deviating from this universal picture of animal mtDNA has been revealed in a few animals. In the phylum Cnidaria, all scyphozoan, cubozoan and hydrozoan species have a single 16 kb or two 8 kb linear mtDNA molecules, while all anthozoan species have a circular molecule (Bridge et al., 1992). In the nematode *Globodera pallida*, and the primitive mesozoan *Dicyema misakiense*, circlular subgenomic mtDNAs have been identified (Watanabe et al., 1999; Armstrong et al., 2000).

Genome reduction

Mitochondrial genomes have experienced tremendous reduction both in size and gene content compared to their free-living relatives (Fraser et al., 1995; Nierman et al., 2001). This reduction is achieved mainly through extensive gene loss and gene transfer to the nucleus after endosymbiosis. Some genes are lost because they either become dispensable or are complemented by nuclear genes (Selosse et al., 2001). Others are transferred to the nucleus due to different evolutionary forces. Several hypotheses suggest that the probability of organelle gene transfer to the nucleus is greater than the reverse (Thorsness and Weber, 1996; Doolittle, 1998). Other hypotheses argue that higher mutation rates and/or the lack of recombination in the mitochondrial genome favor mtDNA gene transfer to the nucleus (Martin and Herrmann, 1998; Berg and Kurland, 2000). Selection for smaller organelle genome size may be another explanation for gene transfer to the nucleus as smaller organelles likely enjoy a replication advantage (Selosse et al., 2001). The competitive advantage of deleted mtDNA over larger wild-types has been demonstrated in yeast (Selosse et al., 2001) and human mtDNA (Diaz et al., 2002). This theory could account for the compactness commonly exhibited by organelle genomes, especially animal mtDNAs, in which introns are usually lacking, and intergenic spacers are relative short.

Mitochondrial genome inheritance in animals

Maternal inheritance of mtDNA

In almost all animals examined to date, mitochondrial DNA is transmitted to progeny only through the mother (Hayashi et al., 1978; Birky, 1995, 2001). The blocking of paternal mitochondrial genome inheritance can take place at various stages through different mechanisms. Most commonly, sperm mitochondria enter the egg but are actively destroyed early by the embryo before the 8-cell stage (Szollosi, 1965; Anderson, 1968; Hiraoka and Hirao, 1988; Shalgi et al., 1994; Sutovsky et al., 1996). In *Drosophila*, sperm mitochondria persist in the midgut of the embryo and are defecated by the larva soon after hatching (Pitnick and Karr, 1998). In other cases, as in the polychaete worm *Nereis limbata* (Lambert and Battaglia, 1993), the tunicate *Ascidia nigra* (Ursprung and Schabtach, 1965) and the chiton *Tonicella lineata* (Buckland-Nicks et al., 1988), sperm mitochondria do not enter the egg at fertilization.

Maternal inheritance is considered advantageous because it can contain the spread of deleterious mtDNA genomes through the population (Grun, 1976). Should a detrimental mutant mtDNA, which can also replicate faster, arise in an individual, due to its replication advantage over its competitors, it will increase in frequency in the germ cells, and hence in its progeny. If the transmission is biparental, it will spread through the entire population and decrease the population fitness. If the transmission is maternal, it will be contained only in its descendants, and the other maternal lineages with fitter mitochondria will be favored by selection.

Doubly uniparental inheritance of mtDNA

A novel mtDNA transmission mode differing from the maternal inheritance of mtDNA was first reported in the marine mussel *Mytilus* (Skibinski et al., 1994a,b; Zouros et al., 1994a,b). Female mussels are normally homoplasmic for one type of mtDNA, the F (from Female) -type, which they transmit to both daughters and sons through the egg. In contrast, males are normally heteroplasmic for two types of mtDNA; the F-type, which they inherit from their mother but do not transmit to offspring, and the M (from Male) -type, which they inherit from their father and transmit to their sons through the sperm (Skibinski et al., 1994a,b; Zouros et al., 1994a,b). In typical adult males, the gonad is dominated by the M-type, and the somatic tissues by the F-type (Stewart et al., 1995; Garrido-Ramos et al., 1998). This system of mtDNA transmission has been termed "doubly uniparental inheritance" (DUI) (Zouros et al., 1994a). The DUI system has subsequently been identified in the marine mussels *M. trossulus*, *M. galloprovincialis*, *M. californianus* and in the closely related *Geukensia demissa* (Mytilidae) (Zouros et al., 1994a,b; Geller, 1994; Hoeh et al., 1996a; Beagley et al., 1997), in the fresh water mussels *Pyganodon grandis*, *P. fragilis* and *Fusconaia flava* (Unionidae) (Lui et al., 1996), and in the clam *Tapes philippinarum* (Veneridae) (Passamonti and Scali, 2001).

Although DUI species harbor and transmit both maternal and paternal mitochondrial genomes, it does not violate the central piece of uniparental transmission through which new mutations are restricted to a single transmission lineage. The DUI system can be considered as two uniparental transmissions operating independently in a single species, with the F genome inherited down

the maternal line and the M genome down the paternal line. The underlying mechanisms of DUI have not been uncovered yet, however, some speculations can be made.

Sperm of *M. edulis* possess five mitochondria (Longo and Dornfeld, 1967), and they enter all the eggs, regardless of whether they would develop into female or male individuals (Sutherland et al., 1998). Hence, paternal mtDNA represents a very small portion of the mtDNA pool in zygotes at fertilization given that eggs contain thousands of mitochondria (Humphreys, 1962). There are two ways through which the paternal genome can overcome this minority and take dominance in the male gonad. Either there is a replication advantage of M mtDNA over F mtDNA at early embryogenesis, which increases the probability of M mtDNA entering the first primordial cells (Skibinski et al., 1994b; Zouros et al., 1994b), or there is a mechanism that directs the paternal mitochondria into the early germ cells (Saavedra et al., 1997).

Properties of the DUI system

Relaxed selection on mtDNA genomes

The separate transmission of F and M mitochondria gives rise to more than 20% sequence divergence between the two genomes in *Mytilus* (Fisher and Skibinski, 1990; Rawson and Hilbish, 1995; Stewart et al., 1995). Comparison of mitochondrial *16S*, *COIII/ND2* and *COIII* genes of F and M showed that the M genome evolves faster than the F (Skibinski, 1994b; Rawson and Hilbish, 1995; Stewart et al., 1996). These results have been explained as the consequence of

smaller effective population size of M and/or a reduced selective constraint on the M genome (Stewart et al., 1996). The latter hypothesis is based on the observations that fixed and segregating sites in the two genomes at synonymous and non-synonymous positions at the *COIII* gene do not follow the neutral expectation and that amino acid substitutions in the M molecule are more common at polypeptide positions that are less conservative in other species (Stewart et al., 1996). In DUI species, the M genome is required to function mainly in the male germ cells, while the F genome functions in soma and female germ cells. Less functional load for M may impose less selective constraint on M than F. By extension, the F genome will be under more relaxed selection than the mtDNA genome with standard maternal inheritance, in which the mtDNA genome performs in soma, female and male germ lines (Stewart et al., 1996). This hypothesis has been supported by the sequence analysis of the *COI* and *COIII* genes from several animal phyla, in which the mussel mtDNA, both M and F, was the most divergent and faster evolving than in other species (Hoeh et al., 1996b).

Role reversals of mtDNA genomes - masculinization and feminization

In surveys of natural populations it has been commonly found that a minority of males carry an F-like sequence rather than typical M at the examined part of the genome (Fisher and Skibinski, 1990; Rawson et al., 1996b; Hoeh et al., 1997; Saavedra et al., 1996; Quesada et al., 1999; Ladoukakis et al., 2002). Two possibilities can be offered for this observation: either these males did not receive a genome from their male parent or their paternal molecule had an F-

type sequence (at least for the examined region). Results from laboratory crosses revealed that some males failed to receive a paternal genome from their father, and the F genome from their mother took over and assumed the male function in their germ cells (Saavedra et al., 1997). This leads to the speculation that the paternally transmitted F-like sequence in natural populations may result from an occasional invasion of the maternal lineage into the male germ line and the subsequent transmission through the male lineage. This hypothesis is consistent with the results of a *COI* gene study in which the paternal transmitted F-like genome has diverged from the typical F by 2.4%, implying that it has been inherited through the male lineage for a period of time since it invaded the male lineage (Hoeh et al., 1997). This role reversal of mtDNA from F to M has been termed "masculinization of the F molecule", similarly, the reverse in which M mtDNA enters eggs and behaves as F mtDNA henceforth will be called "feminization of the M molecule". The latter has not been reported in natural populations.

Origin of DUI: multiple vs. single

Divergence of F and M has been suggested to predate the speciation of *Mytilus edulis, Mytilus trossulus* and *Mytilus galloprovincialis* because M mtDNAs grouped together rather than with their co-species F counterparts (Rawson and Hilbish, 1995; Stewart et al., 1995). However, when sequence comparisons were carried out for six bivalve species with DUI from three distant groups, the six male molecules were not clustered together. Instead, males and females within

each group formed an M-clade and an F-clade, respectively, and the two clades from the same group were more closely related to each other than to other groups (Hoeh et al., 1996a). This observation can be explained by multiple origins of DUI if no role switch between F and M was satisfied. Masculinization of the F molecule has been reported in mussels (Saavedra et al., 1997). If this newly masculinized F molecule happened to replace in evolutionary time all the other paternal lineages in a population, the molecular divergence between maternal and paternal lineages in that population would be wiped out and the date of split between F and M would be reset (Hoeh et al., 1996a). Phylogenetic analyses of collections of paternally and maternally transmitted genomes from different genera of the Mytildae and from species of the Mytilidae and Unionidae are, indeed, consistent with multiple re-settings of the divergence of gender - specific lineages (Hoeh et al., 1997), suggesting DUI arose a single time in evolution with a periodical masculinization mimicking a *de novo* origin event. DUI appears more stable in the freshwater mussel Unionidae, where masculinization seems absent (Curole and Kocher, 2002).

Female-dependent sex ratio

In natural populations, female and male mussels are close to equal in number (Sastry, 1979). However, sex ratios of individual families can vary extremely. Some mothers produced only daughters, others produced mostly sons and still others produced both sexes in intermediate proportions (Zouros et al., 1994b; Saavedra et al., 1997; Kenchington et al., 2002). This property of sex

ratio bias has been proven to be under the control of the female mussel and independent of the male that donates the sperm (Saavedra et al., 1997; Kenchington et al., 2002).

Leakage of DUI

Fidelity of DUI is not perfect. Leakages of paternal and maternal genomes out of their own routes have been reported (Garrido-Ramos et al., 1998; Zbawicka et al., 2003). Specifically, the M molecule was detected at low amounts in various somatic tissues of all male mussels as well as in some female mussels (Garrido-Ramos et al., 1998), meanwhile traces of the F molecule have been found in male gametes (Zbawicka et al., 2003). Moreover, a leaky genome appears functional outside its own line as the expression of the M molecule in soma has been observed (Dalziel and Stewart, 2002). Maculinization can be considered as an extreme case of DUI leakage, in which the F genome not only leaks into the male germ line but also takes over the M regime.

Recombination of F and M mtDNAs

Co-habitation of F and M genomes in the same individual, especially in the male gonad, may facilitate the recombination of F and M mtDNA in DUI species. Recent studies in *M. galloprovincialis* and *M. trossulus* prove that this is the case in *Mytilus*, and the hybrid sequences produced by recombination were only detected in male gonads, but not in soma (Ladoukakis and Zouros, 2001; Burzyński et al., 2003). These observations raise the possibility of the presence

of mosaic (partly F and partly M) genomes in mussel populations, and the transmission of mosaic genomes is more likely through the M lineage.

DUI breakdown and mtDNA introgression

In North American populations, DUI always breaks down when inter-specific hybridization between *M. trossulus* and other *Mytilus* species is made, hence the mtDNA introgression is blocked (Rawson et al., 1996b; Saavedra et al., 1996). In contrast, DUI is not disrupted in European mussel inter-specific hybrids. The mtDNA of *M. edulis* has been introduced into *M. galloprovincialis,* and moreover, replaced the *M. trossulus* mtDNA genome in some European populations (Rawson and Hilbish, 1998; Quesada et al., 1998; Quesada et al., 1999). A recent study by Riginos et al. (2002) showed that European *M. trossulus* is not a pure species, instead, extensive introgression of non-allozyme nuclear loci from *M. edulis* to *M. trossulus* has occurred. Taken together, it suggests that DUI requires high compatibility between the nuclear genome and mitochondria/or mtDNA.

Objective and Experimental Plans

Several peculiarities of DUI and answers to some of them have been revealed since this unique mitochondrial inheritance system was discovered in the last decade. However, the mechanisms of keeping maternal and paternal genomes in lines, the roles of nuclear and mtDNA sex ratio determination and

the molecular features of F and M genomes are still unkown. This study aims to employ the following approaches to address these questions.

1. Trace sperm mitochondria microscopically in male- and female-destined embryos of the mussel *M. edulis* at various cell stages to determine whether any different patterns hold between them. If different observations are identified, the experiment will be extended to embryos from crosses involving several male and female parents. The results are expected to answer the following questions.

a) Does paternal mitochondria have a replication advantage immediately after fertilization, or is there a directing force to lead paternal mitochondria to certain cell lines in male-destined mussel embryos?

b) Are paternal mitochondria actively eliminated in early developing female-destined embryos? These results may also provide clues to the phenomenon of DUI leakage.

c) Is DUI under the control of the mother or father or both?

2. Both pedigree and wild mussels of *M. edulis* will be used in this experiment. Pedigree families will be established for several generations. Parents will be chosen intentionally such that multiple females are mated to a single male and the reverse; sibling, half-sibling and aunt to nephew crosses will be made. The sex ratio of each family will be examined after the animals reach their sexual maturity. Influence of nuclear and mtDNA on sex ratio determination and the correlation between DUI and sex ratio bias will be evaluated, and a

model of sex ratio determination is expected to be proposed. The model will be further tested for its validity using sex ratio data from wild animal crosses.

3. A major portion of the *M. edulis* F-type mtDNA genome has been published, and an area having the potential for the control region has been located (Hoffmann et al., 1992). F- and M-type sequences of this region will be obtained from *M. edulis* and *M. galloprovincialis* in order to compare and determine whether this region is indeed the control region of the mussel mtDNA as well as whether there are gender-specific features in the region.

4. Various genes of F and M genomes (*12S*, *16S*, *tRNA* and protein-coding genes) from three DUI species *M. edulis*, *M. trossulus* and *M. californianus* will be sequenced. Genes are chosen so that all the gene types in mtDNA are included. Sequence comparisons of RNA secondary structures, base compositions, codon usage patterns and nucleotide diversities between F and M molecules will be carried out. Moreover, characteristics of the *M. edulis* F-type genome as a whole will be investigated using the combined sequence data from Hoffmann's and obtained in this study. These analyses will explore the DUI mtDNA features horizontally (between F and M mtDNA genomes) and vertically (between DUI and conventional animal mtDNA genomes). They will provide new insights into understanding the role of mtDNA genome in the unique DUI system.

Chapter 2

Differential Segregation Patterns of Sperm Mitochondria in Embryos of the Blue Mussel (*Mytilus edulis*)

Introduction

Strictly maternal inheritance of mitochondrial DNA (mtDNA) is the general rule in animals (Hayashi et al., 1978; Birky, 1995), even though cases of incidental ("leaky") inheritance of paternal mtDNA accounting for no more than 0.02% of an individual's mtDNA have been reported in fruit flies (*Drosophila*, Kondo et al., 1990), mice (Gyllensten et al., 1991) and anchovies (Magoulas and Zouros, 1993). The processes involved in the recognition and destruction of sperm mitochondria have been examined in only a few species. In mice, destruction occurs in the egg cytoplasm (Kaneda et al., 1995) and involves the recognition of a nuclear-controlled factor in the sperm mitochondrial outer membrane, which is produced in the male germ cells prior to spermatid formation (Shitara et al., 2000). In cattle, an ubiquitin marker is added to the surface of sperm mitochondria prior to sperm formation in the male reproductive tract, but is masked during passage though the epidymis. After the sperm is incorporated into the oocyte, ubiquitin is exposed and attacked by proteasomes and lysosomes of the ovum (Sutovsky et al., 1999, 2000). Ubiquitination of sperm mitochondria also occur in mice, monkeys and humans (Sutovsky et al., 1999, 2000, 2001).

From studies in mice and cattle, we know that the recognition/destruction mechanism of sperm mitochondria is species-specific. In both, paternal mitochondria are eliminated in intra-specific crosses soon after fertilization, but persist in inter-specific hybrids to later stages and may even be detected in adults (Kaneda et al., 1995; Shitara et al., 1998; Sutovsky et al., 1999, 2000).

It is known from early electron microscopy studies (Longo and Dornfeld, 1967) that the mid-piece of *Mytilus* sperm at maturity normally carries five mitochondria that are much larger than egg mitochondria. As in most other animals, the sperm mitochondria penetrate the cell membrane of the ovum at fertilization. However, paternal mtDNA is a minority in the zygote at fertilization, as the egg contains many tens of thousands of F-type mitochondria (Humphreys, 1962). Thus, at fertilization all embryos have a very large bias in favour of the F genome.

Clearly, in species with DUI, different mechanisms of mitochondrial replication and/or destruction must operate between the sexes to produce adult females where the M genome has largely disappeared and adult males where somatic cells are primarily or entirely F-type but the gonads M-type. In females, in a passive model where sperm mtDNA could have the same fate as egg mtDNA with regard to rate of replication and cell distribution in the developing embryo, the M genome would be subject to stochastic elimination or random presence in one or another tissue. This is consistent with observations made in some females using PCR techniques (Garrido-Ramos et al., 1998). Alternatively, active mechanisms may operate whereby M genomes are preferentially

degraded in females. In males, the abundance of paternal mtDNA in gonads must be achieved through active mechanisms given that the M genome is numerically biased against at fertilization. Those might simply involve differential replication and/or destruction of the two genomes, with the M-type being favoured in the gonad and the F-type being favoured elsewhere. Alternatively, there may be preferential partitioning of the paternal mitochondria to the gonads during early development followed by differential replication and/or destruction.

To determine which of these mechanisms are operative, it is necessary to link *in vivo* observations on sperm mitochondria to the sex of the individual. Mussels cannot be reliably sexed histologically, even under the best culture conditions, at less than a year of age. PCR assays using sex-specific mtDNA markers may lead to false characterization of a male animal as female through amplification error, since the female state is determined by absence of amplification product using male-specific markers. Occasionally the paternal mtDNA may have an F-like sequence (Saavedra et al., 1997) or the specific part of the molecule used for the assay might have incorporated an F-like sequence through recombination (Ladoukakis and Zouros, 2001). These difficulties may be largely overcome in *Mytilus* because it is possible to perform selected crosses using pedigreed females that produce progeny of one sex with very high frequency (Saavedra et al., 1997; Kenchington et al., 2002). The fate of sperm mitochondria can be followed in siblings, rather than individuals, facilitating sampling at different developmental stages.

Here I use a combination of genetic and cytological evidence made on selected crosses of pedigreed females, known to produce different proportions of male and female offspring, to obtain strong evidence for differential mitochondrial segregation patterns between the sexes. In female embryos the sperm mitochondria disperse randomly among the cells. However, in male embryos the sperm mitochondria aggregate in the large cell (CD cell) at the 2-cell stage and remain aggregated through subsequent cell divisions. These results are discussed in terms of the DUI model.

Materials and methods

Mussel crosses

Animals were collected from a mussel farm in Country Harbour, Nova Scotia, Canada. Because *M. trossulus*, a sibling species, is known to occur in the region, I confirmed species identity by restriction fragment analyses of the internal transcribed spacer (ITS) region of the nuclear ribosomal RNA genes (Heath et al., 1995). All male mussels and their male progeny carried the Med-1 haplotype of the paternally transmitted genome, which is the typical M-type in *M. edulis* from the Atlantic coast of Canada (Stewart et al., 1995).

The methods used for *in vitro* spawning and subsequent mixing of eggs with sperm were as described in Kenchington et al. (2002), as were the methods for rearing the offspring. Seven females were used, four from pedigreed lines known to produce almost exclusively daughters (> 97%), and three mostly sons (>75%). The seven were individually crossed with one of four males – the use of

multiple females with the same male allowing some examination of the roles of the two parental sexes in determining the fate of the sperm mitochondria. The parents used for each cross and the resulting offspring are here termed a "family" (Table 2.1).

Fluorescence labeling

MitoTracker Green FM (MitoFM) (Molecular Probes Inc., Eugene, OR, USA) is a mitochondria-specific vital dye. It is essentially nonfluorescent in an aqueous solution and becomes fluorescent when its chloromethyl moieties form covalent bonds with protein thiols in the mitochondrion (Haugland, 1996). The intensity of MitoFM fluorescence in living cells depends on the potential of the mitochondrial membrane (Cummins et al., 1997). Loss of mitochondrial membrane potential correlates with the destruction of the mitochondrion (Kaneda et al., 1995; Sutovsky et al., 2000).

MitoFM was diluted to a concentration of 1 mM in dimethyl sulfoxide (Sigma Chemical Co., St. Louis, MO, USA) and added to the sperm suspension to a final concentration of 200 nM. The sperm were incubated in the dye for 20 min at 18°C, then washed thoroughly on a Millipore filter and suspended in seawater. Labeled sperm were added to eggs at a very low concentration of approximately 100 sperm/ml (or 70 sperm per egg) for 10 min. The resulting zygotes were washed three times on a 20 μm filter to remove sperm that did not participate in fertilization but adhered onto the surface of the eggs.

To observe the relative position of the mitochondria with regard to the nucleus, sperm nuclei were counter-stained with propidium iodide (Molecular Probes Inc., Eugene, OR, USA), at a final concentration of 3 μg/ml for 10 minutes.

To examine the possibility that staining of the sperm mitochondria might have an effect on the distribution of paternal mtDNA among progeny, I conducted a subsidiary experiment in which eggs from two females were separately fertilised with stained and unstained sperm from the same male. One of the females (X102E) was known from a previous mating to produce only daughters, the other predominately sons (99wF1).

Visualization of labeled mitochondria in developing mussel embryos

An undetermined number of embryos from each family were extracted from the pool of offspring and mounted in sea water on glass slides under coverslips, sealed and placed in the dark at 18°C for 20 minutes. Epifluorescent microscopy was then used to determine the position of sperm mitochondria in individual embryos. At first I used a polyvar epifluorescent microscope (Reichert-Jung, Wien, Austria) with a 200-W mercury arc lamp and an ORCA 100/C4742-95 Hamamatsu digital camera (Hamamatsu Photonics KK, Japan) operated by SimplePCI software (Compix inc., PA, USA). A combination of a 450-495 nm bandpass exciter filter and a 520-560 nm bandpass barrier filter was employed. Individual embryos were examined by focusing down from one side of the embryo to the other, with several images being acquired at different focal planes.

I subsequently used a Nikon E800 epifluorescence microscope (Nikon, Japan) equipped with a 450-490 nm bandpass excitation and a 520 nm longpass emission filter block set. A series of optical sections were made by focusing down through the embryos at approximately 2-7μm intervals. Images were captured with a Nikon DXM 1200 high-resolution colour digital camera (Nikon, Japan).

Epifluorescent microscopy was used to capture data on embryos at the early stages of development. After the embryos reached the 4- or 8-cell stage, detection of the sperm mitochondria became more difficult with this method as the larvae acquired a background autofluorescence. For the later cell stages (trochophore and D stage), and for capturing images of the sperm alone, confocal microscopy was used. However, the increased time associated with collecting images with this method prohibited its use for extracting data for analytical purposes.

For confocal microscopy, MitoFM stained sperm and trochophore larvae were excited with 488 nm laser light, and the emission was collected with a 515-540 nm bandpass filter. Older D stage larvae were excited with 558 nm laser light and imaged through a 575-640 nm bandpass filter. This was done to distinguish the MitoFM signal from the autofluorescence background. The two images of the same larva were overlain such that orange coloration showed the autofluorescence and green coloration denoted stained sperm mitochondria within the embryo. Images of double-labelled sperm alone were obtained following the same procedure as with D stage larvae.

The number of fluorescent mitochondria and their distribution among the cells were recorded for individual embryos at the 2, 4 or 8-cell stages. Observations of later developmental stages, through to the trochophore stage, were also made from each family with varying degrees of success.

In many embryos the observed number of sperm mitochondria was less than five. This can be attributed to several factors, such as the actual number of mitochondria in the parental sperm, the position of the mitochondrion in the cell interfering with visualization, the density of staining and organelle death. All microscopic observations were made without prior knowledge of the family codes to avoid bias in the results.

Sex determination

Some progeny of all families were raised to sexual maturity and the sex ratio was determined for each family through microscopic examination of the gonads.

Statistical analyses

Chi-square tests were used to evaluate the non-randomness of 1) the distribution of sperm mitochondria among the cells of the embryo within families at the 2- and 4-cell stages according to the number of sperm mitochondria observed, 2) the occurrence of aggregated and dispersed sperm mitochondria distribution patterns among embryos within families irrespective of the number of sperm mitochondria observed, 3) the specific location of sperm mitochondria at

the 2-cell stage (small AB cell or large CD cell) within distribution pattern (aggregated, replicate = embryo; and dispersed, replicate = sperm mitochondrion) within families, 4) the occurrence of aggregated and dispersed sperm mitochondria distribution patterns among families sired by the same male, and among families producing only daughters and those producing mostly sons, and 5) the specific location of sperm mitochondria at the 2-cell stage (AB or CD cell) among families sired by the same male, and among families producing only daughters and those producing mostly sons.

For the first of these tests, the number of possible scenarios of the distribution of sperm mitochondria within families varied according to the number of mitochondria observed and the number of cells. Embryos with one observed mitochondrion were excluded from the data, leaving observations of 2, 3, 4 or 5 sperm mitochondria per embryo. In a 2-cell embryo in which three sperm mitochondria were observed there are two possible scenarios: 3:0 (all three mitochondria in one cell) and 2:1 (two mitochondria in one cell, one mitochondrion in the other cell). Symmetrical scenarios (e.g. 3:0 and 0:3) were pooled together. In most of these tests the expected numbers were small and so a Monte Carlo approach to estimating the probabilities was followed (cf. Manly, 1997). The set of the first 10,000 integers was divided into subsets of sizes proportional to the probabilities of the scenarios of the tested distribution. (In the "two cell, three mitochondria" case 2,500 randomly chosen numbers specified the set of the 3:0 scenario and the remaining 7,500 the set of the 2:1 scenario.) A random-number generating function was employed to draw from the entire set of

10,000 as many numbers as the embryos examined. The numbers drawn from each subset were compared to the expected proportions to generate a chi-square value. This process was repeated 1,000 times. The chi-square values generated in this way were compared to those from the observed numbers. The proportion of the 1,000 chi-square values that were equal or larger than the ones produced from the real observation specified the probability of the null hypothesis.

At the 2-cell stage developmentally-defined cells could be distinguished, and I tested whether one or other of these cells received more or fewer sperm mitochondria than expected by chance alone, by assuming that a mitochondrion had an equal chance to land in one or the other cell.

Because the aggregation of all mitochondria in one cell is the main feature that differentiates the aggregated from the dispersed pattern, I tested whether embryos with all mitochondria in one cell are more than expected by chance. For this I produced the composite "aggregated" class by summing all embryos where the 2, 3, 4 or 5 sperm mitochondria were observed in one cell. All other scenarios were lumped together to form the complementary "dispersed" class.

Results

Sperm of *M. edulis* stained with MitoFM showed a ring of five large mitochondria in the midpiece (Figure 2.1). This is consistent with previous TEM studies (Longo and Dornfeld, 1967). The enlarged size facilitated the tracing of mitochondria in the fertilised ovum. The sperm mitochondria of *M. edulis* are

approximately three times larger than mitochondria in spermatids (Longo and Dornfeld, 1967).

In all families, sperm penetrated the egg at various points in relation to the location of the first meiotic spindle, as observed by Longo and Anderson (1969). Following fertilization, the five sperm mitochondria separated from each other and disperse into the egg cytoplasm (Figure 2.2). At this time the zygote extruded the first and second polar bodies. Subsequently, the zygote formed a protrusion, referred to as the polar lobe, with decreased cytoplasmic density. The position of the polar lobe identified the vegetal pole of the embryo. Concomitantly, a vertical "cleavage furrow" developed, dividing the embryo into two blastomeres. The polar lobe connected to one of the blastomeres and flowed back into it when the first cleavage was completed. As a result, a 2-cell embryo with cells of unequal size was formed.

Following the system of nomenclature developed by Conklin (1897), the larger cell was designated CD, and the other AB. At the second cleavage, the CD blastomere formed a polar lobe again in the same fashion as in the first division. As a result, a 4-cell embryo was produced with three equally sized blastomeres A, B, and C, and one larger D blastomere. At the third cleavage, the D blastomere gave rise to the 1d-micromere and the 1D-macromere. At the fourth cleavage, the 1D-divided into the 2d-micromere and the 2D-macromere. Similarly, 3d-3D and 4d-4D were obtained at the fifth and the sixth cleavages as the embryo developed into a 64-cell stage. It is generally assumed that the germ cells and the gonad originated from 4d (Verdonk and van den Biggelaar, 1983).

The sperm mitochondria retained the vital stain through to the late D stage, indicating that the organelles were still alive. During this phase of larval development, replication or fusion of the sperm mitochondria were never observed. Rare observations of more than 5 mitochondria were made but these were present at fertilization, of equal large size and attributed to polyspermy, which is common in bivalves.

The number of embryos examined and the total number of sperm mitochondria observed in each family at the 2-cell and 4-cell stages are given in Tables 2.1 and 2.2, respectively. At the 2-cell stage the number of embryos in each family varied from 31 to 44 and the average number of sperm mitochondria per embryo varied from 2.97 to 4, with a mean of 3.27 (SD = 0.36). For the 4-cell stage the range was 3 to 3.92 (mean 3.47, SD = 0.46) and for the 8-cell, 3.08 to 3.9 (mean 3.41, SD = 0.30; not in the tables). Thus, the observed number of sperm mitochondria per embryo did not vary among families (or among cell-stages).

As early as polar lobe formation in the first division, sperm mitochondria appear to follow two distinct patterns of behavior, which I have named "dispersed" and "aggregated".

The dispersed pattern

In some embryos the sperm mitochondria appeared to disperse randomly in the egg before cell division (Figure 2.3A) and segregate also at random into different cells at the various developmental stages (Figure 2.3B, C, D). This pattern was observed as late as the early trochophore stage (Figure 2.3E), but

could not be identified at the D stage. This may be because dispersion makes the detection of the mitochondrial dye at late developmental stages more difficult.

The aggregated pattern

In other embryos, all five sperm mitochondria moved into the polar lobe while it was developing. It appeared that these mitochondria were pushed back later by the cytoplasm of the polar lobe as it fused with the CD blastomere, along either the "cleavage furrow" that divided the AB and the CD blastomere (Figure 2.4A) or the margin of the CD cell (Figure 2.4B). This process appeared to occur very quickly, and was observed repeatedly. Consequently, all of the sperm mitochondria were sequestered in the CD blastomere at the 2-cell stage (Figure 2.4C, D). The five sperm mitochondria were also found to aggregate in D and 1D cells at the 4-cell (Figure 2.4E) and the 8-cell (Figure 2.4F) stages, respectively. When all sperm mitochondria aggregated in one cell, they were usually located close to the cell membrane. As with the dispersed pattern, the aggregated pattern was observed to persist as late as the trochophore stage (Figure 2.4G) and even in larvae of the D stage (Figure 2.4H), almost 72 hours post-fertilization.

Because I could not follow the same embryo at successive developmental stages, I cannot assert that an embryo that showed the dispersed or the aggregate pattern in one stage also showed the same pattern at the next stage(s). Yet, two observations provide strong support that this must be the case. The first is the actual nature of the pattern: an aggregate pattern at the four-cell stage is more likely to have resulted from an aggregate pattern at the two-cell

stage than from a dispersed one. The second is the occurrence in high frequency of the same pattern at successive developmental stages in embryos of the same family: when an aggregate pattern was found in high frequency among two-cell embryos of a given family, it was also found in high frequency among the four-cell or eight-cell embryos of the same family.

The distribution of the two patterns in families

The distribution of sperm mitochondria was studied in 2- and 4-cell embryos from the seven families. Table 2.1 lists all possible distribution patterns in 2-cell embryos for each observed number of mitochondria (2 to 5), as well as the probability associated with each pattern, if the distribution is random. Families 1, 2, 3 and 4 produced only daughters and Families 5, 6 and 7 produced mostly sons (Table 2.1). Table 2.2 presents the equivalent information for 4-cell embryos from Families 2, 3 and 5. No observations beyond the 2-cell stage were made in Families 1 and 4, and for Families 6 and 7 the numbers of examined embryos with 4-cells were small and are not given in the Table. There were marked differences between the families. In families that produce no sons, the observed distribution of mitochondria amongst the cells rarely deviated significantly from expectation. There were three exceptions in the sixteen comparisons in Table 2.1 and two in eight in Table 2.2. The distribution of sperm mitochondria among the cells within these families is random. In contrast, the observed distribution of sperm mitochondria in the families producing mostly

sons deviated strongly from expectations in ten of twelve comparisons in Table 2.1 and all four in Table 2.2.

In 2-cell embryos I could consistently record if an observed sperm mitochondrion was in the AB or the CD cell. However, unambiguous cell identification was not always possible at the 4-cell stage. Chi-square tests of the frequency of occurrence of the aggregated and dispersed patterns within families were significant in two of the four daughter-producing families (Family 2 and 4), and for all families producing predominately sons (Table 2.3A). In the four daughter-producing families, an aggregated pattern was equally likely to be found in the AB (small) or CD (large) cell (Table 2.3B). Among embryos with an aggregated distribution of sperm mitochondria, from families producing mostly sons, the aggregation was more commonly found in the CD cell (Table 2.3B). For the dispersed pattern, in all families, there was no tendency for the sperm mitochondria to be delivered preferentially to the CD cell, and location of the sperm aggregation was random (Table 2.3B).

Correspondence between mitochondrial distribution pattern and gender

The joint consideration of the frequency of sperm mitochondrial distribution and sex-ratio in the seven families allows the following hypothesis. Two mutually exclusive processes guide the distribution of sperm mitochondria among the cells of the developing embryo. One represents the default situation, where the mitochondria are assorted at random. This stochastic process is associated with femaleness of the embryo. The other is a decisively non-random

process through which sperm mitochondria follow the same path though cell divisions. This directed process is associated with maleness of the embryo.

There is no strict one-to-one correspondence between the stochastic process and the dispersed pattern or between the directed process and the aggregate pattern. The fact that sperm mitochondria follow a random distribution in female-destined embryos implies that the aggregate pattern may occasionally be seen in families that produce exclusively daughters, because this pattern has a certain probability to occur by chance alone. This must be the case with the several aggregated types that have been seen in the four sonless families. Conversely, not all aggregated types in male-biased families need be associated with maleness. That a certain fraction of offspring in these families are females suggests that in a number of embryos, sperm mitochondria distribution is stochastic, and as a result a fraction of aggregated types could result from this process and be associated with femaleness.

Non-involvement of the male parent

Of the four males used in our families (Table 2.1) one (wm24) was used to produce Families 3 and 5 and another (wm26) to produce Families 4, 6 and 7. Families with the same male parent but different female parents were heterogeneous for two characters, the frequency of the aggregated pattern in 2-cell embryos and the distribution of mitochondria between AB and CD cells (Table 2.4). In contrast, the grouping of the seven families according to whether the female parent produces only daughters or mostly sons, regardless of the

male parent, results in two homogeneous classes for both characters of sperm mitochondria distribution. This test provides further evidence that the distribution of sperm mitochondria in the fertilised egg is under the control of the female parent.

No effect of sperm mitochondria labelling on sex-ratio

All of the progeny of a female known from a previous mating to produce only daughters (X102E) were female, and all of her progeny arising from unstained (N=28) and stained sperm (N= 79) were female. Progeny of a female known to produce mostly sons (99wF1) produced 65% male offspring when eggs were fertilized by unstained sperm (N=29) and 61% male offspring when eggs were fertilized with stained sperm (N=28). These results are consistent with the assumption that labelling the mitochondria with the fluorescent tag did not bias our results (χ^2= 0.141, P = 0.707).

Discussion

The technique I have used to follow the fate of sperm mitochondria in the fertilized ovum has several limitations. It cannot assure that all five mitochondria will be observed, and this will be more so as the embryo cells multiply. This is because mussel eggs are large (~ 65µm) and rich in yolk bodies and lipid droplets so that optical microscopy is unable to scan the whole depth of the embryo with sufficient resolution. Mitochondria fusion could also interfere with the ability to follow the number and fate of sperm mitochondria in the embryo.

Mitochondrial fusion is known to occur in other organisms. In the yeast *Saccharomyces cerevisiae* fusion and fission is continuous during the life of the cell (Jensen et al., 2000). In *Drosophila* mitochondria fuse in early postmeiotic spermatids to form the sperm's Nebenkern (Hales and Fuller, 1997). This process is associated with the expression of the gene *fzo*. In contrast the homologous gene in yeast, *fzo1*, is continuously expressed, in agreement with the fact that in yeast mitochondrial fusion is a continuous process (Hermann et al., 1998). In the diaphragm muscle of the rat the giant reticulum is the result of mitochondrial fusion, which begins after birth (Bakeeva et al., 1981). Thus both in *Drosophila* and rats mitochondrial fusion appears to be developmentally regulated.

When mitochondria fuse their matrix proteins mix rapidly (Nunnari et al., 1997). In principle this means that if I mark sperm and egg mitochondria with a different dye I may be able to detect fused mitochondria with both colours. These trials were unsuccessful because the density of egg mitochondria in the ovum is such that the whole egg is completely stained and detection of sperm mitochondria is impossible. Any evidence for fusion would have to come necessarily from the observation of the sperm mitochondria. I have not noticed any differences in sperm mitochondria size in the embryos I examined, as would be expected if fusion occurred among sperm mitochondria. Additional evidence against mitochondrial fusion comes from the detection of all five sperm mitochondria in larvae of late developmental stages (Figure 2.4). If mitochondrial fusion occurs regularly in early embryonic cells, the probability that all five sperm

mitochondria will accidentally escape fusion with any of the hundreds of thousands of egg mitochondria would be minute. Consistent with these observations are the findings of Ladoukakis and Zouros (2001) in the mussel *Mytilus galloprovincialis*, a sibling species of *M. edulis*. They observed recombinant sequences between maternal and paternal mitochondrial genomes in the gonads of several males, but failed to obtain such sequences from somatic cells from these males. This suggests that sperm mitochondria may fuse with egg mitochondria in the male gonad, where the paternal mtDNA proliferates and becomes dominant, but not before the formation of the first primordial cells. If this was not the case recombinant mtDNA genomes ought to be easily detectable in somatic tissues. Further, early embryonic divisions are not accompanied by mitochondrial division (Chase and Dawid, 1972; Rubenstein et al., 1977; Piko and Taylor, 1987), which is the reason for rapid tissue (Jenuth et al., 1996) or whole-body homoplasmies (Koehler et al., 1991) of individuals resulting from heteroplasmic eggs. Thus, to the extent mitochondrial fusions and fissions are associated with cell division, they would be less likely to occur in early embryos. Finally, even if one accepted that mitochondrial fusion is part of the reason why I was not able to follow the fate of all five mitochondria in all embryos, this would not explain the large and consistent differences between female-biased and male-biased families, unless one assumed that fusion itself is somehow associated with the embryo's gender.

My observations of the distribution of sperm mitochondria in the early developmental stages of mussel embryos allow several inferences to be made

about the fate of these mitochondria and, by extension, about the fate of paternal mtDNA. The first conclusion is that sperm mitochondria destruction in *M. edulis* does not occur during the early developmental stages. A mechanism of this kind involving ubiquitin is known to exist in mice and bovines, where it appears to operate almost immediately after the entrance of the sperm into the ovum (Kaneda et al., 1995; Sutovsky et al., 1996; Sutovsky et al., 1999). In all of our families I could see the persistence of sperm mitochondria to the stage at which the embryo may consist of 128 cells. After this, my data were insufficient to make claims on the fate of sperm mitochondria, however, at least in some embryos, they were observed to persist through to the D stage.

In order to explain the abundance of paternal mtDNA in the male gonad, early publications on the phenomenon of DUI (Skibinski et al., 1994b; Zouros et al., 1994b) speculated that the paternal mtDNA had an early replication advantage over the maternal mtDNA in male-destined embryos. This would increase the amount of paternal mtDNA relative to maternal mtDNA in the mtDNA pool of the early zygote and, thus, increase the chance that it would be included in the mtDNA pool of the first primordial cells. My observations render this hypothesis unlikely on two grounds. First, an early replication of sperm mitochondria in male- but not female-destined embryos would not be compatible with our observation that there is no variation among female- and male-biased families in the number of sperm mitochondria that can be seen at early stages. Second, my observation that sperm mitochondria form an aggregate suggests

another way through which these mitochondria may find their way to the male primordial cells, other than through mere chance boosted by early replication.

The exclusion of an early destruction mechanism of sperm mitochondria in female-destined embryos and of a replication advantage for paternal mtDNA in male-destined embryos leaves unexplained the difference in the fate of the paternal mtDNA in the two genders. My observations suggest a mechanism that would account for this difference and provide a description of the very first steps of that mechanism. The hypothesis is that in early-stage female embryos, sperm mitochondria behave like egg mitochondria. They are assorted at random among the early cells and start replicating at the same time and rate as maternal mitochondria, sometime after the D stage of larval development. In sharp contrast, sperm mitochondria follow a very regimented path in male embryos. They aggregate together just before the first cell division and are sequestered in the large CD cell. This aggregation into the same cell appears to continue for several successive cell divisions. I could not identify, in terms of developmental destiny, the chain of cells through which the aggregate passes, nor do I have data as to when and where the aggregate is dissolved. The fact that germ cells descend from the CD rather than the AB cell is compatible with the hypothesis that the aggregate follows a path that leads to primordial germ cells. However, my observations are not of a nature to provide direct confirmation of this hypothesis.

The existence of a mechanism through which sperm mitochondria are delivered to the first germ cells would explain why the paternal genome is found

almost exclusively in the male gonad, but can not explain why the gonad contains only paternal mtDNA. One way this could happen is if sperm mitochondria are the only mitochondria that enter the first germ cells. If, however, both egg and sperm mitochondria enter the germ cells, then it would require either that the former are actively eliminated from the germ line or that during gonad development the paternal mtDNA multiplies at a much faster rate than maternal mtDNA. There is indirect evidence in support of the latter hypothesis. One well established feature of DUI is that occasionally a maternally-transmitted mitochondrial genome may invade the paternal transmission route. The phenomenon, known as "masculinization" (Saavedra et al., 1997) of the mitochondrial genome, amounts to producing new paternal genomes out of the pool of maternal genomes. It has been observed in laboratory crosses of mussels (Zouros et al., 1994a,b; Saavedra et al., 1997) and in natural populations (Fisher and Skibinski, 1990; Rawson et al., 1996b; Quesada et al., 1999; Ladoukakis et al., 2002). The occasional replacement of paternal by maternal mtDNA suggests that maternal mitochondria can also enter the first germ cells.

Whereas the stochastic behaviour of sperm mitochondria in female-destined embryos requires no further explanation, the aggregation of these mitochondria and the delivery of this aggregate to specific cells in male-destined embryos require the existence of a specific mechanism. At present I have no idea how this mechanism might operate. My study is not the first to report a non-random distribution of mitochondria after cell division. In yeast (*Saccharomyces*

cerevisiae) mitochondria are transmitted from the mother cell to the bud daughter cell through a directed linear movement (Simon et al., 1997). They are first anchored to the bud tip from where they are released during the M phase. Consequently, all inherited mitochondria are retained in the daughter cell. This mitochondrial organization and movement in yeast is actin-associated. Some mutations affecting actin causes the mitochondria to re-enter the mother cell (Simon et al., 1997). Peroxisomes in yeast follow a similar dynamic movement during cell division. They are localized in the cortex of the mother cell by the actin skeleton and slide along the actin filaments to the bud cell during budding and spread over the whole bud cortex later (Hoepfner et al., 2001). This segregation fashion may be responsible for the faithful inheritance of these organelles. It is possible that the similar mechanism has been recruited in mussels to accomplish the transmission of paternal mitochondria within male *Mytilus* embryos.

The existence of a mechanism that operates in male mussel embryos, but not in female embryos, is the central piece of a model that was put forward to explain the two main observations of DUI in mussels: the presence of paternal mtDNA in the male but not female gonads, and the strong female-dependent sex bias. The model (Saavedra et al., 1997; Kenchington et al., 2002) predicts the presence of a factor that is expressed in females, perhaps only during oogenesis. It is encoded by a nuclear locus that segregates for two alleles, the active allele Z and the inactive allele z. Females of genotype zz produce eggs that lack the factor Z (see Chapter 3 for details). In these eggs the sperm mitochondria would have the default fate, which according to our observations is a random

distribution in the developing embryo. Thus zz females produce almost exclusively daughters. Females of genotype ZZ supply their eggs with factor Z. This factor initiates a battery of events the end result of which is the delivery of sperm mitochondria into germ cells. The ZZ females produce mostly sons. The fact that I have rarely seen females that produce exclusively sons suggest that there is always a small probability that eggs with the Z factor will fall back onto the stochastic mode of sperm mitochondria segregation. Females with the Zz genotype produce eggs whose supply with factor Z falls sometimes above and sometimes below the threshold required for the directed mode of sperm mitochondrial segregation. These are the females that produce intermediate sex ratios. The model implies that the development of a mussel egg into a female or male individual is determined at the moment it is produced by the mother.

This study takes this model one important step forward by indicating that the mechanism through which the mitochondrial factor is delivered to the germ cells operates through the aggregation of sperm mitochondria. Further evidence for the model comes from the presence of paternal mtDNA in somatic tissues. As stated, most females contain no paternal mtDNA, but when they do they are found in low amounts and in specific tissues that vary from individual to individual (Garrido-Ramos et al., 1998). This is expected from the stochastic distribution during development of a tiny minority of mitochondria, including their accidental loss. In males, paternal mtDNA is often found in the adductor muscle but rarely if at all in the mantle or the foot (Garrido-Ramos et al., 1998). The adductor muscle originates from the mesoderm, as do the germ cells, whereas the mantle and the

foot originate from the ectoderm (van den Biggelaar et al., 1994). This is what would be expected if a sperm mitochondrion breaks away from the aggregate and is trapped in an adjacent cell.

In conclusion, I have identified two distinct patterns of distribution of sperm mitochondria in developing embryos of the mussel *Mytilus edulis*. One pattern follows the rules of random distribution and the other is characterised by the co-occurrence of all or most sperm mitochondria in the same cell. In the 2-cell embryos this is the cell that initiates the line from which the germ cells descend. The first pattern is much more common in families that produce exclusively daughters and the other in families that produce mostly sons. From the fact that the paternal mtDNA is found almost exclusively in the male gonad I postulate that the end result of the directed process is the delivery of the sperm mitochondria into the first germ cells. I suggest that a factor or factors that are present in the egg control this chain of events and that its end result is the masculinization of the embryo. Females vary genetically in their ability to provide their eggs with this factor, with the results that some females are unable to produce sons, others produce mostly sons and still others produce a mix of both. These results are significant because they provide the first demonstration of a sex-specific behaviour of sperm mitochondria and because they open an important path toward the understanding of the cellular and molecular mechanisms that underlie the phenomenon of doubly uniparental inheritance of mtDNA.

Table 2.1. Distribution of sperm mitochondria in 2-cell embryos of *M. edulis*. For any number of sperm mitochondria (two to five) the possible distribution patterns and their probabilities to occur by chance (random distribution) are given. For each Family, the code for each parent (female, male), the number of daughters and sons (in parentheses), the observed number of embryos by total number of mitochondria observed by distribution pattern (expected number in parentheses), and the probability that the observed numbers are not different from the expected determined from Monte Carlo simulations are listed.

			Family 1		Family 2		Family 3		Family 4	
			X102B98a, wm2 (30, 0)		X102E, wm23 (198, 0)		X102H, wm24 (116, 0)		WF20, wm26 (30, 0)	
No. of mitochondria	Distribution pattern	P of distribution pattern	No. of embryos	P	No. of embryos	P	No. of embryos	P	No. of embryos	P
2	2:0	0.5	2(2.5)		5(4.5)		6(6.5)		4(4)	
	1:1	0.5	3(2.5)	0.377	4(4.5)	0.494	7(6.5)	0.601	4(4)	0.719
3	3:0	0.25	0(0.75)		6(3.75)		3(2.5)		4(2.75)	
	2:1	0.75	3(2.25)	0.156	9(11.25)	0.142	7(7.5)	0.491	7(8.25)	0.313
4	4:0	0.125	4(1.5)		2(1.75)		4(2)		5(1.38)	
	3:1	0.500	6(6)		7(7)		6(8)		2(5.5)	
	2:2	0.375	2(4.5)	0.032	5(5.25)	0.951	6(6)	0.277	4(4.13)	0.001
5	5:0	0.0625	2(0.8125)		2(0.375)		0(0.05)		1(0.25)	
	4:1	0.3125	6(4.0625)		2(1.875)		0(0.625)		1(1.25)	
	3:2	0.6250	5(8.125)	0.107	2(3.75)	0.020	2(1.25)	0.220	2(2.5)	0.240

			Family 5		Family 6		Family 7	
			98A, wm24 (30, 142)		WF21, wm26 (6, 15)		98B, wm26 (9, 39)	
No. of mitochondria	Distribution pattern	P of distribution pattern	No. of embryos	P	No. of embryos	P	No. of embryos	P
2	2:0	0.5	9(5.5)		10(6)		10(7)	
	1:1	0.5	2(5.5)	0.014	2(6)	0.008	4(7)	0.047
3	3:0	0.25	10(2.5)		9(2.25)		6(2.25)	
	2:1	0.75	0(7.5)	0.001	0(6.75)	0.001	3(6.75)	0.003
4	4:0	0.125	3(1)		1(0.875)		4(0.875)	
	3:1	0.500	2(4)		4(3.5)		0(3.5)	
	2:2	0.375	3(3)	0.052	2(2.625)	0.776	3(2.625)	0.001
5	5:0	0.0625	2(0.313)		2(0.188)		2(0.1875)	
	4:1	0.3125	2(1.563)		0(0.938)		1(0.9375)	
	3:2	0.6250	1(3.125)	0.006	1(1.875)	0.003	0(1.875)	0.001

Table 2.2. Distribution of sperm mitochondria in 4-cell embryos of *M. edulis*. See Table 2.1 for column explanations.

No. of mitochondria	Distribution pattern	P of distribution pattern	Family 2		Family 3		Family 5	
			No. of embryos	P	No. of embryos	P	No. of embryos	P
2	2:0	0.25	1(1)		2(1.25)		2(0.75)	
	1:1	0.75	3(3)	0.588	3(3.75)	0.318	1(2.25)	0.012
3	3:0	0.0625	1(0.25)		1(0.5)		3(0.3125)	
	2:1	0.5625	0(2.25)		4(4.5)		2(2.8125)	
	1:1:1	0.3750	3(1.5)	0.039	3(3)	0.744	0(1.875)	0.001
4	4:0	0.015625	1(0.109)		0(0.047)		2(0.109)	
	3:1	0.187500	2(1.3125)		1(0.5625)		2(1.3125)	
	2:2	0.140625	1(0.984)		1(0.422)		0(0.98)	
	2:1:1	0.562500	1(3.9374)		1(1.6875)		3(3.937)	
	1:1:1:1	0.093750	2(0.6562)	0.028	0(0.281)	0.601	0(0.656)	0.005
5	5:0	0.003906	0(0.0117)		0(0.0039)		2(0.035)	
	4:1	0.058594	0(0.1758)		0(0.0586)		3(0.527)	
	3:2	0.117188	0(0.3516)		0(0.117)		1(1.055)	
	3:1:1	0.234375	1(0.7031)		1(0.234)		2(3.164)	
	2:2:1	0.234375	1(1.055)		0(0.362)		0(2.109)	
	2:1:1:1	0.351563	1(0.7031)	0.879	0(0.234)	0.173	1(2.109)	0.001

Table 2.3. Distribution of the aggregated and dispersed sperm mitochondria patterns from 2- and 4-cell embryos within families of *M. edulis* (A) and cell localization of sperm mitochondria in 2-cell embryos within pattern within family (B). For each family, the observed number of embryos (expected number in parentheses) and the probability that the observed numbers are not different from the expected determined from χ^2 test are listed.

	Family						
	1	2	3	4	5	6	7
A. Distribution pattern							
Aggregated	8(5.562)	18(11.746)	16(12.851)	14(8.375)	33(10.520)	22(9.313)	22(10.313)
Dispersed	25(27.438)	44(50.254)	42(45.149)	20(25.625)	25(47.480)	9(21.687)	11(22.687)
Probability	0.257	0.043	0.319	0.025	0	0	0
B. Distribution of sperm mitochondria by cell type in 2-cell embryos							
No. of embryos with sperm mitochondria aggregated							
in AB cell	2	6	4	6	3	5	1
in CD cell	6	9	9	8	21	17	21
Probability	0.157	0.439	0.166	0.539	0.000	0.011	0.000
No. of mitochondria from dispersed pattern							
in AB	47	50	49	30	19	12	15
in CD	55	53	44	38	20	21	19
Probability	0.428	0.768	0.604	0.332	0.872	0.117	0.493

Table 2.4. Probability of homogeneity of families grouped according to common sire (A) and sex-bias of progeny (B) for two characters: frequency of sperm mitochondria pattern (aggregated versus dispersed) and number of mitochondria in the specific cells of 2-cell embryos (AB versus CD).

	Aggregated	Dispersed	χ^2(P)	AB	CD	χ^2(P)
A						
wm24						
F3	13	28	11.241	61	69	12.507
F5	24	10	(0.001)	27	82	(0.001)
wm26						
F4	14	20		48	65	
F6	22	9	7.111	26	68	13.715
F7	22	11	(0.029)	19	79	(0.001)
B						
Female-bias						
F1	8	25		56	76	
F2	15	29		69	80	
F3	13	28	2.23	61	69	0.918
F4	14	20	(0.526)	48	65	(0.821)
Male-bias						
F5	24	12		27	82	
F6	22	9	0.182	26	68	1.872
F7	22	11	(0.913)	19	79	(0.392)

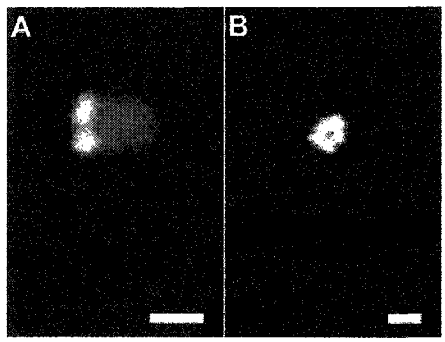


Figure 2.1. Sperm of *Mytilus edulis.* (A) MitoTracker Green FM labelled mitochondria (green) and propidium labelled nucleus (red), lateral view. (B) MitoTracker Green FM labelled mitochondria, showing 5 mitochondrial bundles, posterior view. Confocal microscope. Bars, 2µm.

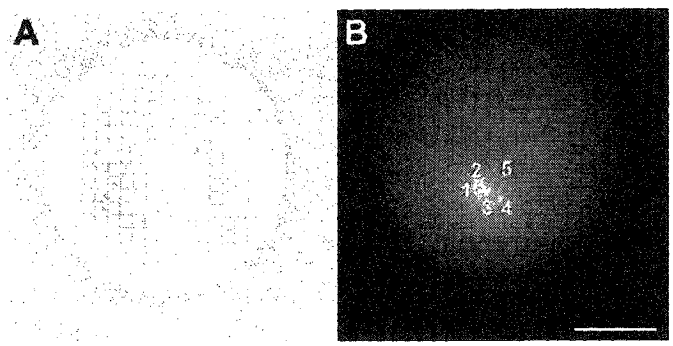


Figure 2.2. Zygote of *Mytilus edulis*. Bright field (A) followed by fluorescent field (B). Sperm mitochondria were labelled with MitoTracker Green FM. Numbers indicate individual sperm mitochondria. Epifluorescent microscope. Bar, 20μm.

Figure 2.3. Dispersed pattern of sperm mitochondria at various cell stages in embryos of *M. edulis*. Bright field followed by fluorescent field for each embryo. Fluorescent images were taken at different focal planes. Numbers indicate individual sperm mitochondria stained by MitoTracker Green FM. (A) One cell zygote with polar lobe (arrow). (B) 2-cell embryo. (C) 4-cell embryo. (D) 8-cell embryo. (E) Trochophore larva. (Epifluorescent microscope, A to D; Confocal microscope, E). Bars, 20μm.

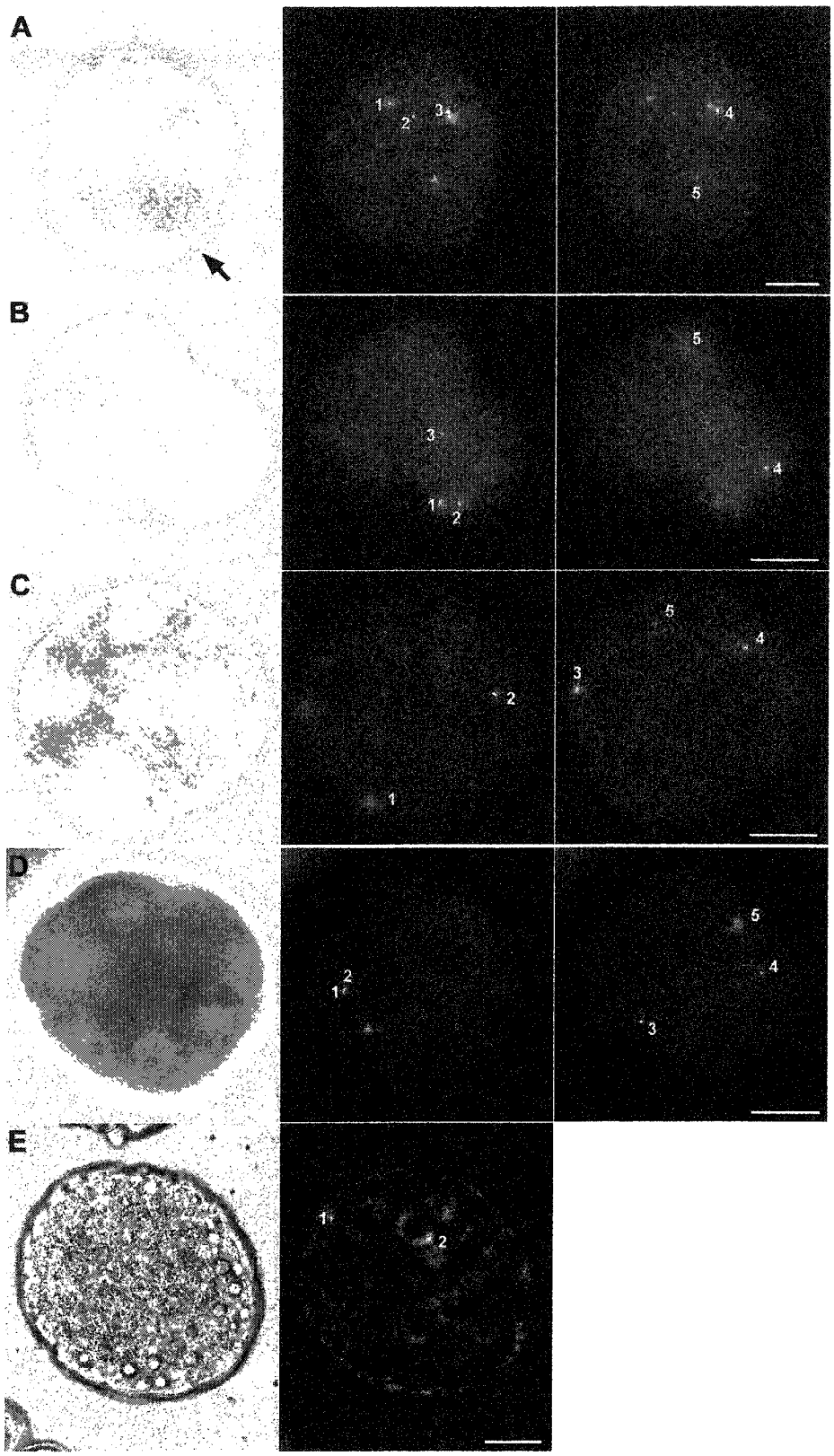


Figure 2.3

Figure 2.4. Aggregate pattern of sperm mitochondria at various cell stages in embryos of *M. edulis*. Bright field followed by fluorescent field for each embryo. Fluorescent images were taken at different focal planes. Numbers indicate individual sperm mitochondria stained by MitoTracker Green FM. (A) Late one-cell zygote with polar lobe (arrow). Sperm mitochondria along the cleavage furrow. (B) Late one-cell embryo with polar lobe (arrow). Sperm mitochondria along the line between polar lobe and the large cell. (C, D) 2-cell embryo. (E) 4-cell embryo. (F) 8-cell embryo. (G) Trochophore larva. (H) D stage larva. (Epifluorescent microscope, A to F; Confocal microscope, G, H). Bars, 20μm.

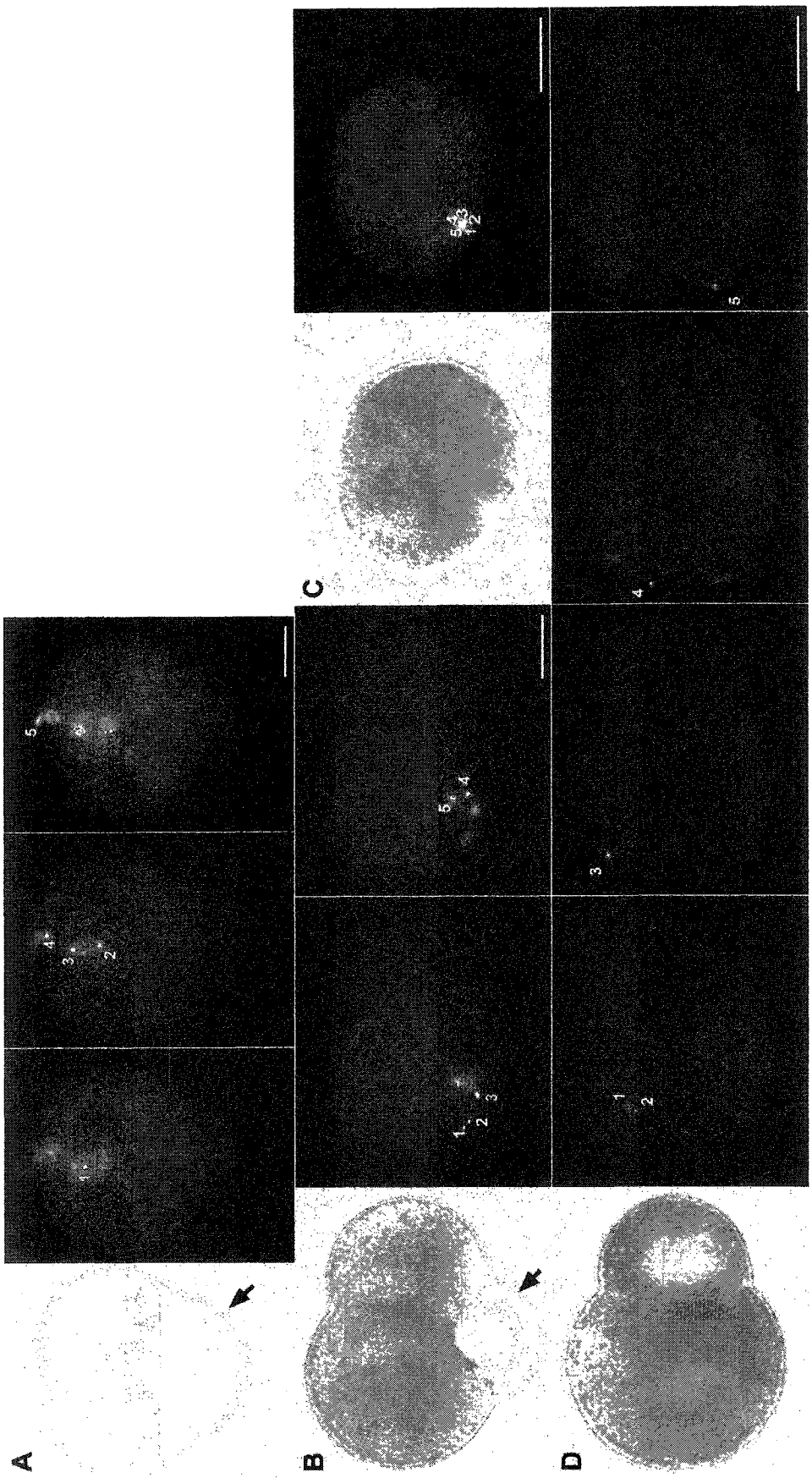


Figure 2.4

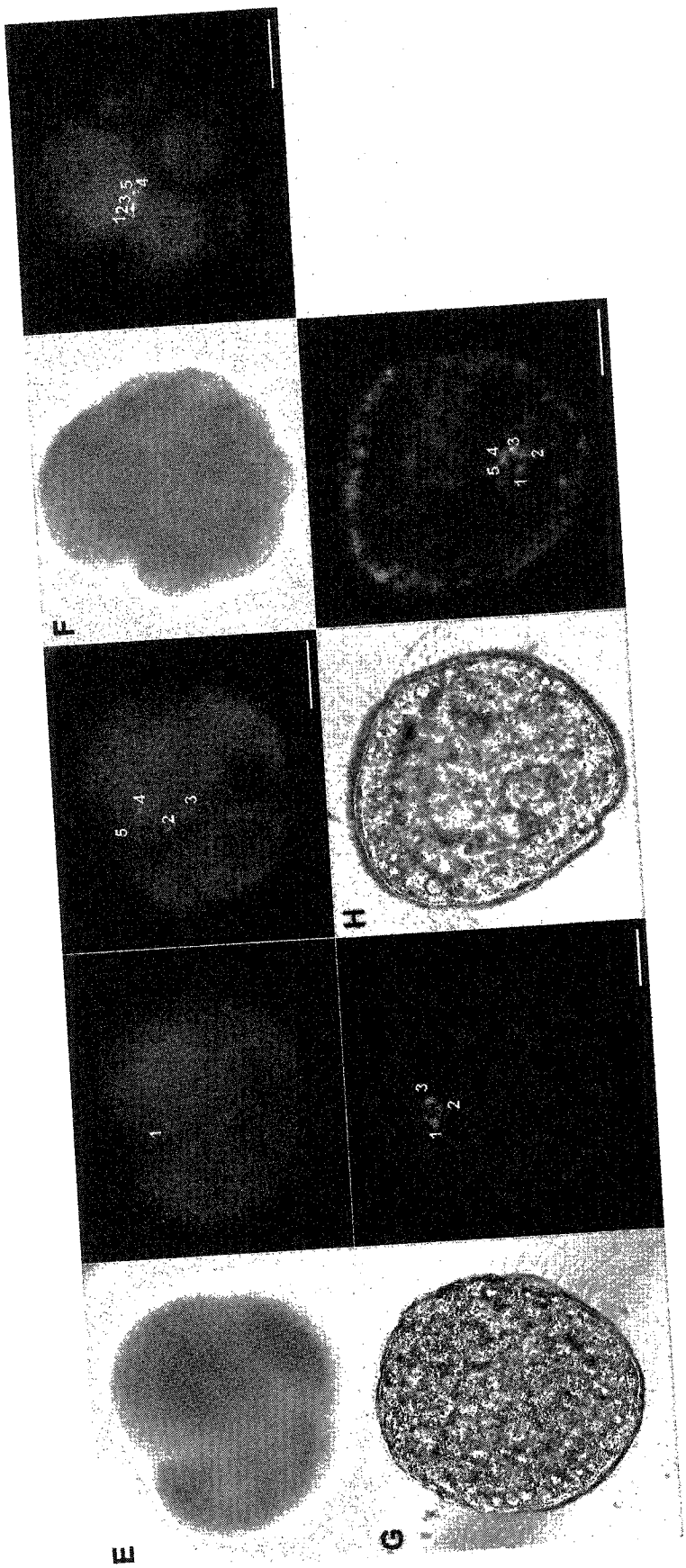


Figure 2.4. (Continued)

Chapter 3

Genetics of Mother-Dependent Sex Ratio in Marine Mussels (*Mytilus* spp.)

Introduction

Sex ratio bias (SRB) has been reported in a variety of invertebrate and vertebrate animals. In these species, the ratio of males and females among progeny deviates significantly from 1:1. Mechanisms responsible for this phenomenon fall into three categories: genetics, environmental effects and cytoplasmic microbes.

Species with multiple genetic sex determination factors produce a between-family sex ratio variance which cannot be accounted for by sampling error (Thompson and Bowen, 1972; White, 1973; Åkesson, 1977; Bull, 1983). Paternal and maternal contributions to family sex ratio and a sex ratio response to selection are characteristic of genetically-determined sex ratio bias.

There are many examples of environmentally-determined SRB. In the nematode *Echinomermella matsi*, sex ratio is density-dependent, which skews towards females at low adult abundance and towards males at high adult abundance (Stien et al., 1996). Other environmental cues inducing sex ratio bias include chemical factors (Leutert, 1974 cited in Pfannenstiel, 1984; Leutert, 1975), temperature (D'Cotta et al., 2001), nutrition (Peterson, 1972, 1977) and

reproductive time in the season (Wright et al., 1995). In some primate species, sex ratio is affected by the female's social rank (Hiraiwa-Hasegawa, 1993). Daughters tend to inherit their maternal rank and may achieve higher reproductive success. Therefore high ranking females produce more daughters, whereas low ranking females produce more sons.

Cytoplasmic transmission of microbes is another factor causing sex ratio bias in many invertebrate species (Bull, 1983; Rousset et al., 1992). The microbes may be protists or bacteria. As those cytoplasmic components transmit vertically from mother to daughters, not sons, they usually drive the sex ratio of offspring to extreme female bias by feminization in which the genetic males are reverted into functional females (Terry et al., 1997; Kageyama et al., 1998). Parthenogenesis that induces infected females to produce female offspring from unfertilized eggs through a modification of the first mitotic division (Stouthamer et al., 1994) and male-killing in which male progeny are killed during embryogenesis due to the bacteria in the cytoplasm (Jiggins et al., 1998; Hurst et al., 1999) also drive the SRB. In some cases, the symbionts can be destroyed by antibiotics or heat, and the normal sex ratio can be restored (Rigaud et al., 1991).

SRB has been revealed in marine mussels of the genus *Mytilus* by Zouros et al. (1994b). A pair-mating study of *M. galloprovincialis* further shows that the sex ratio of progeny is controlled by the mother, and independent of the father. The bias ranges from 0 to 97% male among families (Saavedra et al., 1997). The co-existence of SRB and DUI raises the possibility of causal linkage between the two phenomena. The evidence for this connection has yet to be established.

Research to date has not identified a sex chromosome in bivalve species, nor has the sex determination mechanism been determined. Simultaneous and sequential hermaphrodites prevailing in this group add further complications to interpret sex determination and sex ratio mechanism in bivalves. In the Pacific oyster *Crassostrea gigas*, Guo et al. (1998) observed an increasing proportion of females in older populations and strong paternal but not maternal effects on sex ratio. The authors proposed a single locus model of sex determination corresponding to two types of male, in which heterogametic individuals are permanent males and homogametic individuals are protandric females that may or may not reverse to females at a later age. Study of a European population of this species supported this model and indicated that the sex reversion took place during March to May (Lango-Reynoso et al., 1999).

In the dwarf surfclam *Mulinia lateralis*, Guo and Allen (1994) observed that diploids and triploids had a 1:1 sex ratio, whereas gynogenetic diploids were all females, suggesting an XX-female/XY-male sex determination with Y-domination. In the soft shell clam *Mya arenaria*, all triploids were females, which implied an X/autosome balance sex determination system (Allen et al., 1986).

In the mussel *Mytilus galloprovincialis*, all triploids were identified as males, whereas diploids had a 1:1 sex ratio. Kiyomoto et al. (1996) argued that this species may follow a ZZ male/ZW female sex determination model. However, female heterogamy cannot explain the unique feature of female-controlled SRB in this species (Saavedra et al., 1997).

In this study, I establish pedigreed families and analyze sex ratios in crosses through several generations in *Mytilus edulis*, in an attempt to determine the factors influencing sex ratio and reveal the correlation between DUI and SRB. This work was initiated in 1990 and was largely expanded since I joined the project in 1998. Specifically, I directed third and fourth generation crosses and interpreted the results with my co-authors. Our data show that sex ratio is the property of the female parent and is under the control of the mother's nuclear rather than mitochondrial genes. A single-locus system is proposed as the working model to operate sex ratio bias in *M. edulis.* The majority of these results have been published in Kenchington et al. (2002).

Materials and methods

Wild and pedigreed mussels of *M. edulis* were used in the study. Wild animals were randomly collected from natural populations in Nova Scotia, Canada, and transferred and maintained in the laboratory. All wild mussels were confirmed to be *M. edulis* using a PCR assay (Heath et al., 1995) as their sibling species, *M. trossulus*, which cannot be distinguished morphologically, occurs in the same region. Individual mussels were coded by the year of collection (with the last two digits), the letter W for wild, the letter M or F for male or female, and an identification number. After spawning a wild animal was either discarded or maintained for further use in other crosses. Pedigreed animals were produced in the laboratory, originally from wild stock, and subsequently used as parents. Female animals were coded alphabetically, while male animals were coded

numerically, both being referenced back to their original wild parents in their coding.

Wild animals were mated randomly to produce crosses. Pedigreed females were chosen intentionally from both 100% female and male-biased families, and this selection was continued for several descendent generations. The purpose was to trace the inheritance of the property of sex ratio bias through generations in mussels and detect, if any, the different inheritance pattern of this property between females coming from the two types of families representing the extremes of the SRB. A female could mate with several males and *vice versa*. Some pedigreed crosses were repeated in different years and in different seasons.

Methods of spawning, fertilization and rearing of progeny were generally followed as described in Zouros et al. (1992) and Saavedra et al. (1997). Briefly, animals were kept in individual containers to encourage natural spawning. Eggs and sperm from the desired female and male were transferred and mixed in a new container at an appropriate density. Offspring of each family were held in individual buckets. All apparatus used in maintaining crosses were soaked in a bleach solution before each application. This was done to kill any larvae that may have attached to containers or screens before reuse and eliminate the possibility of contamination among crosses. Postlarvae were maintained in buckets until they reached 2.5 mm in shell length and then transferred into individual silos in an upwelling unit. This upwelling system enhanced the growth of mussels. Animals could be sexed visually at about one year from birth. All sea water used

in the whole process was sand filtered, followed by 2μm and 10μm bag filtration and UV sterilization.

Sex of the progeny was determined by examining the presence of sperm or eggs in the gonads under the microscope, generally one year after fertilization. Published data of four pure *M .edulis* crosses from Zouros et al. (1994b) and four pure *M. galloprovincialis* from Saavedra et al. (1997) were also used in this study.

The homogeneity of sex-ratio among crosses was tested by a simple chi-square test. The population genetics model for the distribution of sex ratio in natural populations was explicitly solved and tested numerically with the EXCEL software program. The estimation of the model's parameter from empirical data was done by the maximum likelihood method.

Results

Pedigreed crosses

Results from 56 pedigreed crosses dammed by ten females and sired by 36 males are listed in Table 3.1. Out of ten female parents, six were daughters from a cross of a sonless female (90WF4) and four were daughters from a cross of a female producing mixed sexs (90WF5). The male parents were either wild animals or sons from two previous crosses. As can be readily seen, there is a considerable variation in the sex ratio among crosses ranging from 0% to 100% male. However, the overall sex ratio is close to 1:1. I define females that produce 5% or fewer male offspring as sonless, and females that produce more than 5%

males as son-bearing. It is notable that none of the three sonless females reverted to son-bearing when they were mated to a different male. The same is true for the seven son-bearing females. In contrast, all nine males that were mated to both sonless and son-bearing females produced both sonless and mixed broods. The sex ratio of their progeny was determined by the female they were crossed to, which implied a maternal but not paternal influence on sex ratio in *M. edulis.*

These results are consistent with the findings of Saavedra et al. (1997) for *M. galloprovincialis* that the sex ratio may vary from zero sons to over 90% sons and that this ratio is a characteristic property of the female parent. Homogeneity of sex ratios among crosses dammed by the same female was tested. It was not rejected in three sonless females (B, D and X), but was in four of the seven son-bearing females (E, Z, H and I). This implies that besides maternal effects other factors (*e.g.,* environmentally induced sex- specific mortality) may affect the final sex ratio in a mixed brood.

The novel observation from Table 3.1 is that the sonless female (90WF4) produced three sonless daughters (B, D and X) and three son-bearing daughters (A, E and Z), whereas the four daughters of the son-bearing female (90WF5) were all son-bearing. Therefore, there was a transition from a sonless mother to son-bearing daughter. To further investigate this phenomenon, two more generations of crosses using females descending from the original crosses in Table 3.1 were made. This extended the pedigreed crosses to four generations. For producing the third generation, 12 daughters of the sonless female X were

used as female parents to mate with the males collected from the wild population (Figure 3.1). Eight of them were also sonless and four were son-bearing. Four of the sonless daughters and one of the son-bearing daughters were crossed to more than one male. There was no reversion of sonless to son-bearing or *vice versa* across all of the cases.

The fourth generation was created using five daughters of three sonless mothers of the third generation (Figure 3.1). Three of the daughters were themselves sonless and two were son-bearing. The sonless trait has been passed down from mother to daughters for four generations. Yet, three transitions from a sonless mother to son-bearing daughters were observed (Table 3.1 and Figure 3.1): from 90WF4 to A, E and Z; from X to X102A, X102D, X102F and X102J; and from X102B to X102B2A and X102B2B.

Figure 3.2A presents four generations of crosses originating from a son-bearing mother (90WF5). Two daughters of the son-bearing female J, which came from a son-bearing family, produced mixed broods when they were mated to wild males. Two daughters from one of these two females (J17A) and three daughters from the other female (J17B) were also son-bearing when mated to their sibling brothers or wild males. Therefore, the son-bearing trait was also transmitted successively for four generations. Figure 3.2B presents the results involving two son-bearing daughters (E and A) of the original sonless female 90WF4 and one daughter from female E and A each. When E and A were mated to their nephews, i.e., sons from their sister Z, they were also son-bearing. Interestingly, a daughter of E produced all males, while a daughter of A also

produced a very high percentage of males. This is noteworthy as these two daughters shared the same father (98WM8), indicating that 98WM8 tended to produce daughters that produced a strong male-biased family.

Two son-bearing females (A and E) stayed as son-bearing when mated to different males at different years (Table 3.1). The same was true for the two sonless females (X102H and X102E) (Figure 3.1). This suggests that the property of SRB is independent of individual spawning and the age of female mussels. Three crosses (female X102E mated to 01WM4; female E to 01WM4; female A to Z101A) were made late in the season in September. There was no reversion from sonless to son-bearing or *vice versa* for the same female (Figure 3.1, 3.2). Therefore, SRB is demonstrated to be a stable trait controlled by the female, which appears unaffected by environmental factors at least under culture conditions. Furthermore, the sonless/son-bearing trait has shown to be transmitted maternally for several generations. Taken together it strongly implies that there is a hereditary basis to female-controlled sex ratio bias in mussels.

The fact that sonless females could produce both sonless and son-bearing daughters made it unlikely that any female's cytoplasmic factors, symbionts or mtDNA, could be the cue affecting mussel sex ratio. Like other animals, female mussels only receive maternal mtDNA. If sex ratio were controlled by cytoplasmic factors, all daughters of the same mother would produce the same sex ratio families, which was not the case in our study. Meanwhile, a single male could have both son-less and son-bearing daughters also excluded the possible influences of paternal cytoplasmic determinants on its daughter's sex ratio. Yet,

two lines of evidence suggest that males may have a genetic contribution to their daughters' sex ratio. A daughter of X102B fathered by 98WM1 was sonless, whereas two daughters fathered by another male (98WM2) were highly male-biased. Another case was that two daughters sharing the same father but having different mothers produced almost only sons. Taken all these observations together, solely female controlled sex ratio bias and involvement of both maternal and paternal nuclear genes in the transmission of this sex ratio bias trait is the possible model operating in mussels.

The model

A working model of genetic sex determination in mussels was proposed first by Saavedra et al. (1997) and revised by Zouros (2000). The authors suggested that femaleness is the default state and maleness requires the occupation of sperm mitochondria in the primordial germ cells. Sex determination operates through the interactions among three factors, a male-specific factor W that resides in the outer surface of sperm mitochondria, a female-specific factor X that resides in the egg cytoplasm and recognizes W and a third female-specific factor Z that is present in the egg cytoplasm and acts as a suppressor of factor X. Interaction between W and X results in femaleness, otherwise maleness when factor X is suppressed by Z. The authors further suggested that factor Z is controlled by a nuclear gene locus with two alleles, the active allele *Z* and the inactive allele *z*. ZZ females produce eggs with enough amount of Z factor in which the presence of sperm mitochondria in early germ cells are guarantied and

will develop into males. The zz females produce eggs with no Z factor in which sperm mitochondria cannot enter the primordial germ cells and will develop into females. Zz females produce eggs with intermediate quantities of Z factor and will develop into either males or females. The model has the assumption that all eggs receive sperm mitochondria at fertilization regardless of their future gender. In this thesis I have proved that it is indeed the case (Chapter 2).

Here I further elaborate this model. If it is assumed that the active allele *Z* and the inactive allele *z* segregate in the population, there will exist three types of females and males, ZZ, Zz and zz. While males have no influence on their offspring's gender, females of different genotypes hold the power to determine the sex of her progeny. I further assumed that ZZ females produce only sons, zz females produce only daughters and Zz females produce daughters with probability k and sons with probability 1-k. Frequencies of females and males with different genotypes at generation t were designated arbitrarily as in Table 3.2. Thus the frequencies of each genotype at generation t + 1 can be deduced as in six recursion equations (Table 3.2). The population reaches a stable equilibrium rapidly with the equilibrium frequencies shown in Table 3.2.

This system has several characteristics at equilibrium. First, males ($\hat{m} = \hat{d} + \hat{h}(1-k)$) and females ($\hat{f} = \hat{r} + \hat{h}k$) are in equal amounts in the population. Second, the male population follows Hardy-Weinberg proportions but the female population does not. Heterozygote frequency takes its maximum value 0.5 (k = 0) and its minimum value 0 (k = 1) in the male population, and 0.586 (k = 0.5) and 0.5 (k = 0 or 1) in the female population. Thus the heterozygous female

frequency always exceeds Hard-Weinberg equilibrium with the maximum excess of 0.125 when k approaches 0 or 1 and the minimum excess of 0.086 when k is 0.5. Third, in the whole population, the frequency of z takes its maximum value 0.625 at k = 0 and remains dominant until k approaches 0.311.

Fitting the model to sex ratios produced by wild-caught animals

Table 3.3 presents the results from 36 pair mating crosses mothered by wild females. In total, 27 crosses were made for the purpose of this study. Four crosses (7, 12, 18 and 28) were extracted from data presented in Zouros et al. (1994b) and five (8, 15, 19, 22 and 26) from data presented in Saavedra et al. (1997), to increase the sample size of the data set. All crosses were pure *M. edulis* crosses, except the five crosses from Saavedra et al. (1997), which were pure *M. galloprovincialis* crosses.

Sex ratio in 32 families of Table 3.3 deviated statistically from 1:1 ratio. However, the total numbers of females and males over all 36 crosses were 823 and 803, which was very close to 1:1. This fits as predicted by the model and is as observed in the wild populations (Sastry, 1979). Using these data to estimate the k value of the model, we need first to divide the 36 females into three classes: daughterless (genotype ZZ), mixed sex (genotype Zz) and sonless (genotype zz). If only females producing all male progeny are considered as daughterless and those producing all female progeny as sonless (zero as the cutoff value), we will get 1 daughterless female, 10 sonless females and 25 females of mixed sex. There are several ways to calculate k. The most

straightforward approach is to get the female frequency of progeny produced by all heterozygous females, which is 0.428 (582 out of 1361) in our case. This value would be a reliable estimator of k (k_1 = 0.428) if the sex ratios are homogenous among 25 heterozygous females, which is not true (Table 3.3). Therefore, a more reliable way to estimate k is to average the ratios over all heterozygous families. This gives the second estimated k_2 = 0.371. Finally, a third estimate of k can be obtained by ignoring the observed sex ratio within families and using instead the observed distribution of families into sonless, daughterless and mixed-sex classes. This can be done through the maximum-likelihood method, using the explicit expressions for the expected frequencies of each type of family in the population (Table 3.2). This produces k_3 = 0.211. On face value these estimates of k are different, but this cannot be supported statistically given that we do not have expressions for the variance of these estimates.

As it can be seen, a few mixed-sex families consisted of a majority of one sex and a low percentage of the other. We may consider it as an incidental leakage of the rare sex in these cases. We then may modify the criterion of classifying three types of females and repeat the above process. When 5% is used as a cutoff point for the rare sex, we get 3 daughterless, 22 of mixed sex and 11 sonless females with k_1 = 0.394, k_2 = 0.333 and k_3 = 0.291. When 15% is used as cutoff point, the numbers of the three classes of females are 5 daughterless, 17 of mixed sex and 14 sonless. The estimators of k are k_1 = 0.373, k_2 = 0.320 and k_3 = 0.305, respectively. There is not much difference

between the k_2 or k_3 estimates, whether one uses the 5% or 15% criterion, but there is a large difference for the values of k_1, which is another reason why this estimate is less reliable. It appears that 0.3 is a good estimate for k as we can obtain from the available data. With k = 0.3, the expected number of *ZZ* (daughterless) females is 3.7; of *Zz* (mixed sex) 20.4 and of *zz* (sonless) 11.9, which compare favorably with the observed 3, 22 and 11 when using the 5% cutoff point, or the 5, 17 and 14 when using the 15% cutoff point.

Having determined 0.3 as the k value, we may examine how many heterozygous females (in Table 3.3) (Zz) have produced a sex ratio in accordance with the model with this estimate. Using zero as the cutoff point for the heterozygous females, 12 (9 after the Bonferroni correction) out of 25 produced a significantly different sex ratio from that predicted from the model. This number was 9 (7 after the Bonferroni correction) out of 22 and 4 (3 after the Bonferroni correction) out of 17 when 5% and 15% cutoff point criteria were adopted, respectively. The variation of sex ratios produced by females of mixed-sex was considered only as a random sampling when estimating k and testing the model. However, it is most likely that other factors may affect the sex ratio as well, such as "leakage" in the sex-determining mechanism or family-specific mortality differences between sexes.

Pedigreed families revisited

Since the model of sex ratio determination has been established and the parameter of k estimated, we may explore it to speculate the genotypes of

mussel parents used in this study. Using 5% as cutoff point, the only three wild females from the original 1990 crosses can be predicted to be zz (90WF4) and Zz (90WF5 and 90WF7). According to the model, the proportion of the three genotypes of females in natural populations is ZZ:Zz:zz = 1:5.6:3.2. Thus, the observation of one *zz* and two *Zz* females in a sample of three is consistent with the model. Of 36 pedigreed female parents, 14 had genotype *zz* (all daughters of *zz* mothers) and 22 had genotype *Zz*.

As males have no effect on the sex of their progeny, genotypes of male parents can be deduced only from the brood information of their daughters, which is not available for most males. Of the three original wild males, 90WM4 can be deduced as genotype Zz since half of his daughters with his zz mate (90WF4) were sonless and another half were of mixed-sex. No genotypic inference can be made for the other two males, 90WM5 and 90WM7, either owing to the lack of the information about their daughters' broods (e.g., in 90WM7) or the information available cannot draw a solid inference (e.g., in 90WM5). Of the 24 other wild males used in subsequent generations, genotypes of only 5 can be inferred. Males 98WM1, 98WM4 and 98WM6 are of type *Zz* or *zz* and males 98WM2 and 98WM8 of type *ZZ* or *Zz*.

Discussion

The results obtained here demonstrate that in *M. edulis*, family sex ratio may skew extremely towards one sex or another and that this is a property of the female parent. This observation is in accordance with the findings made by

Saavedra et al. (1997) in the mussel *M. galloprovincialis* and suggests that it may apply to all species of the genus. The important new finding in this study is that this female-dependent sex ratio trait appears heritable. We have observed that in pedigreed crosses a sonless mother continued to produce sonless female descendants for four generations. The same is true for a son-bearing mother.

We first consider the possibility that sex ratio bias in mussel might be mediated by environmental factors incidentally, which is indeed the case in many invertebrate species. The observations that a sonless female never reverted to son-bearing and *vice versa* regardless of individual spawning year and the time in the season is incompatible with the sex ratio patterns affected by environmental factors, in which the same female can produce sex ratio ranging from one extreme to the other. Moreover, environmental effects also cannot explain the fact that some females are sonless, yet their same brood sisters are son-bearing (Figure 3.2).

Cytoplasmic factor involvement on sex ratio is also worth considering for two reasons. Cytoplasmic symbionts have been reported to drive extreme sex ratio bias in many invertebrate species. The unique mtDNA system of mussels, in which M and F mtDNA are passed down through different lineages separately, makes mtDNA itself a possible candidate as a sex determinant. As I have observed, the transmission of the sonless trait from mother to daughters is not perfectly faithful. There were three cases in which a sonless mother produced both sonless and son-bearing daughters in the same brood. This is not in agreement with the cases of sex ratio being affected by a cytoplasmic factor.

Cytoplasmic symbionts usually drive their host to produce biased sex ratio towards the gender carrying them. Specifically, they will be in favor of daughters if they are transmitted through eggs, and sons if through sperm. The fact that the same male may produce either mostly sons or all daughters depending on the females he mated excluded the possibility of sperm-transmitted symbionts. Egg-transmitted symbionts were also excluded on the basis that the same female may produce both sonless and son-bearing daughters. By the nature of cytoplasmic symbionts, daughters of infected females will receive the symbionts from their mother and continue to produce daughters. Taken together cytoplasmic factors are unlikely the cause responsible for the sex ratio bias in mussels.

Mother's nuclear genotype-dependent sex determination appears the most likely mechanism to explain sex ratio bias in mussels. Contribution of the father's nuclear genotype to sex ratio is excluded by the fact that the same female produced similar sex ratio broods regardless of the males to which she was crossed. The simple one-locus two-allele model proposed that the sex of the offspring is determined by its mother's genotype on an autosomal locus but not by the genotype of the offspring itself. Females homozygous for one allele produce all sons, for another allele produce all daughters. A heterozygous female produces mixed sex in a ratio depending on the degree of dominance of the two alleles. Two alleles have different distributions in the two sexes and different frequencies in the population as a whole when the system reaches a stable equilibrium, at which male population is at Hardy-Weinberg equilibrium but the

female population is not, and yet, the frequencies of female and male in the population are equal at 0.5. This model fits reasonably well to the data from 36 families with wild female parents. The model predicts that males may be involved in the sex determination of grandchildren through contributing alleles to daughters, although the male has no direct influence on the sex of his offspring. We indeed observed two wild males that tended to produce daughters with strong male-biased broods.

Our model estimates the dominance of Z at about 0.7 as Zz females produced about 30% daughters. Although this one-locus two-allele model fits relatively well to the empirical data, the real situation for sex ratio determination in mussels is likely more complicated. It is possible that more than one locus may be involved in the process as well as some external factors that may affect the final sex ratio in the family, i.e. family-specific mortality between sexes. Our model is not provided for quantitative predictions, rather to inspire interest in the investigation of the linkage between sex determination and inheritance of paternal mtDNA in species with DUI.

Table 3.1. Sex ratio of progeny in pedigreed pair matings. The grandparental cross is given above the parents, with the sex ratio in parentheses. The first number of the progeny from each pair mating gives the number of males and the second the total number of progeny scored. The last four male parents were collected in the wild (updated from Kenchington et al., 2002).

	Females										Total
	90WF4X90WM4 (0/20)						90WF5X90WM5 (16/25)				
Males	A	B	D	E	X	Z	F	H	I	J	
90WF5X90WM5 (16/25)											
1		0/23									0/23
2		0/37					7/10				7/47
3							20/24				20/24
4		0/41							24/72		24/113
10			0/52					24/50			24/102
11								35/50		22/29	57/79
12			0/11								0/11
13			0/10					44/61		44/51	88/122
104						33/48					33/48
105					0/50						0/50
106					0/51						0/51
107					0/57						0/57
108									11/17		11/17
90WF7X90WM7 (32/40)											
5		0/37					16/30		26/60		42/127
6		0/19					29/38		18/27		47/84
7		1/46					18/33				19/79
9		0/46					13/20		6/18		19/84
14								24/61			24/61
15								6/12		39/52	45/64
16								29/34		38/47	67/81
17										42/51	42/51

Table 3.1. (Continued)

	Females										Total
	90WF4X90WM4 (0/20)						90WF5X90WM5 (16/25)				
Males	A	B	D	E	X	Z	F	H	I	J	
100						45/52					45/52
101						30/51					30/51
102					0/20	39/50					39/70
103						33/50					33/50
Z101A	10/11[a]			51/55[a]							61/66
Z101A	28/33[b]										28/33
Z101A	19/19[d]										19/19
Z101B	9/10[d]										9/10
Z101C	9/10[d]										9/10
Z103A				19/20[a]							19/20
98WM8	29/37[a]			53/69[a]							82/106
98WM11				26/30[a]							26/30
99WM2	18/24[b]										18/24
00WM22				23/30[c]							23/30
00WM24				19/28[c]							19/28
00WM26	20/20[d]										20/20
01WM4				29/30[d]							29/30
Total	142/164	1/249	0/73	220/262	0/178	180/251	103/155	162/268	85/194	185/230	1078/2024

"a, b, c, d" = Crosses made in 1998, 1999, 2000 and 2001, respectively. Rest of the crosses were made in 1995.

Table 3.2. The model. Females of genotype ZZ produce only sons, Zz females produce sons with probability 1 – k, and zz females produce no sons. The recursion equations for male and female genotype frequencies are obtained from the 3 x 3 matrix of crosses. Numerical equilibrium values are given for k = 0.3 (after Kenchington et al., 2002).

	Female		
Males	*ZZ*, d	*Zz*, h	*zz*, r
ZZ, D	Males	Females with probability k;	Females
Zz, H	Males	Males with probability	Females
zz, R	Males	1-k	Females

Recursion equations

$D' = [Dd + Hd/2 + Dh(1-k)/2 + Hh(1 - k)/4] / [d + h(1 - k)]$

$R' = [Rh(1 - k)/2 + Hh(1 - k)/4] / [d + h(1 - k)]$

$H' = 1 - D' - R'$

$d' = [Dhk/2 + Hhk/4] / (r + hk)$

$r' = [Rr + Hr/2 + Rhk/2 + Hhk/4] / (r + hk)$

$h' = 1 - d' - r'$

Equilibrium frequencies	k = 0.3
$\hat{h} = [1 - (1 - 2k(1 - k))^{1/2}] / [2k(1 - k)]$	0.568
$\hat{d} = [1 - 2\hat{h}(1 - k)] / 2$	0.102
$\hat{r} = (1 - 2\hat{h}k) / 2$	0.330
$\hat{H} = (1 - 4\hat{d}^2) / 2$	0.479
$\hat{D} = (1 + 2\hat{d})^2 / 4$	0.362
$\hat{R} = (1 - 2\hat{d})^2 / 4$	0.158
$\hat{z}_{♀} = (1 + \hat{h} - 2\hat{h}k) / 2$	0.614
$\hat{z}_{♂} = (1 - 2\hat{d}) / 2$	0.398

Table 3.3. Progeny numbers of females from wild populations (after Kenchington et al., 2002).

No.	Female Code	M	N	% M	S (1-k = 0.7)
1	00WF11	24	24	100	-
2	97WF1	29	30	96.7	***
3	00WLF3	23	24	95.8	**
4	00WF16	19	20	95	*
5	00WF3	23	26	88.5	*
6	97WF2	24	30	80	n.s.
7	90WF7	32	40	80	n.s.
8	WGF19	124	156	79.5	**
9	00WF3	19	24	79.2	n.s.
10	00WF7	19	24	79.2	n.s.
11	00WLF5	18	23	78.3	n.s.
12	90WF16	25	33	75.8	n.s.
13	00WF1	18	24	75	n.s.
14	00WF10	17	23	73.9	n.s.
15	WGF20	110	149	73.8	n.s.
16	00WF21	15	21	71.4	n.s.
17	99WF1	82	125	65.6	n.s.
18	90WF5	16	25	64	n.s.
19	WGF66	88	156	56.4	***
20	00WF14	13	24	54.2	n.s.
21	00WF9	11	24	45.8	**
22	WGF53	40	170	23.5	***
23	00WF2	3	21	14.3	***
24	00WF15	3	24	12.5	***
25	00WF18	3	30	10	***
26	WGF31	5	115	4.3	***
27	00WF19	0	10	0	-
28	90WF4	0	20	0	-
29	99WF7	0	23	0	-
30	00WF12	0	23	0	-
31	00WF5	0	24	0	-
32	00WF8	0	24	0	-
33	98WF2	0	28	0	-
34	00WF17	0	29	0	-
35	99WF6	0	30	0	-
36	00WF29	0	30	0	-

M is the number of male progeny produced by the female; N is the total number of progeny scored; %M is the percentage of male offspring; S is the significance of the probability of fit of sex ratio to k (proportion of daughters of a heterozygous mother determined for 0.3) assuming that females were heterozygous; n.s. is not significant; * significant at 0.05; ** significant at 0.01;*** significant at 0.001. Females #8, 15, 19, 22 and 26 are *M. galloprovincialis*; all others are *M. edulis*.

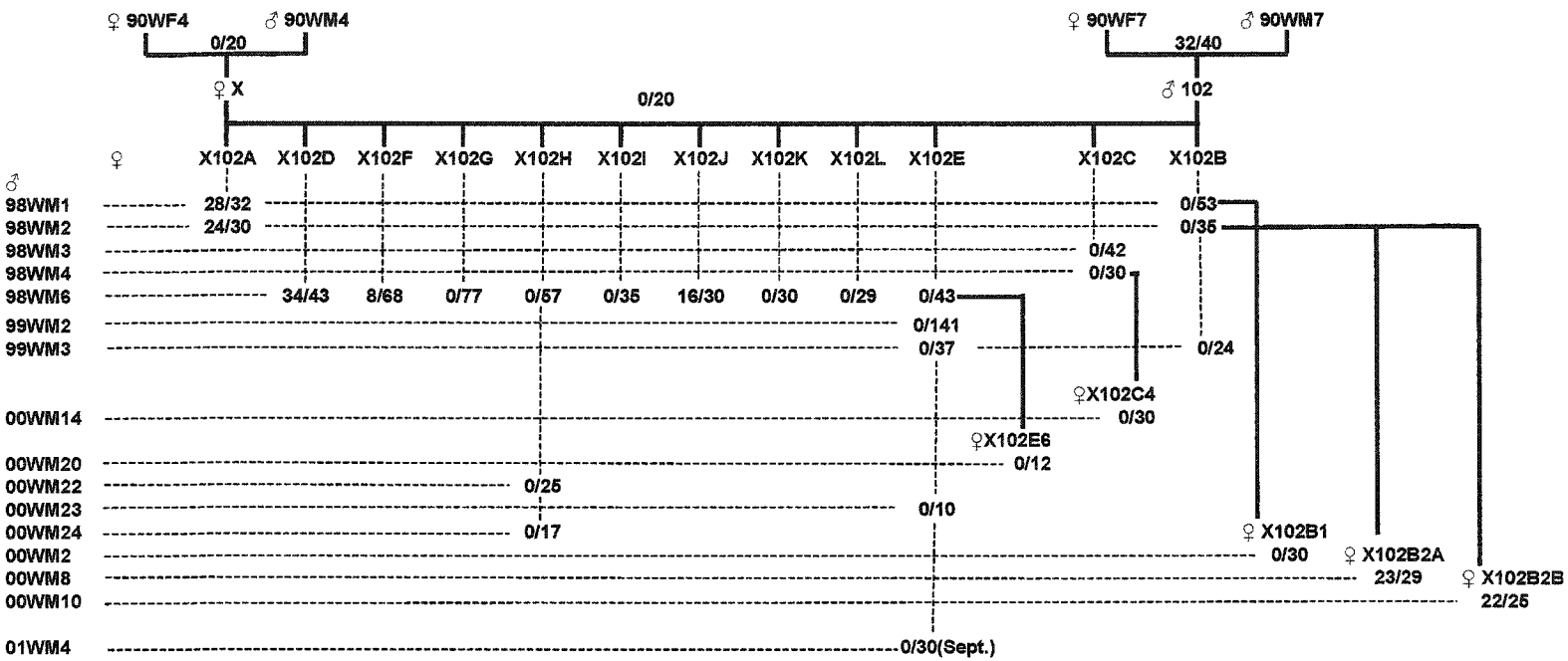


Figure 3.1. Crosses of pedigreed females (and five of their daughters) to wild males. In each sibship, the first number is the number of males and the second the total number of offspring sexed. Bold lines indicate pair matings; dashed lines are to assist in identifying parents (modified from Kenchington et al., 2002). The Cross made in September is indicated. Other crosses were made in June.

Figure 3.2. Crosses of pedigreed females to pedigreed and wild males. (A) females and males originating from son-bearing mothers; (B) females originating from a sonless mother and males from a son-bearing mother. Bold lines indicate pair matings; dashed lines are to assist in identifying parents (modified from Kenchington et al., 2002). All crosses were made in June except two in September (indicated in bracket).

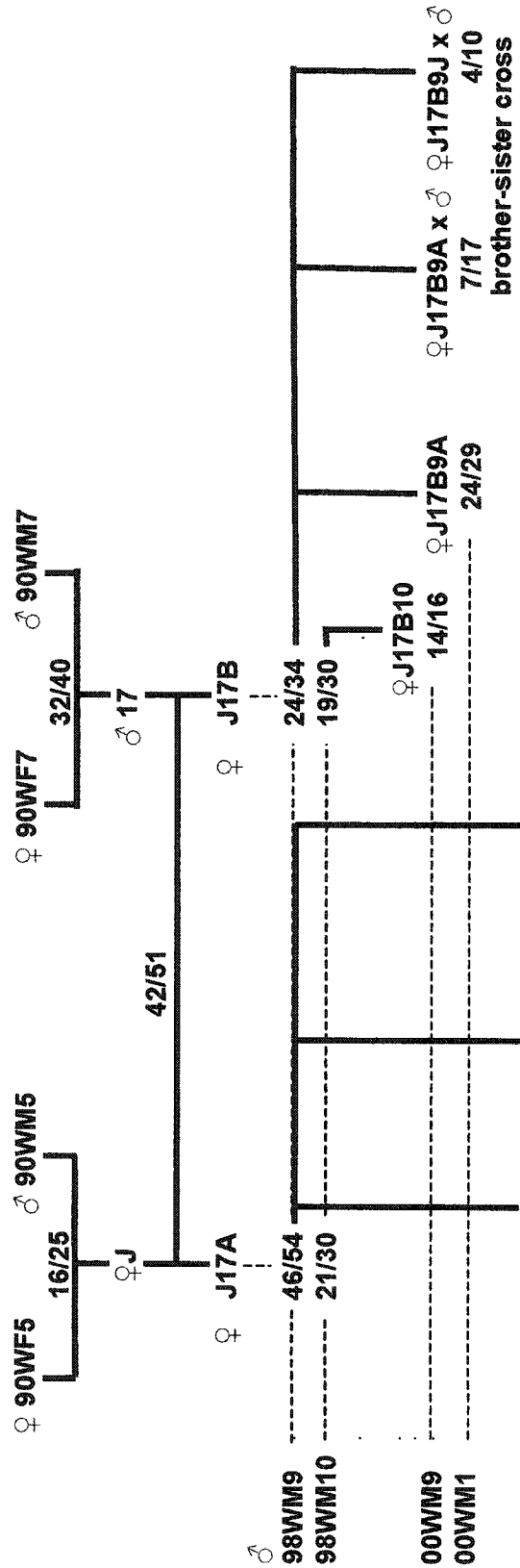

♀ 90WF5
16/25
♂ 90WM5
♀J
42/51
♂ 17
♀ 90WF7
32/40
♂ 90WM7
♀ J17A
♀ J17B
♂
98WM9
98WM10
00WM9
00WM1
46/54
21/30
24/34
19/30
♀J17B10
14/16
♀J17B9A
24/29
♀J17B9A x ♂
7/17
♀J17B9J x ♂
4/10
brother-sister cross
♀J17A9A x ♂
10/19
♀J17A9A x ♂
14/20
♀J17A9B x ♂
16/19
brother-sister cross

Figure 3.2A

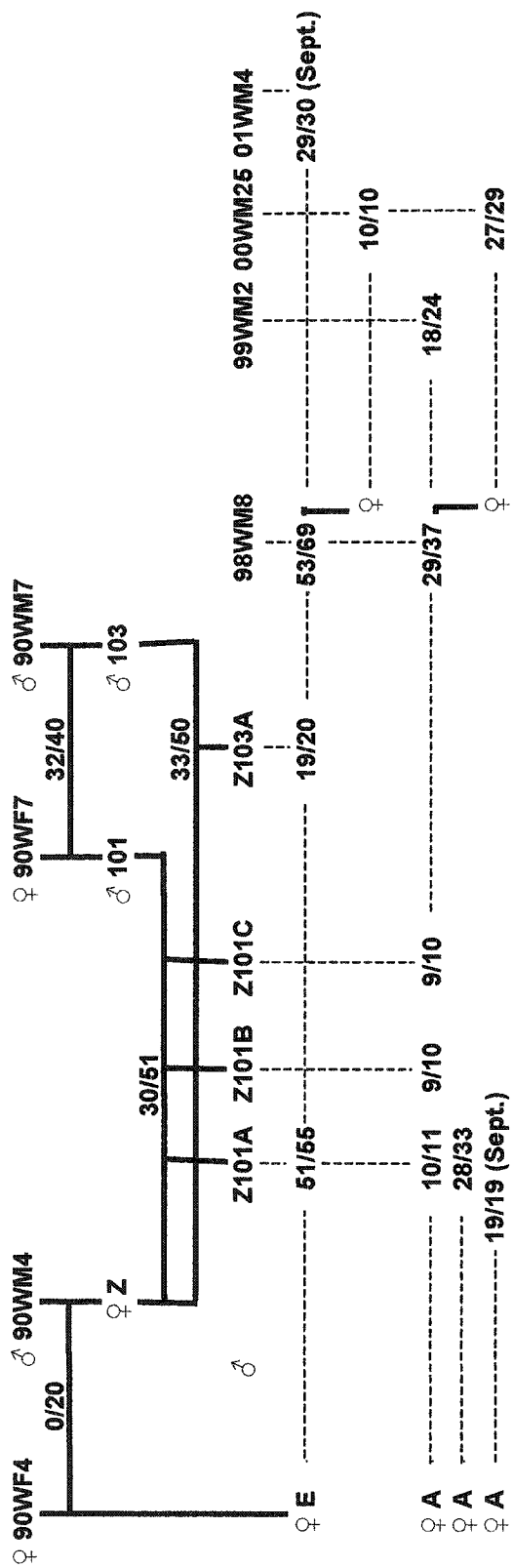

♀ 90WF4
0/20
♂ 90WM4
♀ Z
♂
30/51
♀ 90WF7
32/40
♂ 90WM7
♂ 103
♂ 101
33/50
Z101A
Z101B
Z101C
Z103A
98WM8
99WM2
00WM25
01WM4
♀E
51/55
19/20
53/69
29/30 (Sept.)
♀A
10/11
9/10
9/10
29/37
18/24
10/10
♀A
28/33
♀A
19/19 (Sept.)
27/29

Figure 3.2B

Chapter 4

The Control Region of Maternally and Paternally Transmitted Mitochondrial Genomes in Marine Mussels (*Mytilus* spp.)

Introduction

Hoffmann et al. (1992) published the sequence of 13.9 kb of the total of 17.1 kb of the *M. edulis* mitochondrial genome before the discovery of DUI. Subsequent sequencing of selected mtDNA regions from female and male gonads showed that the Hoffmann et al. (1992) sequence was of the F type (Skibinski et al., 1994b; Stewart et al., 1995; Rawson and Hilbish, 1995). As all segments that were unsequenced were located in the middle of the genes, the complete gene map of the *M. edulis* F-type mtDNA can be determined. Apart from the genes, the genome contains five intergenic sequences of no apparent function, four of which are relatively small (80 to 120 bp) and one of large size (approximately 1.2 kb). The latter sequence, which we call here "large unassigned region" (LUR) to distinguish it from the smaller sequences of unassigned function, is the focus of this study.

Previous studies showed that the two sex-specific sequences differ by more than 20% and that the M genome evolves at a faster rate than the F genome (Skibinski et al., 1994b; Stewart et al., 1995; Rawson and Hilbish, 1995). The faster evolution of the M genome was subsequently observed in freshwater unionid mussels (Liu et al., 1996; Hoeh et al., 1997) and venerid clams

(Passamonti and Scali, 2001). In spite of their large differences in nucleotide substitutions and amino acid replacements, these studies suggested that the F and M genomes do not differ in gene content and arrangement. In view of the different transmission mode, it is important to know whether the similarity of the F and M genomes in terms of protein, rRNA and tRNA content extends to regions that do not code for a gene, but may have regulatory functions. The LUR is the most obvious region for this examination. Because it is the only reasonably large segment of the mussel genome of unknown function, the LUR is the primary candidate for the site of regulatory elements of replication and transcription of mussel mtDNA. Indeed, Burzyński et al. (2003) have suggested that in addition to being the control region, the LUR may also contain F-specific and M-specific sequences that are responsible for their sex-specific transmission.

Here I produced strong evidence in favor of the hypothesis that the LUR is the control region of both the F and M genomes. Further I show that the F and M LURs have features that differentiate them from the rest of the genome. The middle part of LUR has remained practically undifferentiated in the two genomes, whereas another segment has evolved at a much higher rate than any other part of the genome. These features are discussed in terms of the constraints that are placed by the nuclear genes on the control region of mtDNA and of the possibility that the LUR may contain, in addition to control elements, sequences that determine the transmission route of the molecule.

Materials and methods

Sample collection and identification

Mussels were collected from Canada (Lunenburg Bay, Nova Scotia, individuals: w22, w24, w128 and w143), USA (Totten Inlet, Puget Sound, Washington State, individual #6) and France (Morgat, Bay of Douarnenez, West Brittany, individuals #1 and #12), and transported alive to the laboratory, where their sex was determined by examining the presence of sperm or eggs in the gonad under a light microscope (Table 4.1). Total DNA was extracted from gonad tissue using a modified salt extraction procedure (Miller et al., 1988). Species identification was performed by PCR amplification of the internal transcribed spacer (ITS) of the nuclear ribosomal RNA genes and *Hha*I restriction of the PCR product to produce species-specific restriction patterns (Heath et al., 1995). As a second assay for species identification, I used the polyphenolic adhesive protein *Glu* gene as described by Rawson et al. (1996a).

DNA amplification, cloning and sequencing

Total DNA from gonad tissue was used as the template to amplify the large unassigned region (LUR) with primers UNFOR1 5'-TTG CGA CCT CGA TGT TGG C- 3' and UNREV1 5'-AGC TCA CCA CCT ATT CCT C-3', which correspond to nucleotide positions 3218-3236 and 4654-4636 of segment-5 of the *M. edulis* mitochondrial F genome (Hoffmann et al., 1992). One band was obtained in females, and two in males. Of the two male bands, one was weaker and of the same size as the female band, the other was stronger and of smaller

size. To enhance the larger band in males, I used the primer UNFOR1 given above in combination with primer UNREV2 5'-GCG TTA GTG TTA TAT GCA G-3', which corresponds to position 4726-4708 of segment-5 in Hoffmann et al. (1992). The PCR mixture consisted of 2 µl of template DNA, 0.8 mM of each primer, 1 mM dNTP, 2.5 mM $MgCl_2$ and 1 U *Taq* polymerase (MBI Fermentas) in the buffer supplied by the company in a total 25 µl volume. The cocktail was heated initially at 94°C for 3 min and then incubated at 94°C for 1 min, 53°C for 1.5 min and 72°C for 1 min for 40 cycles and 72°C for 6 min for a final extension.

The amplified products were visualized on a 1% regular agarose gel, and the corresponding bands were excised. DNA was recovered using the UltraClean DNA purification kit (Mo Bio), and cloned on pGEM-T vector (Promega) following the procedure provided by the supplier. Positive clones from each individual were confirmed by PCR amplification with the same primers and conditions as above. One randomly chosen clone from each female and two from each male, one with the small and another with the large band, were sequenced commercially from both directions with 80%-85% overlap using either LICOR 4200 or ABI 373 automated sequencer.

Sequence analysis

Sequences were aligned with the aid of the computer program ClustalX (Thompson et al., 1997). Further corrections of the alignment were performed manually in order to maximize the sequence similarity, particularly for the two size-variable regions that flank the relatively conserved central region (Appendix

A). Kimura 2-parameter genetic distances (Kimura, 1980) and Neighbor-Joining (Saitou and Nei, 1987) trees were calculated using the computer program MEGA Version 2.1 (Kumar et al. 2001). Pair-wise alignments used to calculate K-distances are available upon request. The DnaSP version 3.53 program (Rozas and Rozas, 1999) was used to carry out the sliding-window plot (Figure 4.1). Prediction of potential secondary structures was made using the Web version of the *mfold* software version 3.1 by Michael Zuker (Zuker, 2003).

Results

I have included in this study seven individual mussels, three females, two of which belonged to the species *M. edulis* and one to its sibling species *M. galloprovincialis,* and four males, two of which belonged to *M. edulis* and two to *M. galloprovincialis*. I focused on a segment of the mitochondrial genome, which includes part of the *16S-rRNA* gene, the LUR and a small part of the tyrosine tRNA gene. Each sequence was characterized as F type or M type on the basis of the part of the *16S* gene, whose F type and M type sequences are known from previous studies (Rawson and Hilbish, 1995). Each female yielded an F-like sequence, as is the case with most female mussels. Male mussels normally contain an M-type and an F-type sequence, but exceptional males that lack an M-type sequence at the examined part of the molecule can be found in varying frequencies in natural populations. All four males used in this study yielded an F-type and an M-type sequence. Thus, the whole set of data consisted of four M-type and seven F-type sequences. The eleven nucleotide sequences are given in

Appendix A together with the corresponding sequence from Hoffmann et al. (1992).

The three parts of the large unassigned region (LUR)

The large unassigned region of the mussel mtDNA was first identified by Hoffmann et al. (1992) as the non-coding region between the *16S-rRNA* and the tyrosine tRNA gene. The 3'-end of the LUR can be easily defined as the nucleotide preceding the first nucleotide of the tyrosine tRNA gene, which itself can be identified from the tRNA folding pattern. In contrast, the 3'-end of the 16S-*rRNA* gene cannot be readily identified because the length of the gene as well as the sequence of its 3'-end varies among species. This introduces a degree of arbitrariness regarding the 5' starting point of the LUR. For consistency, we have used the same starting point as Hoffmann et al. (1992).

Simple visual inspection of the twelve aligned LUR sequences (Appendix A) suggests that there is a high degree of differentiation between the set of F sequences from the set of M sequences. The first and most marked difference is the presence of a large number of indels of varying length. Deletions are more common in the M sequences. Among the four M sequences examined the length varied from 925 to 937 bp, whereas among the eight F sequences the length varied from 1156 to 1194 bp. Further inspection shows that the differences in indels are not randomly distributed over the whole length of the LUR. Starting from the 5'-end, indels are commonly found for about half of the length. This first part of LUR is followed by a stretch of more than 300 bp in which indels are

practically absent. Indels become again prominent in the remaining part of LUR. This part is dominated by strings of adenines and guanines. This indel-based division of LUR in three parts, is further supported by the difference in the degree of nucleotide differentiation between the F and M sequences. Figure 4.1 gives the two-parameter Kimura distance for consecutive lengths of 25 bp of one F and one M sequence, after removal of indels. The compared sequences, em.w143-F and em.w143-M, were chosen because they were derived from the same individual, but the result is basically the same for any pair of F/M sequences (see below). The distances are consistently high in the first part of the LUR, they drop dramatically in the second and rise again in the third, even though not as high as in the first.

Based on these observations, I divided the LUR into three regions, using arbitrary internal points (see Appendix A). As the 3'-end of the first indel-rich region I have chosen the last indel that differentiates the two sets of sequences. This is a one-nucleotide deletion that is fixed in the F sequences but does not occur in any M sequence. As the 3'-end of the second region I defined the last pyrimidine before the start of the first adenine string. I refer to the three segments as variable domain 1 (VD1), conserved domain (CD) and variable domain 2 (VD2).

VD1 differs most between the two types of genomes, both in terms of total length and nucleotide divergence. Its length in the F genomes is 654 bp (Table 4.1). An exception occurred in a *M. galloprovincialis* sequence (sequence gm.6-F) that carries a 36 bp insert (Appendix A). The insert is an almost perfect repeat

of the preceding 36 bp, and has most probably resulted by replication slippage (Levinson and Gutman, 1987). The length of VD1 is 490 bp in two M genomes, 488 bp in a third and 500 bp in the fourth. The latter size was found again in a *M. galloprovincialis* individual (sequence gm.12-M) and is due to an insert of ten adenines that is flanked by strings of adenines, also a possible result of replication slippage. In total, I found 48 fixed deletions in the M sequences, ranging from one to 13 nucleotides in length, with the majority of them of one or two nucleotides. In contrast, I found only nine fixed deletions in the F genome, of which seven are of one nucleotide. This makes the VD1 of the M type genome shorter by about 160 bp, which accounts for most of the length difference between the two types. Sequence divergence in this domain from all pair-wise F/M comparisons, after removal of positions with gaps, varied from 0.468 to 0.535 with a mean of 0.496 (SE = 0.008).

The high divergence of VD1 is abruptly terminated at the conserved domain (CD). There are no fixed indels in this domain for either the M or the F sequences, but only occasional small (one or two-nucleotide) deletions in some sequences of both the M and F types. As a result, the length of the domain is similar in both types varying between 335 to 337 pb. Sequence divergence, after removal of positions with gaps, from all pair-wise F/M comparisons varied from 0.006 to 0.025 with a mean of 0.016 (SE = 0.001).

The third domain (VD2) is dominated by strings of purines (74.4% A+G, compared to 55.2% for the first region and 51.9% for the second) and contains one 67 bp deletion that is present in all M sequences and absent in all F

sequences. This is the largest indel for the whole LUR. There is only one other fixed indel, a one nucleotide deletion in the F sequences at the very end of the segment. Sequence divergence, after removal of positions with gaps, from all pair-wise F/M comparisons varied from 0.079 to 0.184 with a mean of 0.122 (SE = 0.006).

Phylogenetic trees based on the Kimura two-parameter distances for the three parts of LUR and the sequenced part of the *16S-rRNA* gene produced a clear separation of the F and M sequences for all four parts, except for one F sequence (gm.6-F) which clustered with the M sequences in the tree based on the CD region (Figure 4.2). As expected, the F and M sequences form distinct and compact clusters separated by long branches in the tree based on VD1. In contrast, the distance that separates the two clusters is only twice as long as the mean distance between same type sequences in the tree based on the CD. The third region produced an intermediate pattern. The tree based on the *16S-rRNA* gene is typical of the trees for the F and M lineages of *Mytilus edulis* observed for various other regions of the genome, such as parts of the *COI*, *COIII* and *16S*-rRNA genes (Rawson and Hilbish, 1995; Stewart et al., 1995; Hoeh et al., 1996a).

The LUR as the control region of the mussel mtDNA

Because the LUR is, as noted, the largest segment of the mussel mtDNA that does not code for a protein or an RNA, it was considered as the prime candidate for the region that contains the control elements for the replication

and/or transcription of the genome. In order to produce evidence for this hypothesis I have used information from published sequences of well-characterized control regions in other animal species.

Direct comparison of the whole LUR sequence with mtDNA non-coding sequences of other molluscan species (*Albinaria coerulea*, Hatzoglou et al., 1995; *Katharina tunicata*, Boore and Brown, 1994; *Loligo bleekeri*, Tomita et al., 2002) produced no significant degree of similarity. Comparisons with well studied control regions of other species, such as *Drosophila* (Goddard and Wolstenholme, 1978, 1980; Clary and Wolstenholme, 1985, 1987; Lewis et al., 1994), sea urchin (Jacobs et al., 1989; Mayhook et al., 1992; Elliott and Jacobs, 1989; Cantatore et al., 1990) and human (reviewed by Clayton, 2000; Saccone et al., 1999, 2000; Shadel and Clayton, 1997; Taanman, 1999) also did not produce any evidence of similarity. However, the clear recognition of three parts in the LUR, of which the central part is conservative and the two flanking parts are variable, is very suggestive of a structural similarity with the human mtDNA control region (Saccone et al., 1991, 1999; Sbisà et al., 1997; Stoneking, 2000). This tri-partite structure is well established in mammals (Sbisà et al., 1997; Saccone et al., 1991, 1999) and has also been observed in other vertebrates (Saccone et al., 1987; Randi and Lucchini, 1998; Marshall and Baker, 1997; Delport et al., 2002; Ray and Densmore, 2002), but there is no evidence for its presence in invertebrates.

Figure 4.3 contrasts the three parts of LUR with those of the human control region. It can be observed that the central domain is about the same

length in the mussel LUR and the human control region and that the flanking regions are of different length, with the VD1 being longer and the VD2 being shorter than the corresponding human domains. In each domain of the human control region I have marked the position of the element for which there is wide consensus that they play a key role in the replication of the heavy strand and the transcription of both strands. Gray boxes below the line indicate binding sites of transcription factor mtTF1 and black boxes indicate the binding sites of other factors (Fisher et al., 1987). Black boxes above the line are motifs that are relatively conserved in mammals, despite the fact that they are located within fast evolving domains. Open bars indicate the secondary structures. The TAS (termination associated sequence, Doda et al., 1981) element is considered as the termination site of heavy strand synthesis, even though Sbisà et al. (1997) have suggested that this role is played by two longer motifs (which they have called ETAS1 and ETAS2), both within the first domain. The CSB (conserved sequence blocks, Chang and Clayton, 1985) elements are assumed to play a role in replication and/or transcription owing to their conserved sequence. This role is better established for CSB1 than CBS2 or CBS3, both of which may even be absent in some mammalian species (Saccone et al., 1991; Sbisà et al., 1997).

I have examined the F and the M LUR sequences in an attempt to identify motifs that resemble the above elements of the human control region. Figure 4.4 shows the parts of LUR sequences that produce the best matches with the corresponding human motifs. There is a considerable danger that some of these matches may result by chance alone owing to the small length of the motifs.

However, this danger is considerably reduced by the fact that there is a good agreement between the positions of the putative motifs in the F and M LUR and the positions of the corresponding motifs in the human control region (Figure 4.3).

I have found no sequence in the LUR that would match the ETAS1 or ETAS2, as defined by Sbisà et al. (1997), but I have identified a sequence in both the F and M LURs that resembles the conventional TAS element and is located in the same domain as in the human control region. In the M LUR there is, however, another small sequence that produces a better match (TAS(alt), Figure 4.3). I have found CSB1 and CSB3 in the same order and in the distal part of LUR as in the human control region, but I could not find a sequence that would sufficiently match the CSB2 motif. This must not come as a surprise since it is known that the number of CSB elements varies among mammalian species and in some cases it is reduced to only one (Saccone et al., 1991; Sbisà et al., 1997). With regard to the binding sites of transcription factors, I have identified four of them [mt3(L), the first binding site of mtTF1, the second binding site of mtTF1 and mt4, Suzuki et al., 1991]. Three of them are in the right order and in the same region as in the human control region. This is not the case with the putative second binding site of mtTF1, which is located before rather than after the first binding site of mtTF1. These two binding sites may play the role of LSP and HSP (major light strand promoter and major heavy strand promoter, Chang and Clayton, 1984), for which we did not find LUR sequences with sufficient similarity.

In the human control region there is a tRNA-like secondary structure covering approximately the entire length from the beginning of the first region to the beginning of TAS (Brown et al., 1986). I have identified a similar structure in both LURs that is also in the area of TAS (Figure 4.5). I note that these structures do not start exactly at the beginning of the LUR and that in the case of the M molecule the structure includes the entire TAS. Moreover, an alternative tRNA-like structure is found in the M molecule downstream of the alternative TAS. Interestingly, the sequence between the two tRNA-like structures of the M LUR is of the form $A_n(TA_m)TA$ (where n and m vary from 11 to 36 and from 13 to 14, respectively, between M sequences). A sequence of this form was shown to correspond to the second termination point of transcription in mice (Camasamudram et al., 2003). The existence of two putative TAS, two tRNA-like structures and of a $A_n(TA_m)TA$ sequence in the M LUR is a marked difference between the LURs of the two mtDNA genomes.

No other secondary tRNA-like structure could be identified in either the F or M LUR, but several stem-and-loop structures can be recognized. The longest of these is located in the CD region and covers about 51% of the domain (Figure 4.6). Interestingly, a similar structure can be found in the human central domain (covers about 67% of CD). The functional role of this structure, if any, is unknown. We note that, even though of considerable size, the structure contains only one functional element, mt3(L). Therefore, it may not interfere with the functions of other elements, such as that of TAS or CSB, which are located in different domains. On the other hand, a possible role for this long stem-and-loop

structure may explain why the sequence of the central domain is highly conserved.

Discussion

Here I examined the assumption that the LUR is the control region of the mussel mitochondrial genome. The comparison with the human mtDNA control region produced evidence in favor of this assumption. The LUR has a tri-partite structure, with the central domain being much more conservative than the flanking domains – a feature characteristic of the mammalian control region (Figure 4.3). I have identified in the LUR seven sequence motifs that occupy similar positions and show considerable homology with motifs that are known to play a crucial role in the replication and transcription control of the human mtDNA (Figure 4.3 and 4.4). Finally, I have found sequences in the LUR that have the potential to form similar secondary structures at about the same positions as in the human control region (Figure 4.3 and 4.5). All these features are present in both the maternally and paternally transmitted mussel genomes and imply strongly that the LUR contains the sequence elements that control replication and transcription of the mussel mtDNA.

Even though both LURs have the landmarks of the control of replication and transcription, they differ in many other aspects. A certain amount of divergence is expected from the mere fact that the two genomes have been separated before the emergence of *M. edulis, M. galloprovincialis* and *M. trossulus* from their common ancestor (Rawson and Hilbish, 1995; Stewart et al.,

1995). The family Mytilidae first appears in the fossil record in the Devonian, 400 million years ago (MYA), while the divergence of the subfamilies Mytilinae and Modiolinae occurred sometime in the late Paleozoic (approx. 250 MYA) (Soot-Ryen, 1955). *Mytilus* has been recorded from the Jurassic (approx. 155 MYA; Pojeta et al., 1987) with *M. edulis, M. galloprovincialis* and *M. trossulus* emerging as separate species in the late Tertiary, with *M. edulis* being ancestral and first appearing in the Miocene (approx. 30 MYA) (Soot-Ryen, 1955; Bernard, 1983). An upper age estimate for this event would be the early Eocene (40-50 MYA), when the common ancestor of these taxa is assumed to have originated in the North Pacific (Kafacov, 1987, cited in Vermeij, 1992). *M. galloprovincialis* may have more recent antecedents coincident with the formation of the Mediterranean Sea 1-2 MYA in the Pleistocene (Barsotti and Meluzzi, 1968). Kenchington et al. (1995) using 18S RNA sequences have estimated that the divergence of the *Mytilus* species occurred in a star radiation event about 33 MYA in the early Oligocene. On the other hand, Varvio et al. (1988) suggested from electrophoretic enzyme data that these sibling species have evolved roughly 1-3 MYA.

As noted, the ~20% divergence at the sequenced *16S*-rRNA part of the gene is typical of coding parts of F and M mtDNA. The divergence of the second variable domain of LUR is not statistically different from this, but the first variable domain and the central domain differ strongly, in opposite directions. To evaluate the significance of these deviations I collected data on the divergence of the homologous regions of mtDNA of seven primate species from the human mtDNA

(Table 4.2). I have also obtained estimates of intra-genomic variation of the human and mussel mtDNA. The first observation is that intra-M, intra-F and intra-human diversity is comparable for all four segments listed (Table 4.2), except HVS1 (or VD1) in M molecules, where it appears to be higher, but not by a large margin. The comparison of regional divergences across primate species and gender-specific genomes leads to some interesting results. Again using the sequenced part of the 16S-*rRNA* gene as reference, I asked if, within primates, there is linear correlation between divergence at this part of the mtDNA and the three domains of the control region. This correlation is very good for the CD domain ($r = 0.93$, $P = 0.002$; Figure 4.7), reasonably good for HVS1 ($r = 0.77$, $P = 0.042$; regression not shown) and poor for HVS2 ($r = 0.64$, $P = 0.122$; regression not shown). Given that the divergence at both variable regions reaches high values for species outside the human-pan-gorilla triad (making the possibility of substitution saturation high) the linearity of the last two comparisons is questionable.

Using the regression of the conservative domain, the expected divergence between the F and M genomes is 0.308, *i.e.* approximately twenty times more than observed. Theoretically, the discrepancy can be the result of an abnormally high rate of divergence of the F and M molecules at the *16S-rRNA* gene (and also at other coding regions) and a normal divergence at the CD, or of an abnormally strong selection against nucleotide substitution at the CD. It is difficult to decide between these two alternatives, given the observation that the mussel mtDNA evolves faster than in other animals (Hoeh et al., 1995). Yet the fact that

in primates the conserved domain has diverged at about the same rate as the 16*S*-rRNA (Figure 4.7) suggests that there is no particularly strong pressure against substitutions at this domain across primate species. In view of this, the observation that the intra-F or intra-M diversity barely exceeds the inter-genome divergence and it is of the same magnitude as the intra-human diversity, suggests that there is a constraint against divergence of the F and M central domain of LUR. The obvious difference between the F and M genomes of *Mytilus* and the human and lemur (two pairs with the same divergence at the *16S*-rRNA gene) CD domains is that in the first pair the CD has been evolving under the same nuclear background, and in the second under nuclear backgrounds that have separated from each other (and have evolved independently) since 81.5 MYA or more (Tavarè et al., 2002).

Our explanation of the high conservatism of the CD segment of the mussel mtDNA is that this domain contains sequences that are highly tuned to interactions with products encoded by nuclear sequences. This imposes a strong co-evolutionary constraint between nuclear and mtDNA sequences, the consequences of which can be seen only if the two genomes are co-separated into different evolutionary paths. I have no idea how this constraint may operate at the molecular level, but the observation that almost the entire CD can form a stem-and-loop structure (Figure 4.6) may have something to do with this. Three of the 61 base-pairs of the F stem and 4 of the 60 base-pairs of the M stem are imperfect purine-pyrimidine pairings (in the human stem the number is 1 out of 56). If this low rate of mismatching causes no substantial impairing of the function

of the secondary structure, it would generate a pressure for restoring normal pairing by mutation at either the purine or the pyrimidine site. This will result in gradual divergence, which, if secured by the severance of gene flow through speciation, could accumulate to the levels seen among species of primates (Table 4.2).

This same logic may be extended to the other domains of LUR. From Table 4.2 it can be seen that the first domain evolves at about the same rate as in primates (using always *16S-rRNA* as reference), and the third domain at a lower rate, resembling the central domain. Taken on face value, these observations mean that the second variable domain is also under a nuclear, species-specific constraint, which agrees with the high concentration of control motifs in this region (Figure 4.3). The high divergence rate of the first variable domain is also consistent with the low density of elements for the control of replication and transcription, which may "leave room" for sequences that may play a role in sex-specific transmission of mtDNA. These observations are not, however, as reliable as for the CD, first because the amount of divergence for four of the seven primate comparisons is too high to obey a linear correlation and second because the two variable domains of LUR are internally heterogeneous in divergence rate (Figure 4.1).

My results are consistent with the hypothesis that the first domain of LUR may contain sex specific transmission sequences, as proposed by Burzyński et al. (2003). These authors examined the LUR from several female and male individuals of *M. trossulus* from the Baltic Sea. In this population an introgression

of the *M. edulis* mtDNA has resulted in a nuclear-mtDNA mosaic, in which the nuclear genome is of the *M. trossulus* and the mtDNA of the *M. edulis* type. The majority of LURs obtained by Burzyński et al. (2003) were of the F or M *M. edulis* type, as expected. But they also found two recombinant types in male gonads. Both these types carried a segment of M type sequence at the first variable domain in LURs that otherwise were of the F type. This observation and the results we report here about the function of LUR and its regional divergence between F and M molecules give a high priority to the search for sex specific transmission sequences (SSTS), as one of the most promising avenues to the understanding the molecular mechanism of DUI.

Table 4.1. Origin, type and the parts of eleven mtDNA sequences studied.

Species	Sex	Individual Code	Sequence Code	mtDNA Type	Length (in bp)						
					16S-rRNA	LUR				*tRNA-Tyr*	Total
						VD1	CD	VD2	Total		
M. edulis	♀	w22	ef.w22-F	F	227	654	337	168	1159	12	1398
		w24	ef.w24-F	F	227	654	335	167	1156	12	1395
	♂	w128	em.w128-F	F	227	654	337	167	1158	12	1397
			em.w128-M	M	228	490	337	101	928	12	1168
		w143	em.w143-F	F	227	654	337	168	1159	12	1398
			em.w143-M	M	227	490	336	102	928	12	1167
M. galloprovincialis	♀	#1	gf.1-F	F	227	654	337	169	1160	12	1399
	♂	#6	gm.6-F	F	227	690*	336	167	1194	12	1433
			gm.6-M	M	227	488	336	101	925	12	1164
		#12	gm.12-F	F	227	654	337	167	1158	12	1397
			gm.12-M	M	227	500**	337	100	937	12	1176

*: contains an insert of 36 bp long; **: contains an additional 10 base-long A-stretch.

Table 4.2. Kimura two-parameter nucleotide divergences (with standard error in parentheses) at four mtDNA regions. Intra-F, intra-M and F versus M divergences are the means of comparisons involving all F and M sequences of Appendix A. The intra-human divergence was obtained by using all sequences in Ingman et al. (2000). For the divergence of human and other primates, the following sequences were used. Human: the revised Cambridge reference sequence, GenBank Accession No. NC 001807, corrected according to Andrews et al. (1999); chimpanzee: GenBank Accession No. NC 001643; gorilla: GenBank Accession No. NC 001645; orangutan: GenBank Accession No. NC 001646; gibbon: GenBank Accession No. NC 002082; baboon: GenBank Accession No. NC 001992; lemur: NC 004025; tarsius: GenBank Accession No. NC 002811. For lemur the divergence at HVS2 was calculated after removal of a sequence of 336 nucleotides, which consists of multiple repeats of CAYAYG.

	16S	HVS1 (VD1)	CD	HVS2 (VD2)
among F (overall mean)	0.017 (0.005)	0.018 (0.003)	0.009 (0.003)	0.023 (0.007)
among M (overall mean)	0.007 (0.004)	0.054 (0.008)	0.004 (0.003)	0.036 (0.016)
F vs. M (mean)	0.197 (0.034)	0.504 (0.049)	0.016 (0.005)	0.123 (0.035)
among humans (overall mean)	0.002 (0.001)	0.027 (0.004)	0.006 (0.002)	0.013 (0.003)
human vs. chimpanzee	0.085 (0.021)	0.198 (0.027)	0.069 (0.016)	0.180 (0.024)
human vs. gorilla	0.106 (0.024)	0.256 (0.038)	0.058 (0.014)	0.198 (0.027)
human vs. orangutan	0.089 (0.021)	0.262 (0.036)	0.075 (0.016)	0.284 (0.035)
human vs. gibbon	0.138 (0.028)	0.452 (0.047)	0.117 (0.021)	0.303 (0.036)
human vs. baboon	0.130 (0.026)	0.480 (0.049)	0.229 (0.031)	0.358 (0.039)
human vs. lemur	0.193 (0.034)	0.405 (0.046)	0.294 (0.038)	0.348 (0.037)
human vs. tarsius	0.205 (0.035)	0.486 (0.052)	0.337 (0.041)	0.307 (0.032)

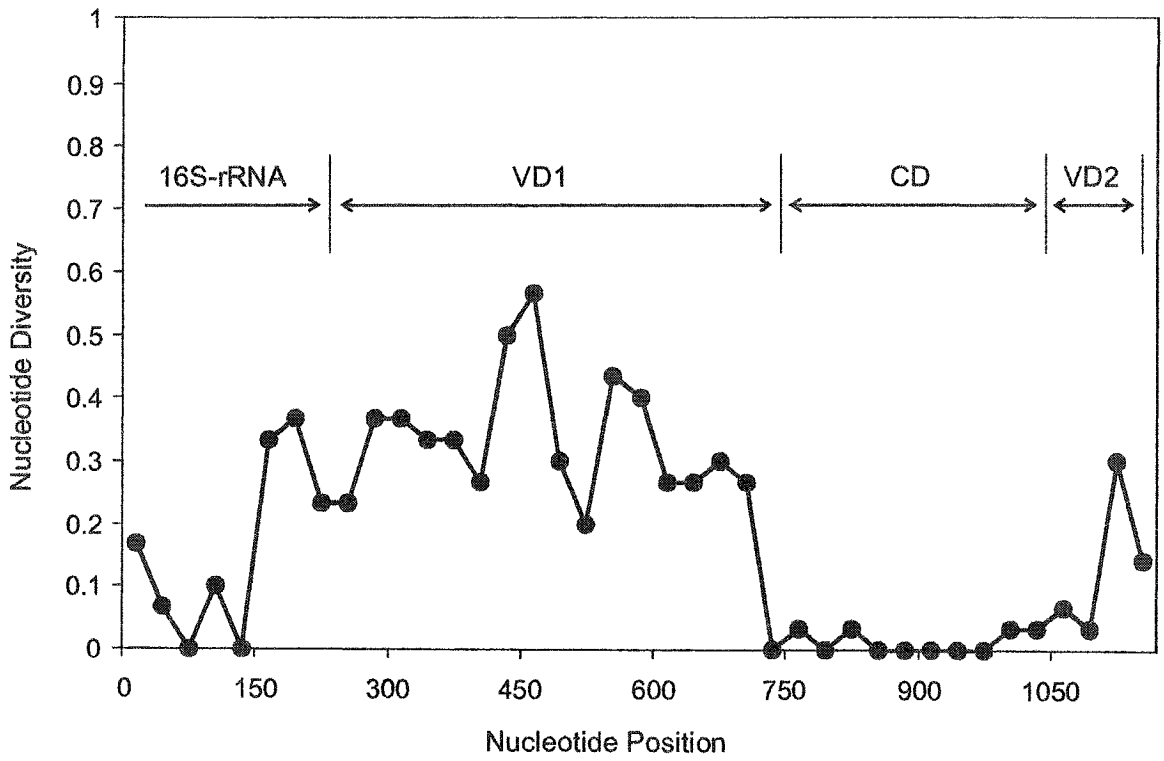


Figure 4.1. Nucleotide divergence in steps of 30 bp between the F and M sequences of individual w143 after removal of gap positions. Dots correspond to the midpoint of the sliding window. Arrows mark the four segments that comprise the aligned sequences.

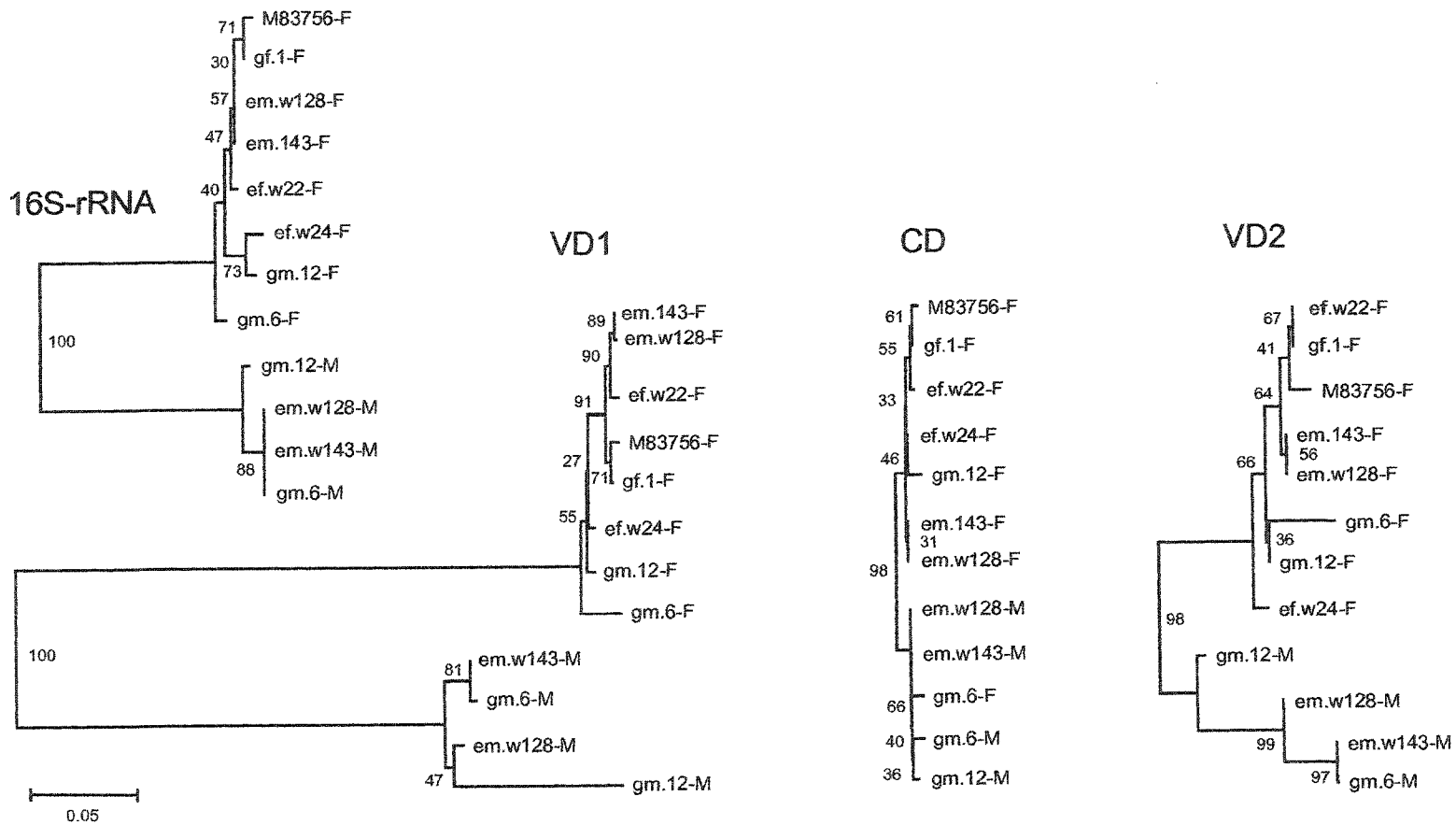


Figure 4.2. Neighbor-joining trees of the twelve sequences of Appendix A for each of the four segments (16S-rRNA and the three regions of LUR). Numbers indicate percent bootstrap support for 1000 replicates .

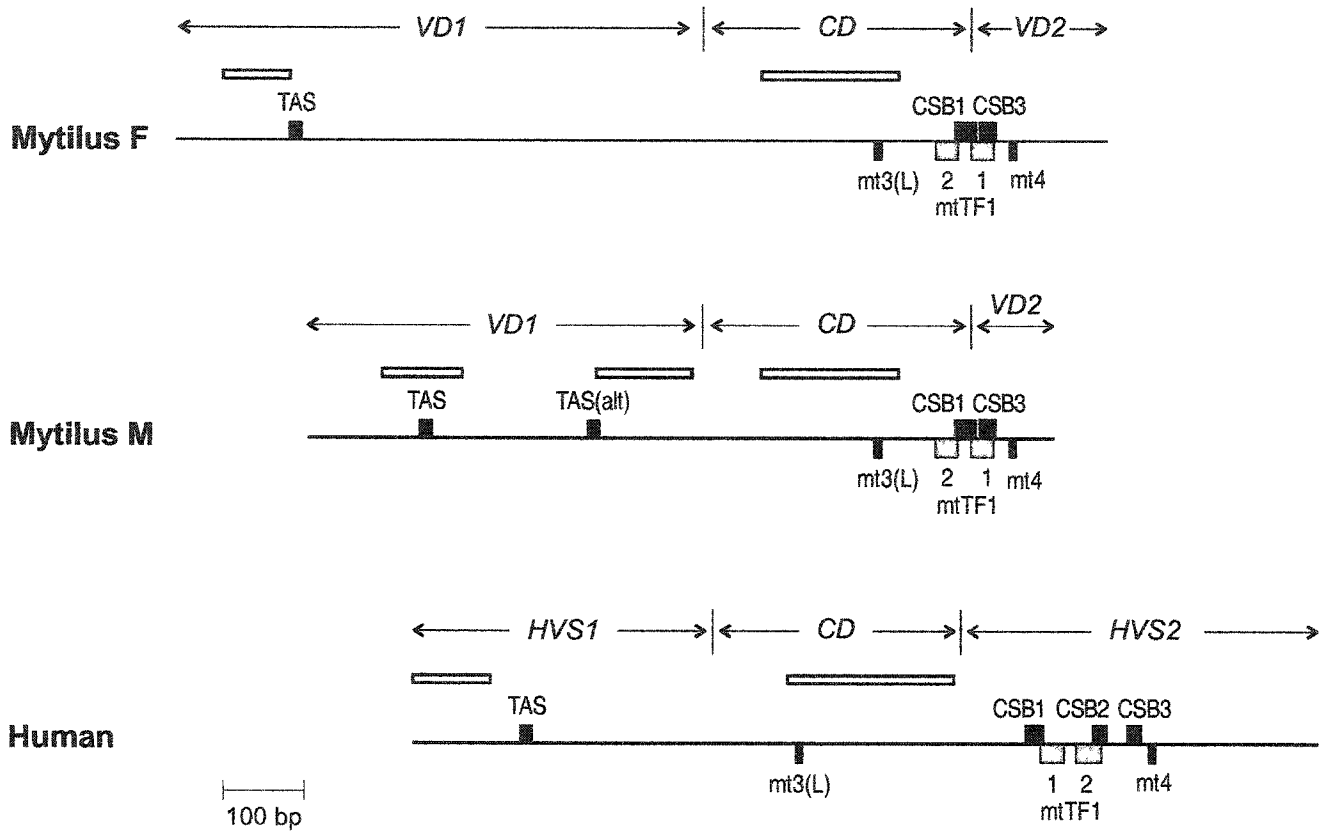


Figure 4.3. Graphical presentation of the corresponding domains of the human mtDNA control region and the LUR of the F and M genomes (see Figure 4.4 and 4.5). Human sequence used is the revised Cambridge reference sequence, GenBank Accession No. NC 001807, corrected according to Andrews et al. (1999) and annotated by Wallace and Lott (2003). See text for explanation of symbols.

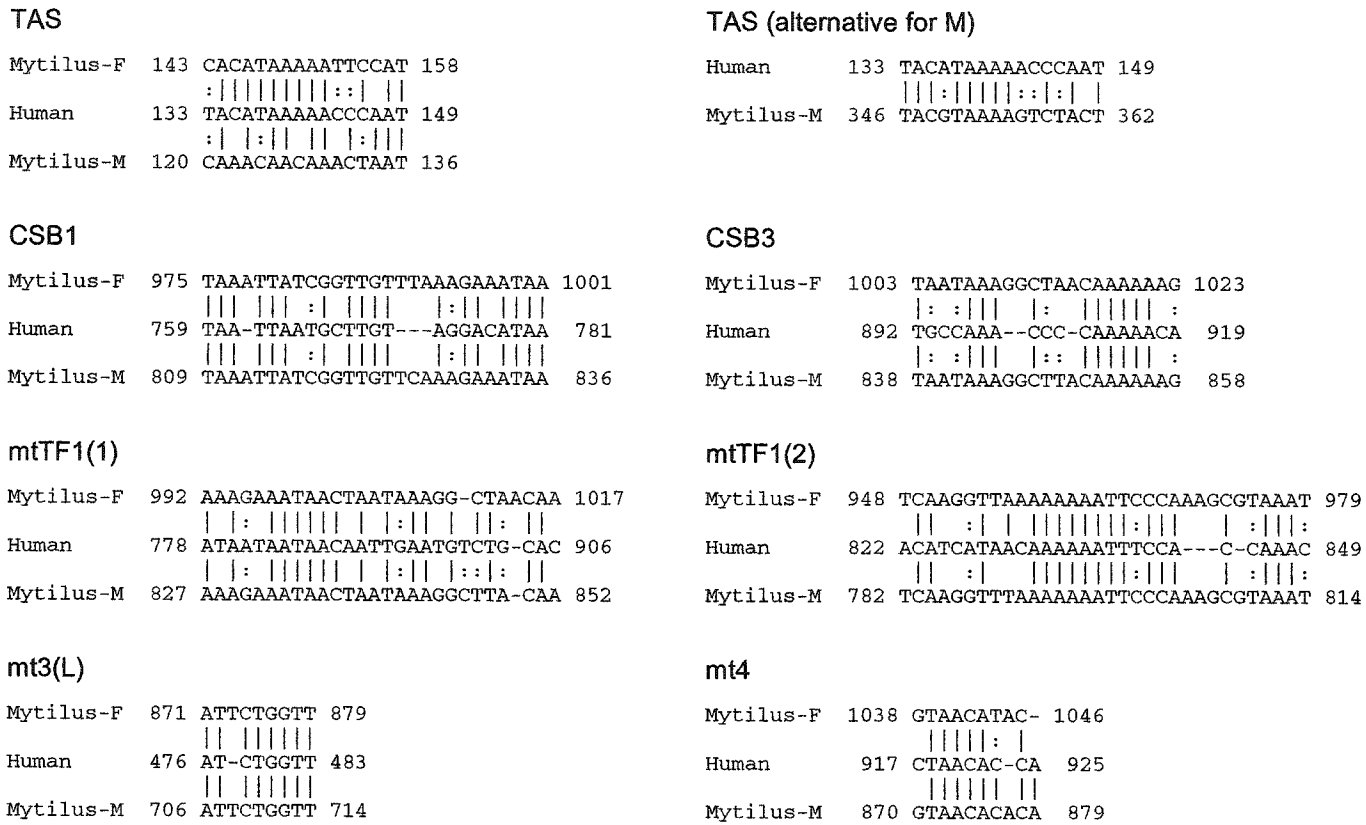


Figure 4.4. Sequence matching of some well defined functional elements of the human mtDNA control region with corresponding putative elements of the LUR of the F and M genomes. The actual sequences shown are from ef.w22-F (F genome) and em.w143-M (M genome) in Appendix A. See text for explanation of symbols.

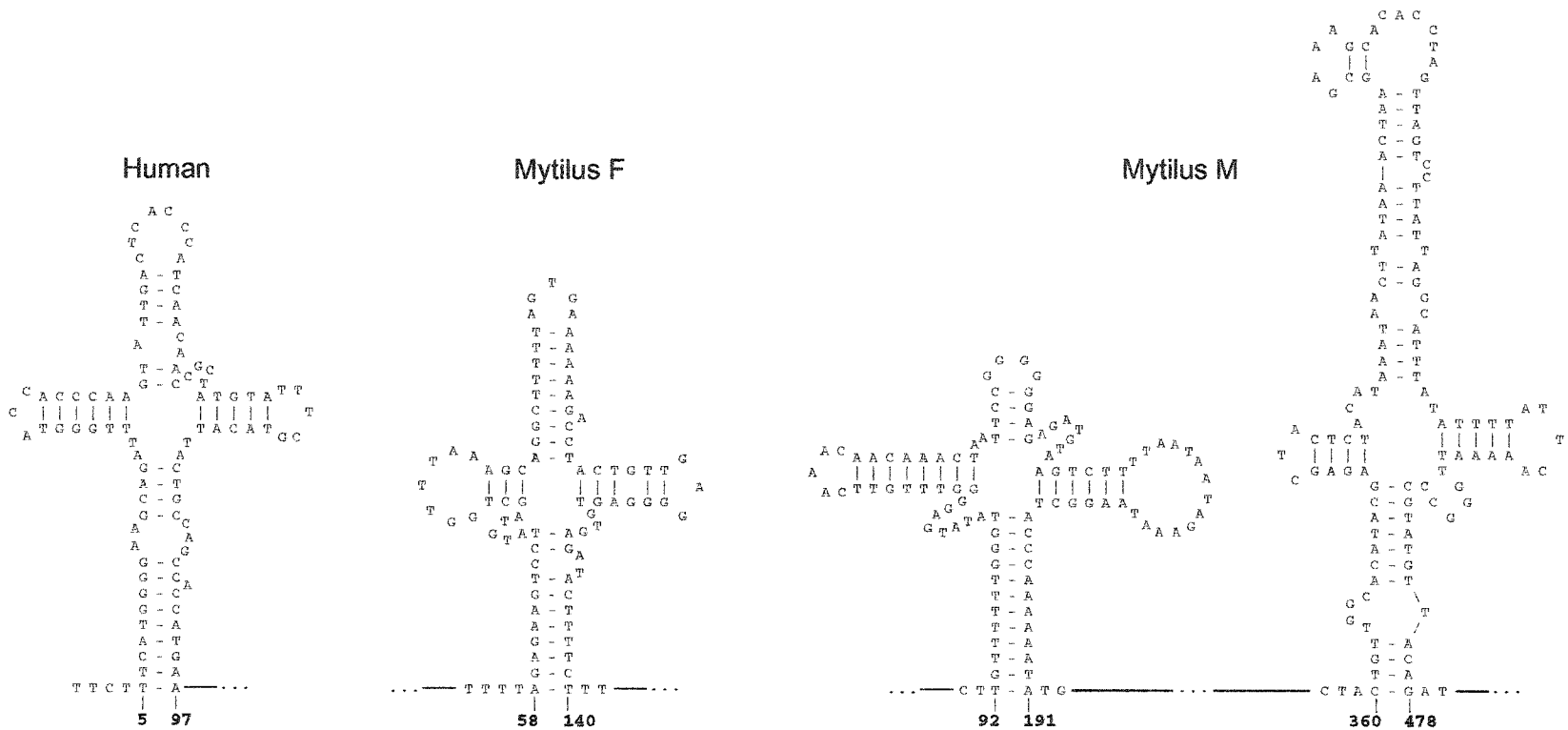


Figure 4.5. tRNA-like structures in the HVS1 of human mtDNA control region and the VD1 of the F and M LUR. The actual sequences shown are from ef.w22-F (F genome) and em.w143-M (M genome) in Appendix A.

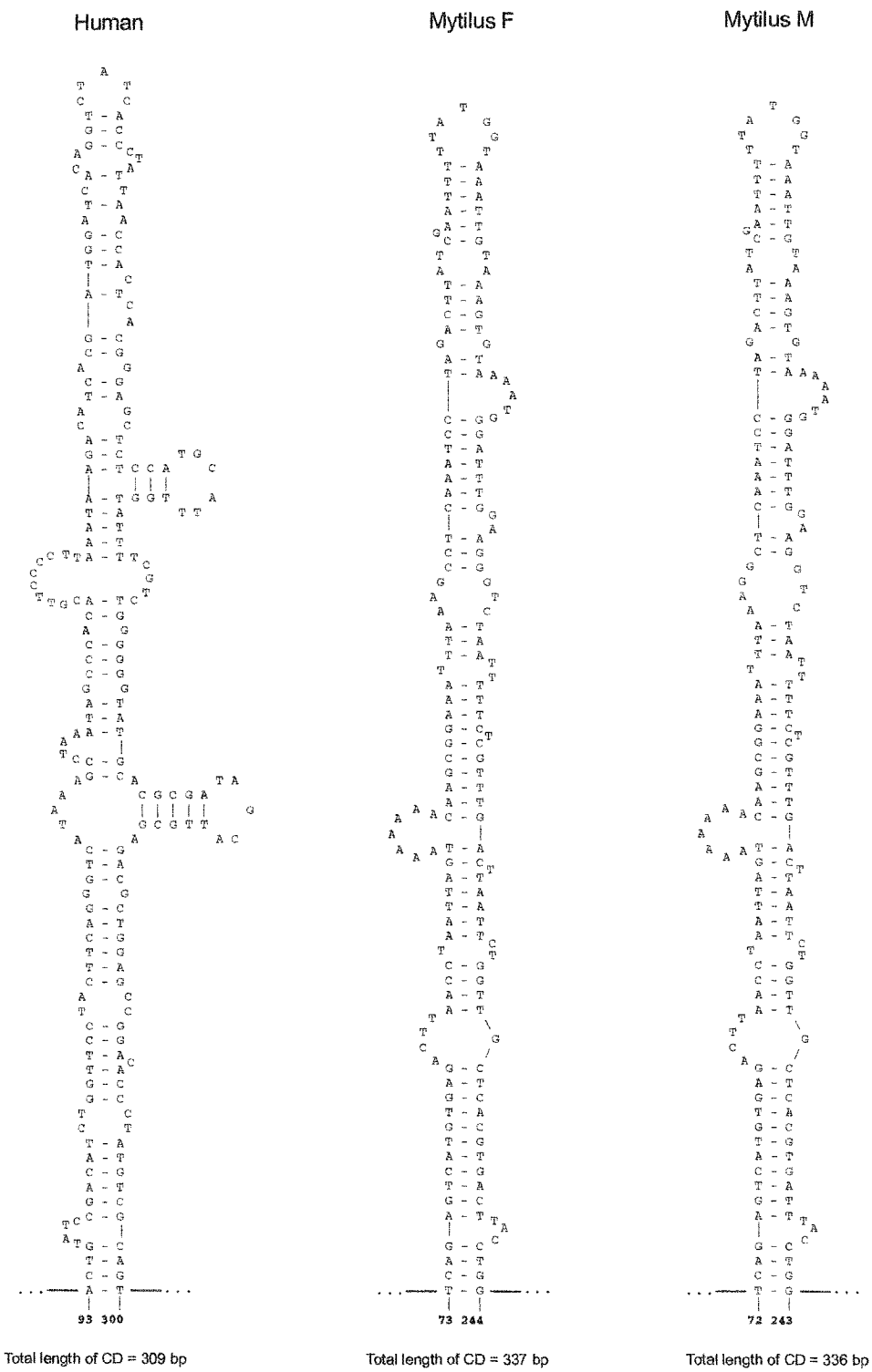


Figure 4.6. Stem-and-loop structure in the CD of the human control region and in the CD of the F and M LUR. The actual sequences shown are from ef.w22-F (F genome) and em.w143-M (M genome) in Appendix A.

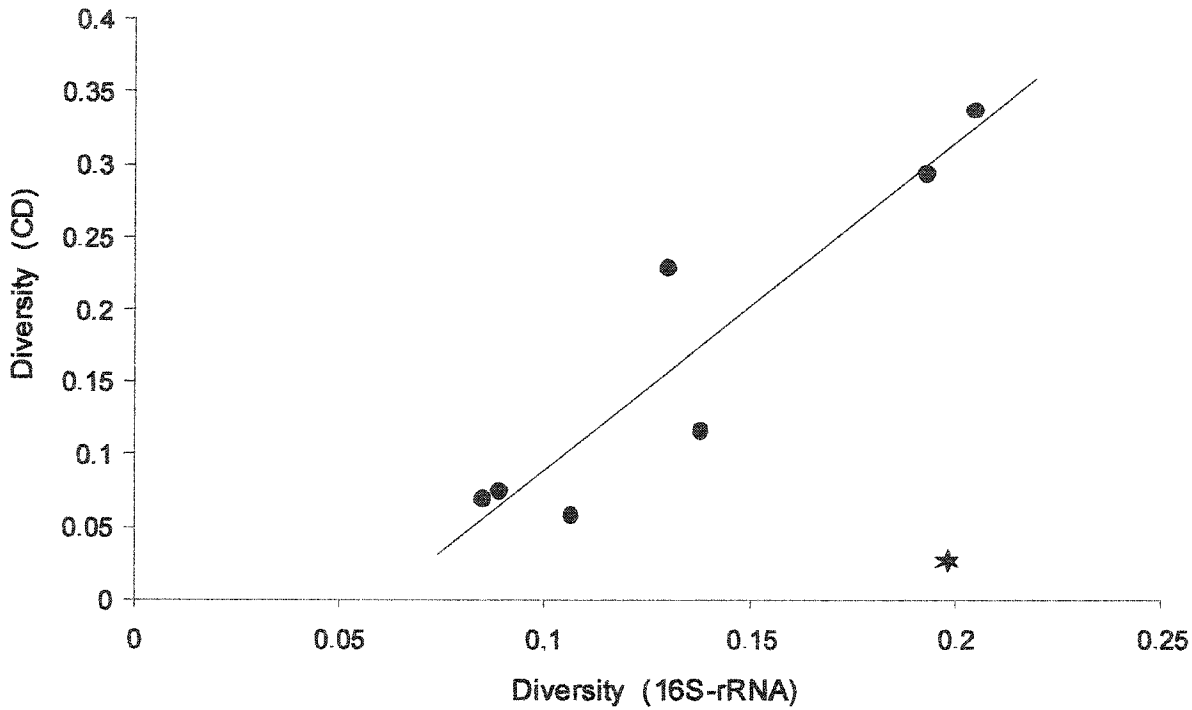


Figure 4.7. The divergence at the CD against the divergence at the 16S-rRNA of human mtDNA and seven other primates (data from Table 4.2). The star shows the F/M comparison.

Chapter 5

Characterization of Maternally and Paternally Transmitted Mitochondrial Genomes in Mussels of the Genus *Mytilus*

Introduction

Hoffmann et al. (1992) published the sequence of 13.9 kb of the 17.1 kb length of the female *M. edulis* mitochondrial genome. The most interesting feature of the genome is its gene order, which seems to have undergone a large number of re-arrangements compared to mtDNA of other metazoans. It contains the full complement of genes of the metazoan mitochondrial mtDNA and an extra tRNA for methionine, but lacks the *ATPase8* subunit gene. This raises the possibility that mussel mtDNA genomes may be drastically different from other animal mtDNA because of its unusual mitochondrial transmission system, and that F and M genomes each have distinguishing sequence features. To investigate this possibility, a comprehensive sequence comparison between F and M molecules is necessary. Previous studies between F and M sequences were limited to a small portion of protein-coding genes and the *16S* gene, and therefore could not provide a comprehensive assessment of the organization of the mussel mitochondrial genome.

In this study, I compare partial *12S*, complete *16S*, eight *tRNA*s and two protein coding genes between F and M molecules for three *Mytilus* species, *M.*

edulis, *M. trossulus* and *M. californianus*. This is the first comparative study between males and females of DUI species employing all the gene types.

Materials and methods

Sample collection and DNA extraction

Specimens of *M. edulis* and *M. trossulus* were collected from Lunenburg Bay, NS, Canada, and *M. californianus* from Neah Bay, WS, USA. As *M. edulis* and *M. trossulus* cannot be distinguished from each other morphologically, the identity of these two species was determined by restriction fragment analysis of the internal transcribed spacer (ITS) region of the nuclear ribosomal RNA genes as described in Heath et al. (1995). Animals were sexed by examining the gametes microscopically. Total DNA was extracted from gonad tissue using a modified salt extraction procedure (Miller et al., 1988).

PCR amplification, cloning and sequencing

Primer sequences and primer sets used in each species are given in Table 5.1, and the locations of all primers in the genes are shown in Figure 5.1. Total DNA from each individual served as the PCR template, and the full sequence length from the 3' end of the *12S* gene to the 5' end of *Cytb* was obtained from all of the animals. PCR amplifications were carried out in 25 µl reaction volumes containing 2 µl of template DNA, 0.8 mM of each primer, 1 mM dNTP, 2.5 mM $MgCl_2$ and 1 U *Taq* polymerase (MBI Fermentas) in the buffer supplied by the company. Details regarding PCR conditions for each reaction are

listed in Appendix B. PCR reagents, cloning and sequencing are referred to in Chapter 4. Primer walking was applied when cloned fragments were greater than 1.2 kb to ensure the sequence reading quality.

Data analysis

Multiple DNA sequences were aligned initially using Clustal X, version 1.8 (Thompson et al., 1997). Subsequent adjustments were made manually based on amino acid alignment for protein genes and secondary structures for RNA genes with the constraint that corresponding stem regions were aligned to one another while gaps were placed in other regions to maximize the degree of sequence similarity with minimum insertion/deletion events.

RNA secondary structure construction

The 16S and the third domain of 12S rRNA secondary structures were modeled from the proposed secondary structures of their *M. edulis* and *Drosophila* homologs, respectively (Cannone et al., 2002). Compensatory substitutions were used as evidence to evaluate the validity of stems. Domains and stems of 12S and 16S are named after the numbering system of Neefs et al. (1993) and Ban et al. (2000), respectively. Transfer RNAs were referred to their counterparts of *M. edulis* to form cloverleaf structures.

Base composition, codon usage and sequence diversity analysis

Eight tRNA genes in the same individual were concatenated together and treated as one single unit in the tests. For each gene, base pair compositions were examined and homogeneity of nucleotide frequencies was tested by χ^2 analyses among taxa. Stem and non-stem regions of RNA genes and each code positions of protein genes are also tested for homogeneity.

Twelve mitochondrial protein gene sequences (completed or partial) of *M. edulis* females were obtained from GenBank: M83756, M83757, M83758, M83759, M83760, M83761 and M83762 (Hoffmann et al., 1992). The percentage of each base was calculated on the third positions of four-fold degenerate codons for each gene, excluding the uncertain 5' or 3' ends, using Mega2 (Kumar et al., 2001). *M. edulis* female partial *Cytb* sequence from this study, *COI* sequence from Hoeh et al. (1996a) and *ND4* (accession number AF315165) were incorporated into the analyses.

The effective number of codons (Nc), GC content at the third position of synonymously degenerate codons (GC3s) and the relative synonymous codon usage (RSCU) were calculated using the program CodonW (written by John Peden, available from http://www.molbiol.ox.ac.uk/cu/). A χ^2 test was carried out to test the significance of codon usage differences between data sets from paternal and maternal mtDNAs. Nucleotide diversity between sequences was calculated using DnaSP 3.53 (Rozas and Rozas, 1999).

Results

Table 5.2 summarizes the lengths for each gene sequenced in this study except for the LUR region, which has been analyzed in Chapter 4. Three *Mytilus* species and the six mitotypes, share the same gene order in this segment (Figure 5.1). Males have fewer nucleotides in total compared to their female counterparts. Deletions in males occur mainly in rDNAs and spacers (Table 5.2).

Secondary structure features

Figure 5.2 shows the F type *M. edulis* 16S rRNA secondary structure model proposed by Cannone et al. (2002). My models of complete 16S rRNA and the third domain of 12S rRNA for the F *M. edulis* are presented in Figure 5.3 and 5.4, other mitotypes are given in Appendix C and D. The mitochondrial rRNAs studied here appear structurally conserved among *Mytilus* species and between genders, and in general conventional throughout most of the structural cores (see Lydeard et al. (2000) for 16S consensus diagram of mollusc and Hickson et al. (1996) for the third domain of animal 12S rRNA). However, some discrepancies exist. For 16S, 39 of 48 stems in Cannone's were supported by at least some compensatory evidence in *Mytilus* spp., nine were not. Of those, six stems (11, 13, 40, 86, 97, 100) were delimited from the model as no compensatory substitutions supported the stem formation in these regions among taxa or sex-specific groups, and three stems (2, 3, 41) were reconstructed at slightly different regions from the originals (Figure 5.3).

For the 12S third domain, two major insertions which are not in other animals were found. The first is the region between stem 38 sequences, which is

usually less than 60 bp and form stem 39, 40 and 42 in animal mtDNA. In *Mytilus*, this region varies from 60 bp to 152 bp, and no consensus structure can be identified among all species. The second lies between stem 45 and stem 47. An extra stem, named helix 46.1, can be determined for this region in all taxa.

The eight inferred tRNA secondary structures of *M. edulis* are given in Figure 5.5, those of *M. trossulus* and *M. californianus* are presented in Appendix E. They all share the cloverleaf structures similar to other animals (see Helm et al. (2000) for the consensus mitochondrial tRNA structures). Acceptor stem and anticodon domains are of typical animal mt tRNAs and are very conserved. D-stems usually have 4 bp (3 bp for a few tRNAs), and the D-loop has large variability in size in *Gly* and *Cys* among species, which is also observed in their counterparts of mammalian mt tRNA (Helm et al., 2000). T-stems formed of 5 bp, 4 bp or 3 bp and T-loops variable in size to a certain extent are found in mussels, which are the common features found in mammalian mt tRNAs too.

Base composition among taxa

Table 5.3 lists the results of χ^2 test for homogeneity of nucleotide base frequencies among *Mytilus* spp. males and females (six molecules) for each gene. These analyses demonstrated that for all genes base compositions did not differ significantly either for the gene as a whole or for specific regions (stem or non-stem in RNA genes) or positions (first, second or third code position in protein genes).

For rRNA and tRNA genes, base frequencies are more similar in stem regions than in non-stem regions across taxa. It is noteworthy that the mean GC content is higher in stems in all RNA genes. Further tests revealed that the mean base compositions in stem regions varied significantly from non-stem regions in *12S*, tRNA and *16S* genes (Table 5.3). For protein genes, while base compositions at the first and second codon positions are highly conserved among species and sex, third positions have large variation. The third codon positions are subject to less selective pressure as they consist of all four-fold sites which are under less selection constraint.

Compositional asymmetry in mtDNA genes

Heterogeneous base composition for genes in the same mtDNA genome has been reported in mammals. For the 12 H-strand (G rich) protein-coding genes, it appears that the compositional asymmetry correlates with the time the corresponding region of the paternal H strand remained in the single-stranded state during replication. This phenomenon was explained by an asymmetrical directional mutation pressure in mammal mtDNA genome (Reyes et al., 1998). The replication of the mammalian mitochondrial genome is asymmetric, with the H strand being synthesized earlier from one direction and the L strand (C rich) later from the opposite direction (Clayton, 1982). Therefore, during replication the paternal H strand stays single-stranded until paired with the newly synthesized L strand, whereas the paternal L strand is always paired with newly synthesized H strand. The H strand replication origin (O_H) is located in the main non-coding

region, and the L strand replication origin (O_L) is about 11 kb downstream of the O_H. Assuming the same replication rates for both H and L strands, the duration of the paternal H strand genome remaining single-strand state (D_{ssH}) can be calculated as described in Reyes et al. (1998):

$$D_{ssH} = \frac{2(O_L - \bar{x})}{L}$$

for the genes downstream of O_H and upstream of O_L, and

$$D_{ssH} = \frac{L - 2(\bar{x} - O_L)}{L}$$

for the genes downstream of O_L and upstream of O_H, where L is the total length of the genome, O_L is the distance of O_L from O_H, and $\bar{x}$ is the distance of the middle position of the gene from O_H.

Heterogeneity of base composition has been found among 12 genes of the *M. edulis* mtDNA genome ($\chi^2 = 53.67$, $P = 0.01$). To investigate whether base composition asymmetry also holds in *M. edulis*, replication origins in H and L strands have to be determined. The LUR in *Mytilus* has been identified as the putative control region (O_H) (Chapter 4), however, L strand replication origin (O_L) has not been determined. All mtDNA genes of *M. edulis* are located on the same strand (Hoffmann et al., 1992), the percentages of C and G are 14.2% and 23.6% in the 15 kb sequenced genome out of 17.1 kb. GC content of mtDNA is less than 50% in all metazoans (Saccone et al., 1999), therefore the strand holding all the genes is the H strand. A simple genome map of *M. edulis* including only 12 coding genes, 2 rRNA genes and 5 non-coding regions is given in Figure 5.6, and frequencies of nucleotide base composition of each gene in

Table 5.4. For testing the composition asymmetry, I first designated each coding gene a number corresponding to their position relative to the LUR in the direction of H-strand replication such as 1 for *Cytb*, 2 for *COII* and so on, and then plotted the percentages of A, C, G and T on four-fold positions (P_{4FD}) against their positions in the genome for each gene. It appears that the frequencies of four nucleotide base compositions in genes follow a pattern along the mtDNA genome (Figure 5.7). Specifically, percentages of A in genes increase with their distance from O_H first and then the value drops at one position and rises again afterwards. The same is true for G and C as well as T but in the opposite direction. What is interesting is that the turnover points for the four bases are all at position 8 except for A which is at position 7. Two non-coding regions lie between *ND3* (gene 7) and *COI* (gene 8) (Figure 5.6). The one of 119 bp is capable of forming secondary structure and significantly more stable than random sequences with the same nucleotide composition (Hoffmann et al., 1992). These properties make it a possible candidate of O_L. If it is the replication origin of light strand, D_{ssH} of each gene can be calculated using one of the two equations given above according to their positions in the genome. The duration order of each gene remaining single-stranded during replication will be: *ND3* < *ND2* < *COIII* < *ND6* < *ND4* < *ND1* < *ND5* < *ND4L* < *COII* < *ATP6* < *Cytb* < *COI*. Figure 5.8 shows that correlations between the base composition on P_{4FD} of genes and D_{ssH} were found for all four bases. A, C and G were negatively correlated, while T was positively correlated with the single-strand state.

Correlations in A and T were significant for G ($r = -0.81$, $P < 0.01$) and T ($r = 0.70$, $P < 0.05$).

Codon usage pattern

Nc is an estimator to measure the codon usage bias. For mussel mtDNA genes, it can range from 20 at which the bias is at a maximum and only one codon is used for its corresponding amino acid, to 62 which indicates no bias in codon usage and all codons are used for translation. Codon usage variation is influenced mainly by two factors: mutational bias (base compositional bias) and natural selection (optimize the efficiency/accuracy of translation). Under no selection, Nc is correlated to GC3s, and a plot of Nc vs. GC3s can be drawn as described in Wright (1990). Figure 5.9 displays the codon usage patterns of *COI* and *Cytb* genes for three *Mytilus* species and 12 genes for *M. edulis* F mtDNA. The curve shows the relationship between Nc and GC3s under no selection. All dots for *COI* and *Cytb* are close to the curve, suggesting that base composition itself is likely the main cause for the codon usage bias in *COI* and *Cytb* genes for *Mytilus*. As the base composition is homogenous among *Mytilus* species, it is expected that genes from F and M molecules will use the same codon at similar frequency. To test this hypothesis, sequences of *COI* and *Cytb* were pooled together for all F and M, respectively. A χ^2 test was performed for each codon between two data sets (Table 5.5). The results show that the frequency of certain codons used in F molecules is not different from the same codon in M molecules for all codons except codon AAG, which is used more often in F molecules.

Codon usage of 12 genes from *M. edulis* F mtDNA also appears to be constrained by the base composition bias (Figure 5.9). The 5 genes deviating from the curve the most have relative small sequences sizes, which may be responsible for the greater deviation. Figure 5.10 represents the relationship between frequencies of nucleotide bases at the third codon position in the gene and their D_{ssH}. It is noteworthy that plots of third codon position vs. D_{ssH} follow the same trend as of the four-fold site vs. D_{ssH}. Specifically, the T-ending codon appears to be used more frequently in the gene having larger D_{ssH}, while the opposite is found for the G- and A-ending codon. The C-ending codon seems to be used at a similar frequency in all genes. This further suggests that codon usage is mainly under the influence of mutational bias. Correlation between base frequency and D_{ssH} was found to be significant for the G-ending codon ($r = 0.75$, $P = 0.005$) and almost significant for the T-ending codon ($r = 0.55$, $P = 0.06$).

The GTG codon appears to be the start codon for the *Cytb* gene in the *M. trossulus* M molecule, while the ATG codon may be the start codon for its counterpart in *M. trossulus*, and for *M. edulis* and *M. californianus* F and M molecules. To exclude the possibility of PCR error, an extra clone of the *Cytb* gene for *M. trossulus* M and a PCR product amplified at a different time were sequenced. The same results were obtained.

Search for distinct domains between genders

The analyses carried out above suggested that there is no major difference between the F and M mtDNA molecules in terms of secondary

structures, base composition and codon usage for rRNA, tRNA and protein coding genes. However, it is still possible that the distinct sex-specific domains are contained in a limited portion of the genes. To examine this, all sequences from the same individual were concatenated, and a global alignment of all concatenated sequences was constructed. Nucleotide diversity was measured among all M molecules, all F molecules and between F and M in the same species. The estimates were made using a sliding window of 50 bp, moved by steps of 25 bp along the whole alignment (Figure 5.11). The plots show that the diversity among M molecules is always greater than among F over the whole sequence, and the diversity between F and M from the same species is in general smaller than among M and larger than among F groups in *M. edulis* and *M. trossulus*, while it is almost always the greatest in *M. californianus*.

Discussion

I have compared 3.8 kb of nucleotide sequence from F and M mtDNA molecules from three *Mytilus* species, covering rRNAs, tRNAs and protein coding genes, and also analyzed 15 kb out of 17.1 kb of the published *M. edulis* F molecule. This represents the first systematic analysis of sequence comparison for all types of mtDNA genes between genders and for all 12 mtDNA genes in the same genome in the species with DUI. This comparative study between genders allows several conclusions to be made.

First, M molecules appear to have a tendency for a smaller genome size. In *Mytilus* rRNAs, M molecules are constantly shorter than their F counterparts,

and this trend is also observed in spacer sequences (Table 5.2) and the LUR (Chapter 4 and unpublished data). The deletions in M rRNAs usually occur in structurally variable regions. Smaller mtDNA genomes may facilitate the paternal mtDNA to overcome their paucity at fertilization and take the dominancy in the male gonad. Replication advantage associated with smaller mtDNA has been proven in human mtDNA (Diaz et al., 2002). The existence of such a trend is also supported by reduced length of the M genome in a distant DUI species, *Venerupis philippinarum*, in which both F and M genomes have been sequenced completely (accession numbers AB065374 and AB065375).

Secondly, comparative analyses of 16S, 12S domain III and eight tRNA genes in *Mytilus* demonstrate that mussels basically share the same structure models of other animals, although some refinements can be made. There is no distinguishing structure between F and M molecules. Two large insertions, which are not present in other animal mtDNA, were found in 12S domain III in *Mytilus*. One forms an extra stem between stems 45 and 47. It is, however, unlikely to play a DUI related special role as this structure is absent in *V. philippinarum*. In addition, it is also found in a species of fungi, *Suillus sinuspaulianus* (Cannone et al., 2002). Another large insertion is located between the stem 38 two sequences. Due to its considerable size variation among species and the difficulty in identifying consensus structure in this region, this region seems more likely to be a variable region under relaxed functional constraint than a unique functional motif. Base compositions are homogeneous among all mitotypes for the above genes. Stem regions are significantly rich in G and C nucleotides,

which is consistent with the findings for other metazoans (Vawter and Brown, 1993). Higher GC content in paired regions is predicted to give a lower free energy value for the structures (Turner et al., 1988).

Thirdly, homogeneity of base composition in mussels also holds for protein coding genes. Codon usage appears affected mainly by the base bias of the gene. Out of 62 codons, 61 are used at a similar frequency in F and M molecules. Only one codon (AAG) seems to be used more in F molecules, but this exception could be caused by a small sampling size in the test. Based on the DNA sequences, GTG is predicted as the start codon of the *Cytb* gene for M *M. trossulus*, while ATG is predicted for F *M. trossulus* and other species. This is the first case showing that different start codons are used by a single species for a single gene. In *Drosophila,* two species in the same genus using different start codons for *COI* has been reported (Satta et al., 1987). Considering that the *M. trossulus* M molecule branched off before the speciation of *M. edulis* and *M. trossulus* (Stewart et al., 1995), this different start codon usage in *M. trossulus* F and M seems not an unusual exception.

Fourthly, there are no gender-specific motifs in *Mytilus* spp. sequences. Figure 5.11A shows that M molecules are more divergent than F over the whole alignment, which suggests that M is under more relaxing selection constraints than F. In addition, the two lines representing the diversities among each gender display a parallel pattern. This suggests that no distinct gender-specific regions exist in mussel sequences and functional constraint sites in M and F are at the same positions, with some regions under higher constraint and having lower

diversity values, and some regions under less selective pressure and having higher diversity values. This is further enforced by the plots in Figure 5.11B, C, D. If there is a male-specific conserved region, the diversity value among males will be lower than between F and M from the same species in this region, which is not the case seen in our study.

Sequence analyses of 88% *M. edulis* F mtDNA have revealed several new features in this genome. I suggest that the replication origin of L strand is located between *ND3* and *COI* genes based on two lines of evidence. One is that nucleotide composition bias patterns changed between these two genes for all four bases (Figure 5.7), which is consistent with the case for adjacent genes of O_L in mammalian mtDNA (Reyes et al., 1998). The other is that the 119 bp non-coding region found between *ND3* and *COI* can form secondary structures and have lower free energy than random sequence, making it a likely candidate for the O_L in mussels. Assuming this region is the O_L and the replication proceeds in the same fashion as in mammalian mtDNA, a correlation between nucleotide percentage and the duration of the gene remaining in the single-strand state (D_{ssH}) was detected for all four bases (Figure 5.8). This asymmetry is more pronounced on the less constrained four-fold sites than on all the sites (P_{123}), which have smaller variability in the base composition (Table 5.4), suggesting that an asymmetrical directional mutation pressure rather than selection is responsible for the base composition asymmetry in mussel mtDNA.

The mechanisms that lead to the asymmetrical base bias in mtDNA could be the spontaneous decomposition involved in transcription and replication

(Reyes et al., 1998). The H strand is exposed as a single strand for a considerable time during highly asymmetrical replication, and suffers decomposition caused by oxidation and hydrolysis. The longer the H strand stays single-stranded, the higher chance it has to experience these mutations. In mussels, all coding genes are located on the H strand, therefore it will undergo the same mutation pressure during transcription as in replication. Oxidation converts G into 8-hydroxyguanine, which pairs with A rather than C (Lindahl, 1993). Hydrolysis changes C into T, and A into hypoxanthine by deaminations, the latter preferentially paired with C rather than T (Lindahl, 1993). Figure 5.12 illustrates how these two forces lead to the base asymmetry pattern in *M. edulis* mtDNA. Deamination predicts a decrease of C and A frequencies on the H strand according to D_{ssH}, which is consistent with this data (Figure 5.8). It appears that the oxidation effect is much stronger than deamination in mussel mtDNA. Although deamination tends to increase G frequency, the end result is the declination of G caused by oxidation and the drastic increase of T resulted from both oxidation and deamination. The major cause of the compositional asymmetry in mtDNA is not the same for different species. In mussels the dominant force is oxidation, while in mammals it is deamination. It might explain the phenomena of lineage-specific GC asymmetry of mtDNA in metazoans (Saccone et al., 1999).

Compositional bias and translational selection are known as the two major factors to affect codon usage. Our analyses of codon effective number and regressions between the base frequencies on the third codon positions and D_{ssH}

imply the compositional bias caused by asymmetrical mutation pressure is the most important source to shape the mtDNA codon usage in *M. edulis*. Nc appeared correlated to GC3s, which indicated no external influence was imposed on Nc (Wright 1990). The variance in the data can be attributed to gene length. That regressions of the base composition on the third codon positions follow the same patterns as on P_{4FD} is also strong evidence to support this conclusion. The increase of T-ending codons according to D_{ssH} could be argued as the result of NNY amino acid accumulation, but it can not explain why T-ending codons were favored even more over their synonymous C-ending codons when D_{ssH} increased. To my knowledge, this is the first report to describe strand-specific asymmetrical directional mutation pressure and the codon usage shaped by this mutational bias in invertebrate mtDNA, although these observations have been reported in mammalian mtDNA, bacteria and chloroplast genomes (Reyes et al., 1998; McInerney, 1998; Saccone et al., 2002; Morton et al., 1999).

Based on these results, mussel mtDNA appears to be a regular mtDNA, although it is transmitted through a unique system. DUI is unlikely operating through mtDNA genes.

Table 5.1. Primer sequences (part A) and combination of primers for PCR amplification (part B).

A

Primer	Sequence[a]
F1	5' - ACCAGGATTAGATACCCTGT - 3'
F2	5' - CTAAACATTTGGTTGTCAAAGAG - 3'
F3	5' - AGCTACTCTAGGGATAACAGC - 3'
F4	5' - TTGCGACCTCGATGTTGGC - 3'
F5	5' - GAGAAAGTGGTGAATTTGACC - 3'
16SAR[b]	5' - CGCCTGTTTATCAAAAACAT - 3'
Ucytb144F[c]	5' - TGAGSNCARATGTCNTWYTG - 3'
R2	5' - AGGTCTTGCTTGTGAGCAA - 3'
R3	5' - ACTTTGCATTACAGACTGTC - 3'
R3b	5' - CCYARRGGRTTATTDCTHCC - 3'
R4	5' - AGCTCACCACCTATTCCTC - 3'
R5	5' - CATACATAGCGGAGCATACTC - 3'
16SBR[b]	5' - CCGGTCTGAACTCAGATCACGT - 3'
Ucytb270R[c]	5' - AANAGGAARTAYCAYTCNGGYTG - 3'
Ucytb272R[c]	5' - GARTGGTAYTTYCTNTTYGC - 3'

B

Species	Gender	mtDNA type	Code	Primer set
M. edulis	♀	F	ef	F1-16SBR, F4-Ucytb270R
M. edulis	♂	M	em	F1-16SBR, F4-Ucytb270R
M. trossulus	♀	F	tf	F1-16SBR, F4-R4, F5-R5, Ucytb144F-Ucytb272R
M. trossulus	♂	M	tm	F1-16SBR, F4-Ucytb270R
M. californianus	♀	F	cf	F1-16SBR, F3-R3, Ucytb144F-Ucytb272R
M. californianus	♂	M	cm	F2-R2, 16SAR-16SBR, F3-R3b, Ucytb144F-Ucytb272R

[a] D = A, G, or T; H = A, C, or T; N = C, T, A, or G; R = A or G; S = C or G; W = A or T; Y = C or T.
[b] Designed by Palumbi et al. (Palumbi et al., 1991).
[c] Designed by Merritt and Shi (Merritt and Shi, 1998).

Table 5.2. Sizes of the genes sequenced in the present study.

Species	Gender	mtDNA type	Code	*12SrRNA* (bp)	*16SrRNA* (bp)	*Cytb* (bp)	*tRNA* (bp)								Spacer (bp)	Total (bp)
							Asn	*Asp*	*Cys*	*Gln*	*Glu*	*Gly*	*Ile*	*Tyr*		
M. edulis	♀	F	ef	530	1244	849	65	65	68	67	65	66	67	67	22	3175
M. edulis	♂	M	em	509	1242	852	65	65	68	67	65	66	67	66	22	3154
M. trossulus	♀	F	tf	532	1243	849	65	65	68	67	65	66	67	67	22	3176
M. trossulus	♂	M	tm	481	1238	849	65	65	68	67	64	65	67	66	18	3113
M. californianus	♀	F	cf	529	1246	849	65	65	68	67	65	66	67	67	34	3188
M. californianus	♂	M	cm	417	1214	846	65	69	69	67	63	63	69	68	12	3022

Table 5.3. Base composition comparison among taxa for genes. For each gene mean sequence length and nucleotide base frequencies, χ2 value and the probability for bias in base composition among *Mytilus* spp. are listed. The same test results are also given for the stem and non-stem regions for rRNA and tRNA genes, and each code position for protein-coding genes. χ2 tests for homogeneity of mean base composition between stem and non-stem regions for RNA genes are also given (last column).

Region	Length[a] (bp)	A	C	G	T	G + C (%)	χ^2	P	χ^2/(P)
12S									
Stems	217	0.210	0.204	0.283	0.303	48.7	1.870	1.000	
Non-stem	282.7	0.429	0.098	0.186	0.287	28.4	6.640	0.967	194.126
All	499.7	0.334	0.144	0.229	0.294	37.2	2.430	1.000	(0.000)
tRNAs									
Stem	318	0.295	0.162	0.200	0.343	36.2	5.030	0.992	
Non-stem	211.8	0.391	0.109	0.164	0.336	27.3	7.610	0.938	41.502
All	529.8	0.333	0.141	0.185	0.340	32.6	4.240	0.997	(0.000)
16S									
Stems	562.3	0.277	0.170	0.214	0.340	38.4	3.000	1.000	
Non-stem	675.5	0.381	0.112	0.181	0.326	29.3	10.430	0.792	115.628
All	1237.8	0.334	0.138	0.196	0.333	33.4	9.930	0.824	(0.000)
Cytb									
First	283	0.292	0.178	0.254	0.277	43.2	1.350	1.000	
Second	283	0.250	0.173	0.190	0.387	36.3	0.540	1.000	
Third	283	0.220	0.157	0.206	0.417	36.3	20.340	0.159	
All	849	0.254	0.169	0.216	0.360	38.6	8.430	0.905	
COI									
First	206	0.296	0.168	0.307	0.230	47.4	1.260	1.000	
Second	206	0.182	0.202	0.202	0.413	40.5	0.380	1.000	
Third	206	0.326	0.134	0.180	0.361	31.3	18.570	0.234	
All	618	0.268	0.168	0.230	0.335	39.7	8.230	0.914	

[a] Average length of all species.

Table 5.4. Base composition of *M. edulis* F-type mtDNA genes.

				P_{123}[b]				P_{4FD}			
Gene	Position[a]	D_{ssH}	Length (bp)	A	C	G	T	A	C	G	T
Cytb	1	0.850	1044	23.6	17.4	22.9	36.1	20.0	11.9	25.6	42.5
COII	2	0.720	579	26.2	13.7	23.4	36.7	27.3	5.2	31.2	36.4
ND1	3	0.584	663	22.2	14.2	26.4	37.3	27.2	11.4	27.2	34.2
ND4	4	0.446	1188	24.9	15.7	25.1	34.3	34.0	13.2	31.6	21.2
COIII	5	0.306	792	22.5	16.7	25.5	35.4	31.8	9.1	28.0	31.1
ND2	6	0.184	651	28.6	12.8	27.2	31.4	35.5	13.6	32.7	18.2
ND3	7	0.057	348	20.7	13.2	28.4	37.6	24.6	12.3	36.8	26.3
COI	8	0.922	954	25.7	15.4	24.3	34.6	29.8	6.8	22.4	41.0
ATP6	9	0.797	714	22.8	14.1	24.4	38.7	26.4	12.4	21.5	39.7
ND4L	10	0.718	279	24.4	13.3	23.3	39.1	32.5	15.0	25.0	27.5
ND5	11	0.587	1566	26.8	14.9	23.0	35.4	33.6	9.7	24.6	32.1
ND6	12	0.444	474	25.2	9.5	26.2	39.1	20.5	11.0	27.4	41.1

[a] Gene position to LUR.
[b] All three code positions.

Table 5.5. Cumulative codon usage of *Mytilus* spp. F-type and M-type mtDNA from *COI* and *Cytb* genes. The number of codons and the relative symonymous codon usage (RSCU) are given.

		F-type		M-type				F-type		M-type	
AA	Codon	N	RSCU	N	RSCU	AA	Codon	N	RSCU	N	RSCU
Phe	UUU	75	1.52	84	1.58	Tyr	UAU	50	1.43	41	1.3
	UUC	24	0.48	22	0.42		UAC	20	0.57	22	0.7
Leu	UUA	70	2.21	67	2.15	TER	UAA	0	0	0	0
	UUG	37	1.17	32	1.03		UAG	0	0	0	0
	CUU	32	1.01	36	1.16	His	CAU	32	1.31	30	1.25
	CUC	9	0.28	6	0.19		CAC	17	0.69	18	0.75
	CUA	25	0.79	30	0.96	Gln	CAA	6	1.33	5	1.11
	CUG	17	0.54	16	0.51		CAG	3	0.67	4	0.89
Ile	AUU	59	1.39	60	1.46	Asn	AAU	56	1.32	44	1.09
	AUC	26	0.61	22	0.54		AAC	29	0.68	37	0.91
Met	AUA	54	1.37	54	1.23	Lys	AAA	26	1.18	41	1.74
	AUG	25	0.63	34	0.77		AAG*	18	0.82	6	0.26
Val	GUU	43	1.37	35	1.17	Asp	GAU	28	1.3	23	0.98
	GUC	4	0.13	10	0.33		GAC	15	0.7	24	1.02
	GUA	39	1.24	44	1.47	Glu	GAA	16	1.1	15	1.15
	GUG	40	1.27	31	1.03		GAG	13	0.9	11	0.85
Ser	UCU	25	1.98	28	2.09	Cys	UGU	14	1.4	14	1.4
	UCC	4	0.32	5	0.37		UGC	6	0.6	6	0.6
	UCA	9	0.71	5	0.37	Trp	UGA	23	1.02	20	0.89
	UCG	1	0.08	3	0.22		UGG	22	0.98	25	1.11
Pro	CCU	45	2.28	35	1.84	Arg	CGU	12	1.6	15	1.88
	CCC	8	0.41	15	0.79		CGC	3	0.4	5	0.63
	CCA	12	0.61	15	0.79		CGA	9	1.2	7	0.88
	CCG	14	0.71	11	0.58		CGG	6	0.8	5	0.63
Thr	ACU	37	2.03	33	1.91	Ser	AGU	13	1.03	17	1.27
	ACC	4	0.22	8	0.46		AGC	5	0.4	6	0.45
	ACA	21	1.15	16	0.93		AGA	17	1.35	22	1.64
	ACG	11	0.6	12	0.7		AGG	27	2.14	21	1.57
Ala	GCU	38	1.95	32	1.54	Gly	GGU	38	1.2	24	0.77
	GCC	7	0.36	18	0.87		GGC	11	0.35	14	0.45
	GCA	24	1.23	26	1.25		GGA	27	0.85	30	0.96
	GCG	9	0.46	7	0.34		GGG	51	1.61	57	1.82

* is used significantly more frequently in F mtDNA genes ($P<0.05$)

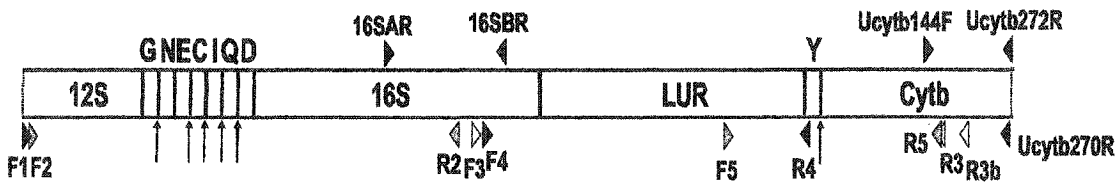


Figure 5.1. Gene map of sequenced segment of *Mytilus* spp. mtDNA. Horizontal arrowheads show the locations and directions of primers used for PCR amplifications. Vertical arrows below the map identifies the spacers. Locations of tRNAs are marked by their one-letter amino acid codes.

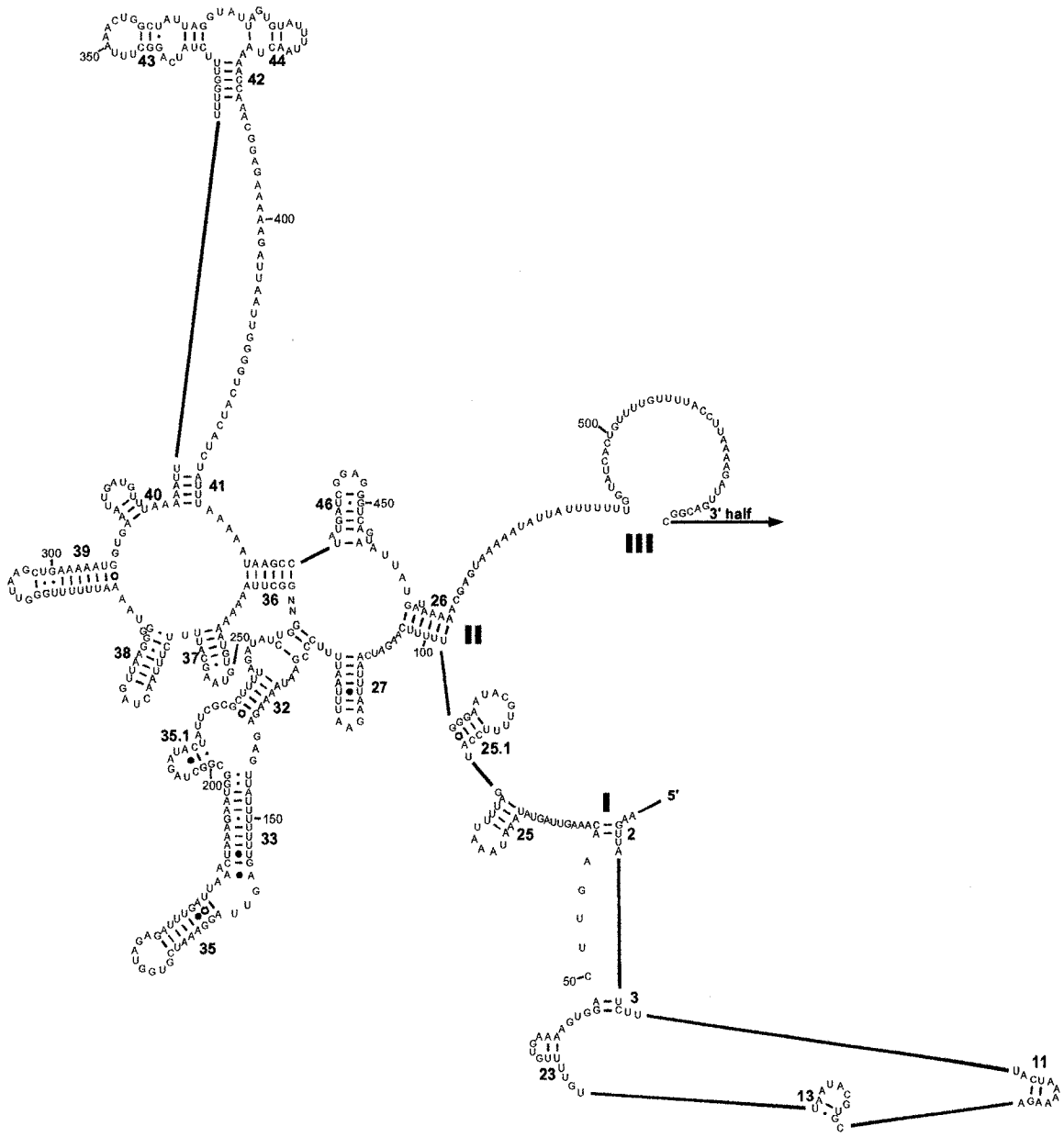


Figure 5.2A. Secondary structure model of *M. edulis* F-type 16S mitochondrial rRNA (5' half), from Cannone et al. (Unpublished). Bold numbers indicate stem numbers.

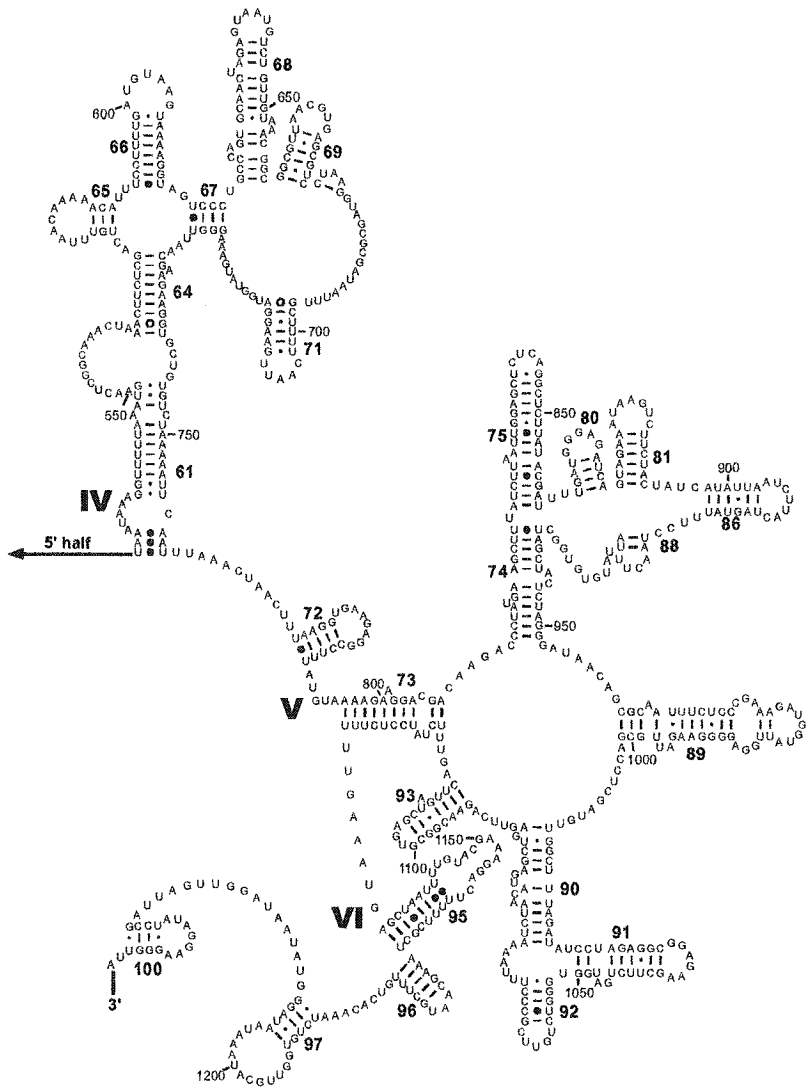


Figure 5.2B. Secondary structure model of *M. edulis* F-type 16S mitochondrial rRNA (3' half), from Cannone et al. (Unpublished). Bold numbers indicate stem numbers.

Figure 5.3. Potential secondary structure models of *M. edulis* F-type 16S mitochondrial rRNA from this study. (A) 5' half and (B) 3' half. Roman numerals denote the six domains. Bold numbers are stem numbers. Thin line boxes indicate regions variable among 16S rRNAs of *M. edulis* F-type (Mef), *M. edulis* M-type (Mem), *M. trossulus* F-type (Mtf), *M. trossulus* M-type (Mtm), *M. californianus* F-type (Mcf) and *M. californianus* M-type (Mcm). Number of nucleotides in this region is given for Mef and what has different number from Mef. Arrows attached with a small square identify the locations of one extra nucleotide in the indicated 16S rRNAs. Nucleotides with an arrow pointing out are absent in the indicated 16S rRNAs.

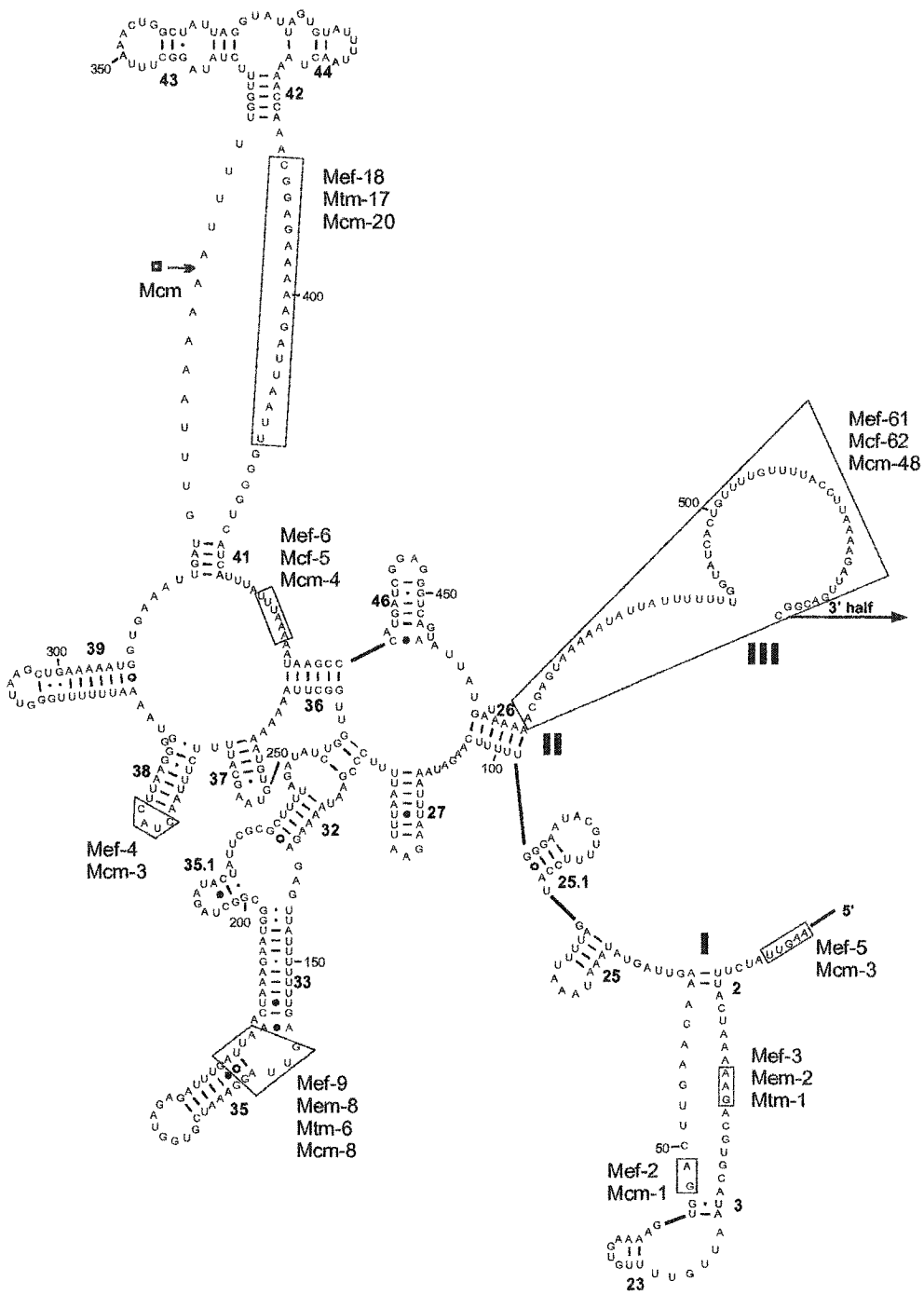

Mef-18
Mtm-17
Mcm-20
Mcm
Mef-6
Mcf-5
Mcm-4
Mef-61
Mcf-62
Mcm-48
3' half
Mef-4
Mcm-3
Mef-9
Mem-8
Mtm-6
Mcm-8
Mef-5
Mcm-3
Mef-3
Mem-2
Mtm-1
Mef-2
Mcm-1
5'
3
I
II
III

Figure 5.3A

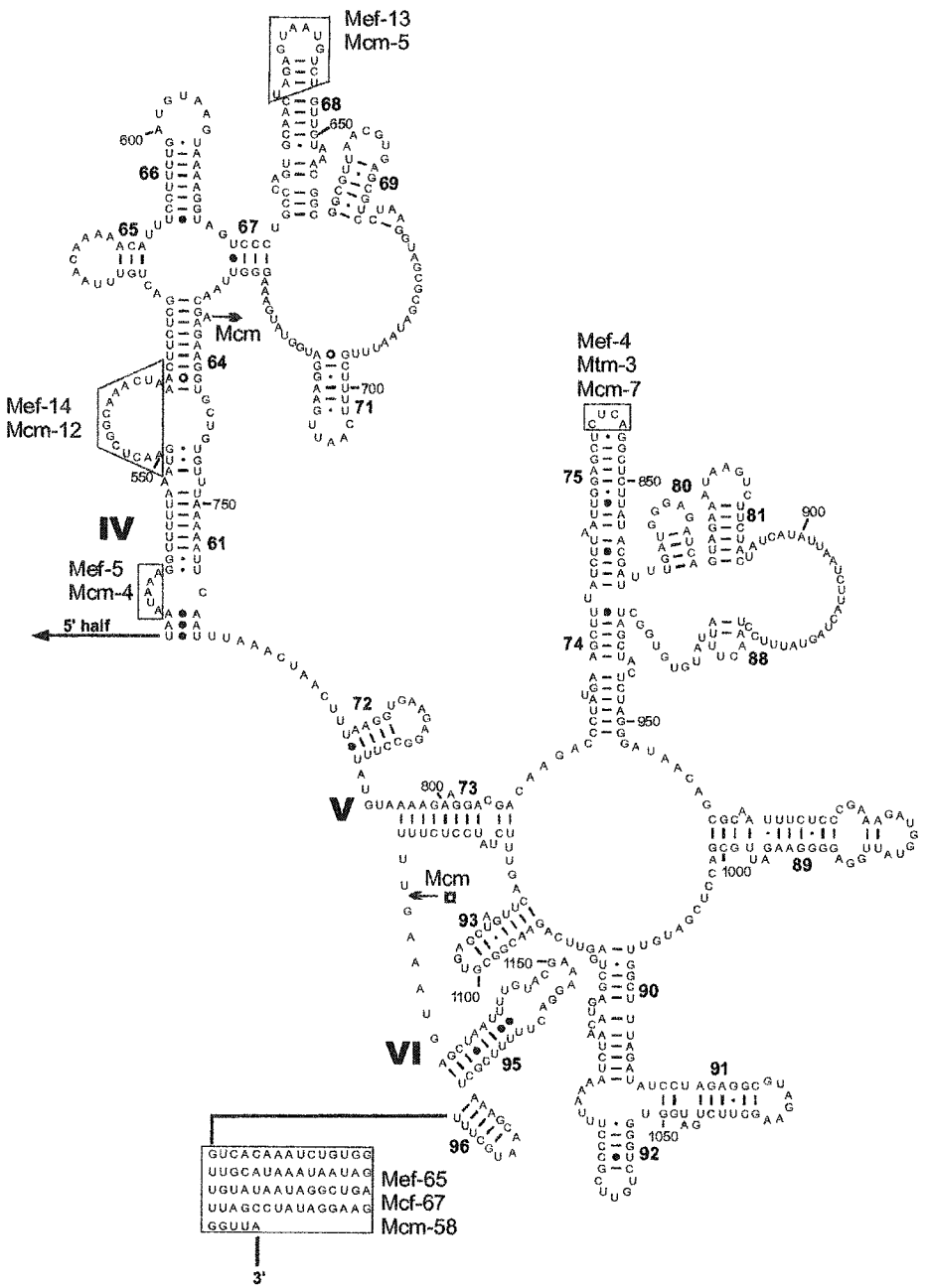

Mef-13
Mcm-5
Mcm
Mef-14
Mcm-12
IV
Mef-5
Mcm-4
5' half
Mef-4
Mtm-3
Mcm-7
V
Mcm
VI
GUCACAAAUCUGUGG
UUGCAUAAAUAAUAG
UGUAUAAUAGGCUGA
UUAGCCUAUAGGAAG
GGUUA
Mef-65
Mcf-67
Mcm-58
3'

Figure 5.3B

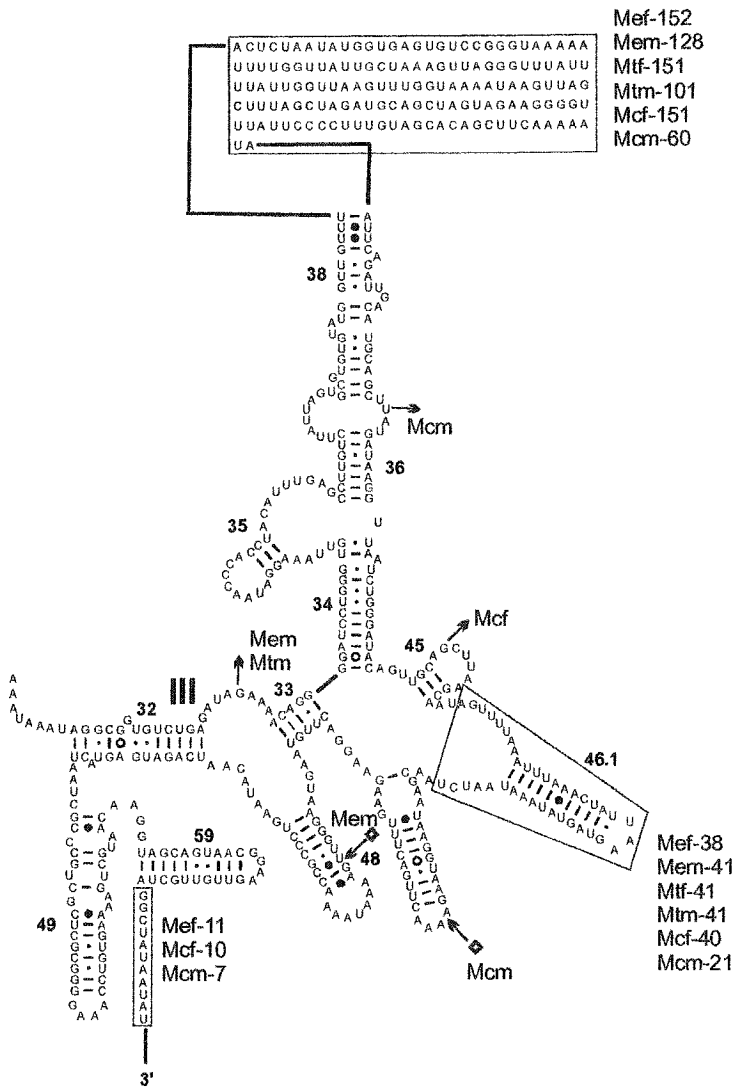


Figure 5.4. Potential secondary structure model of the third domain of *M. edulis* F-type 12S mitochondrial rRNA. See Figure 5.3 legend for labeling explanation.

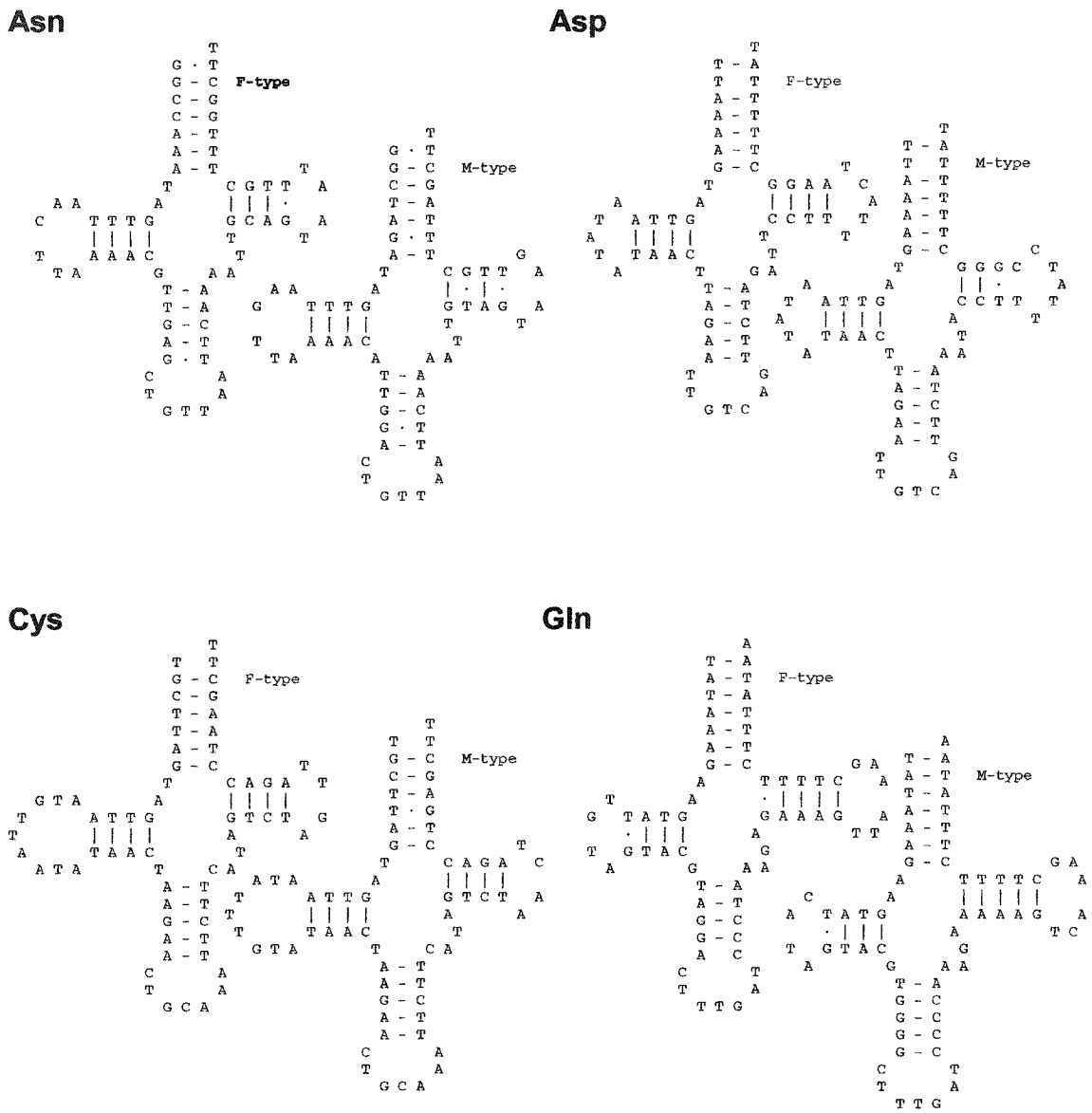


Figure 5.5. Secondary structures of 8 F-type and M-type tRNA genes of *M. edulis*.

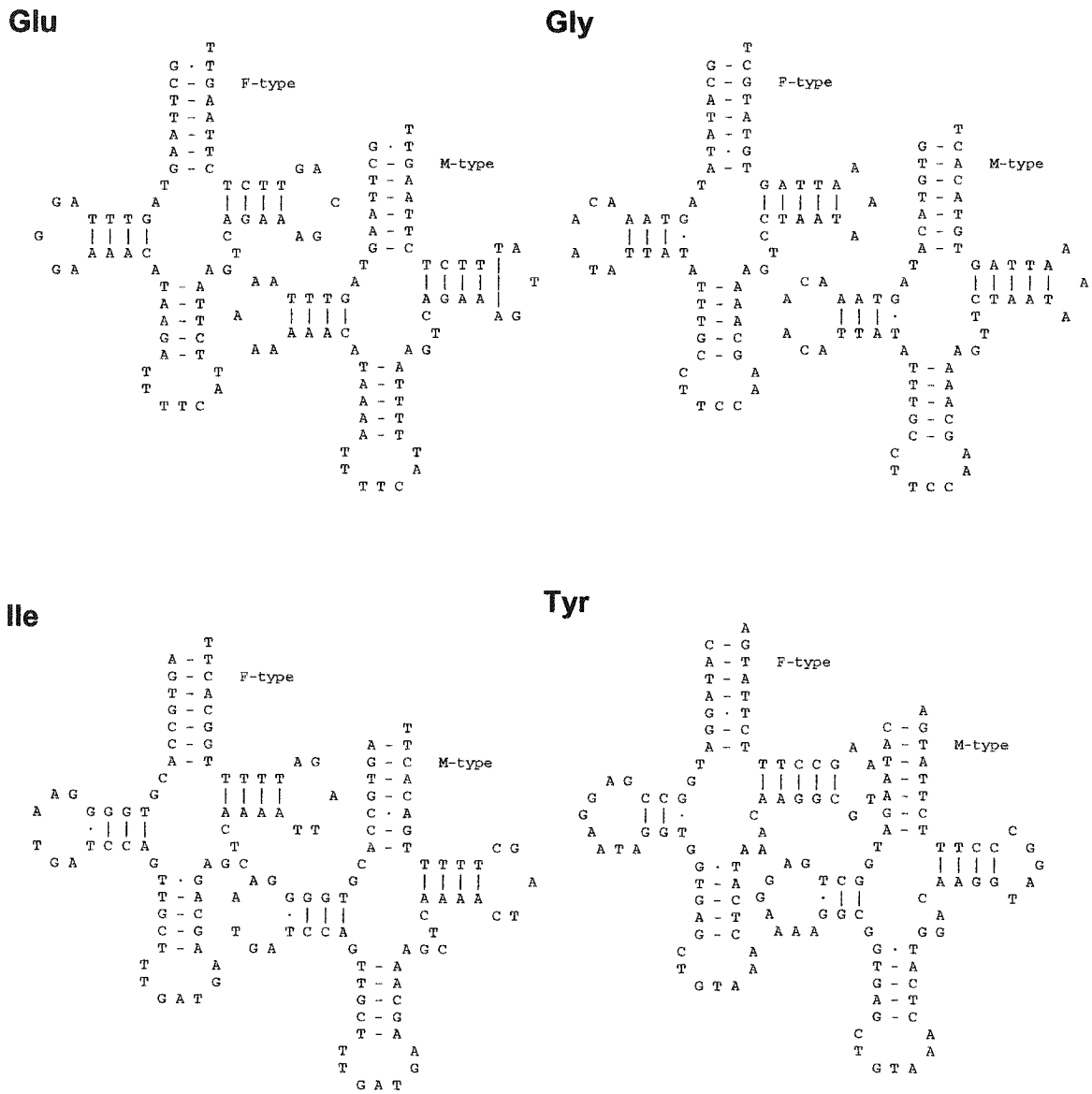


Figure 5.5. (Continued)

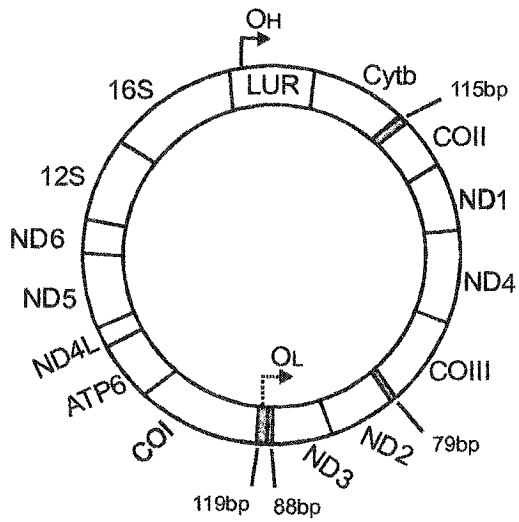


Figure 5.6. Simplified genome map of *M. edulis* F-type mtDNA. Arrows indicate the replication direction of the H and L strands. Unassigned regions are shown as shaded areas with their size numbers.

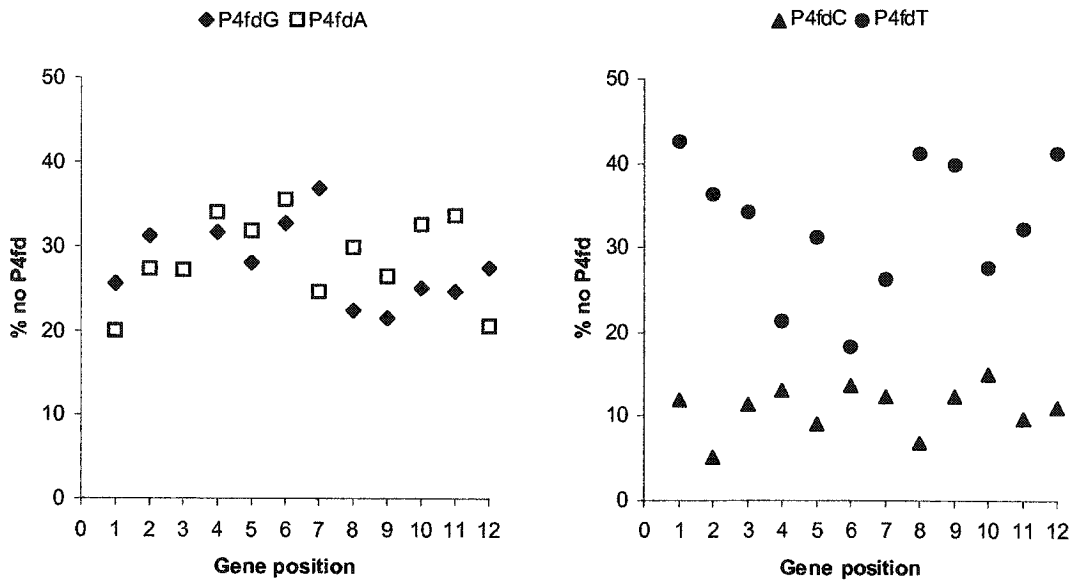


Figure 5.7. Plot of nucleotide frequency on the third position of four-fold degenerate codons vs. gene location away from the replication origin of H strand (*M. edulis* F genome).

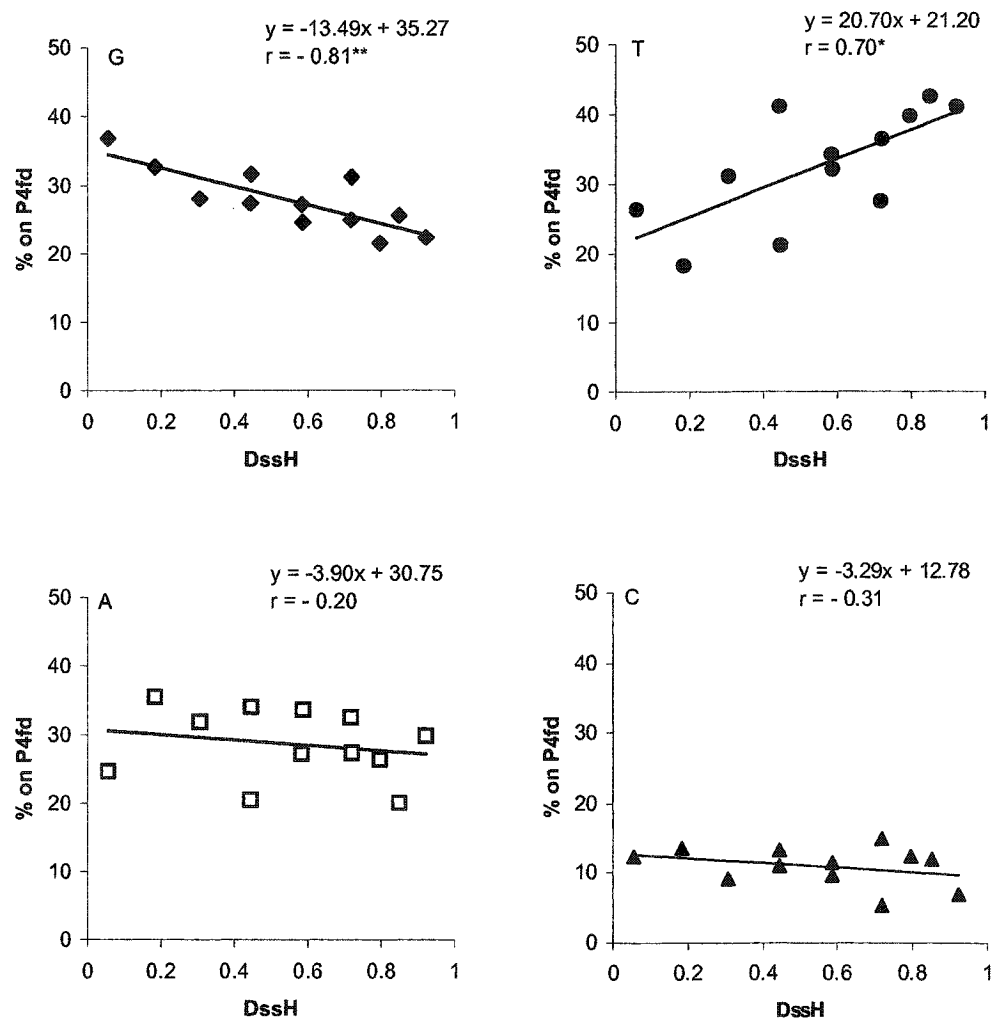


Figure 5.8. Correlation between the nucleotide frequency at the third position of four-fold degenerate codons and the duration of single-stranded state (DssH) for each gene in the *M. edulis* F genome (* $P < 0.05$; ** $P < 0.01$).

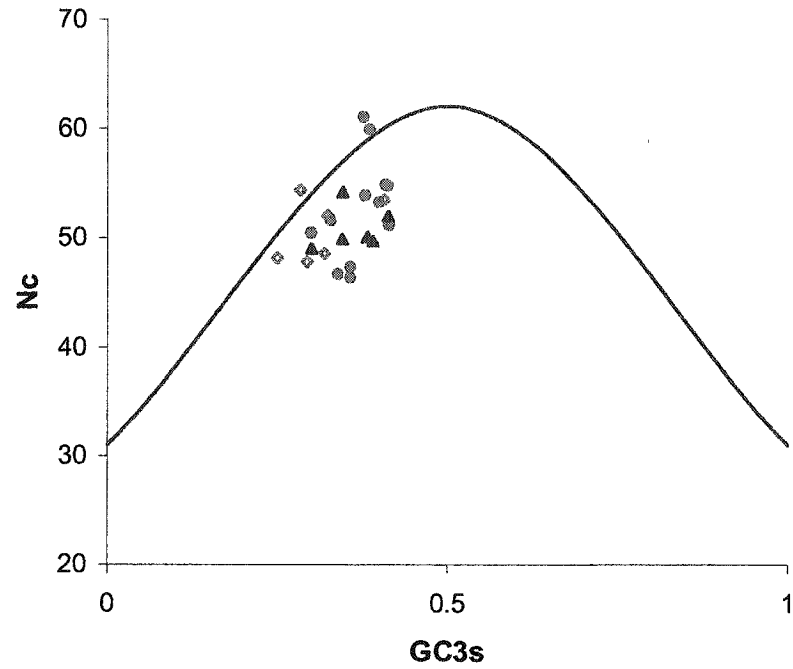


Figure 5.9. Plot of effective number of codons (Nc) against GC content of the third codon position. Solid circle represents the 12 genes of the *M. edulis* F-type genome. Triangles and open diamonds represent *Cytb* and *COI* genes, respectively, for six mitotypes from three *Mytilus* spp..

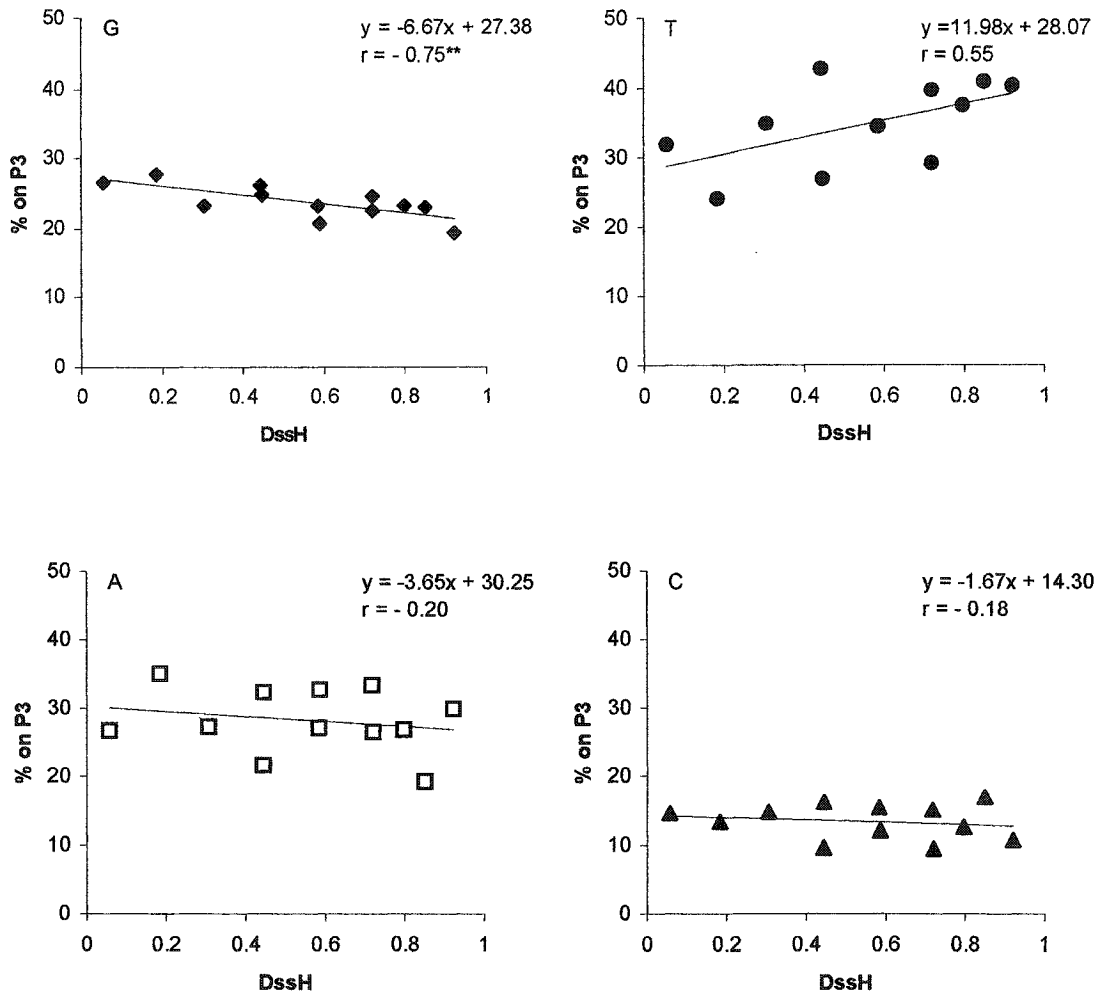


Figure 5.10. Correlation between the nucleotide frequency at the third codon position and the duration of single-stranded state (DssH) for each gene in the *M. edulis* F genome (** $P < 0.01$).

Figure 5.11. Nucleotide diversity among the sequences of the same gender or between the pairs of sequences from the same species. Slide window is of 50 bp. Symbols (dots or triangles) correspond to the midpoint of the slide window. Blue line stands for the nucleotide diversity among three *Mytilus* M sequences, pink for among three F sequences, green for between F and M of *M. edulis*, violet for between F and M of *M. trossulus* and dark red for between F and M of *M. californianus.*

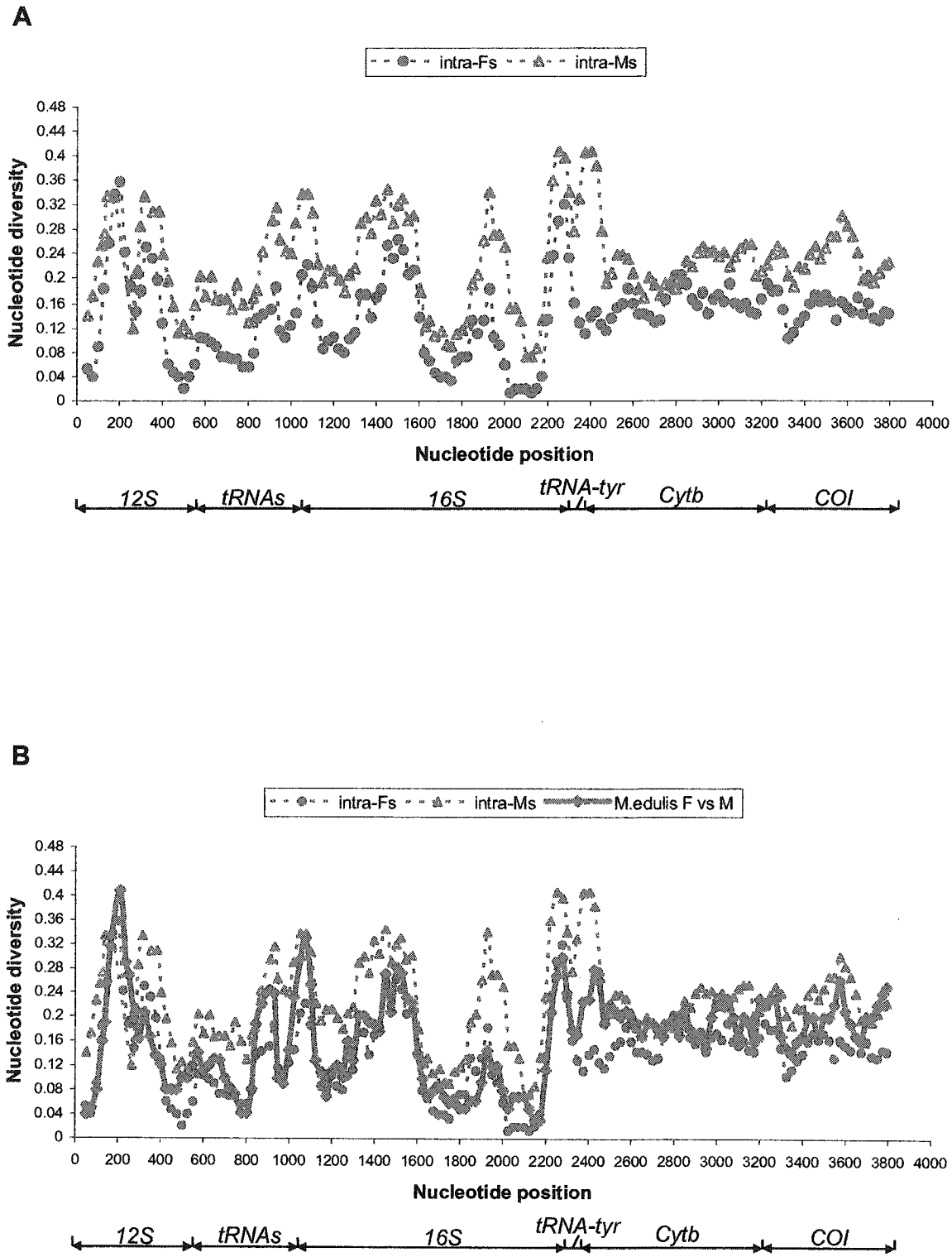
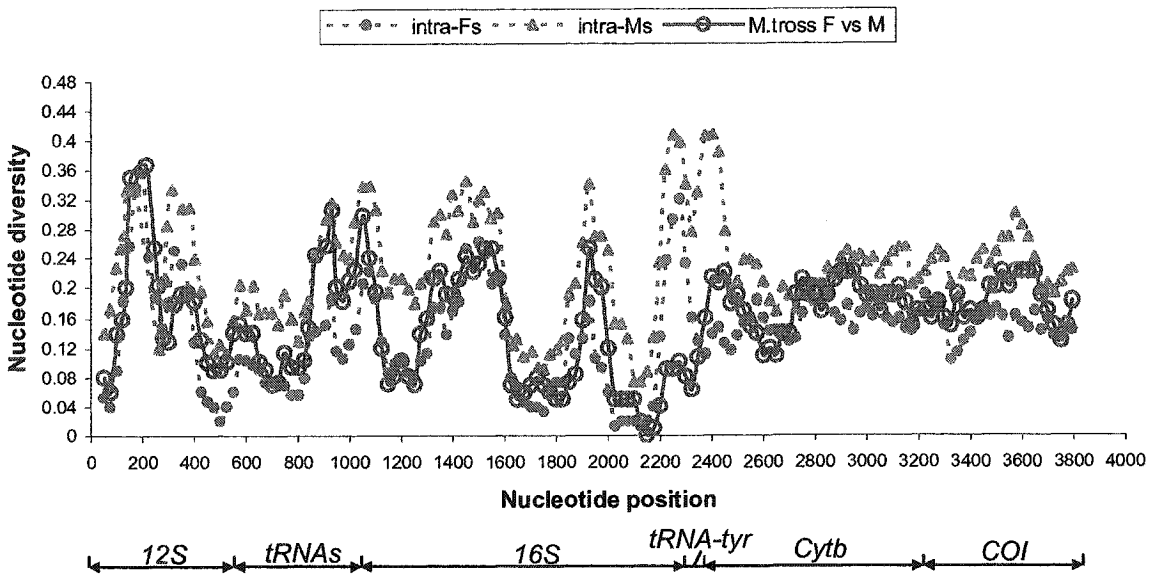

A
intra-Fs
intra-Ms
Nucleotide diversity
0.48
0.44
0.4
0.36
0.32
0.28
0.24
0.2
0.16
0.12
0.08
0.04
0
0 200 400 600 800 1000 1200 1400 1600 1800 2000 2200 2400 2600 2800 3000 3200 3400 3600 3800 4000
Nucleotide position
12S
tRNAs
16S
tRNA-tyr
Cytb
COI
B
intra-Fs
intra-Ms
M.edulis F vs M
Nucleotide diversity
0.48
0.44
0.4
0.36
0.32
0.28
0.24
0.2
0.16
0.12
0.08
0.04
0
0 200 400 600 800 1000 1200 1400 1600 1800 2000 2200 2400 2600 2800 3000 3200 3400 3600 3800 4000
Nucleotide position
12S
tRNAs
16S
tRNA-tyr
Cytb
COI

Figure 5.11

C



D

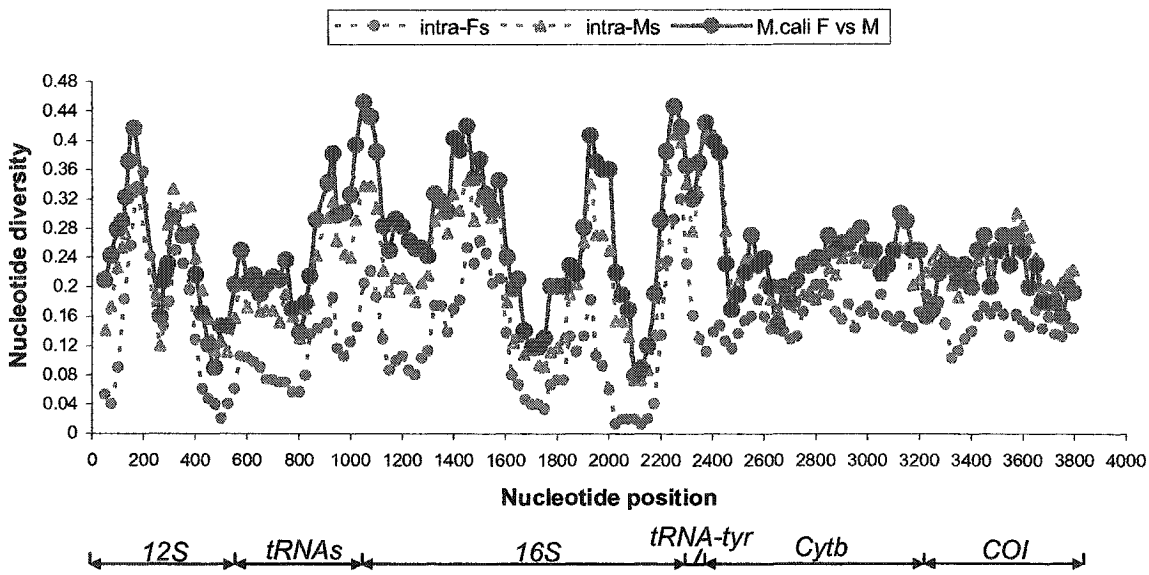


Figure 5.11. (Continued)

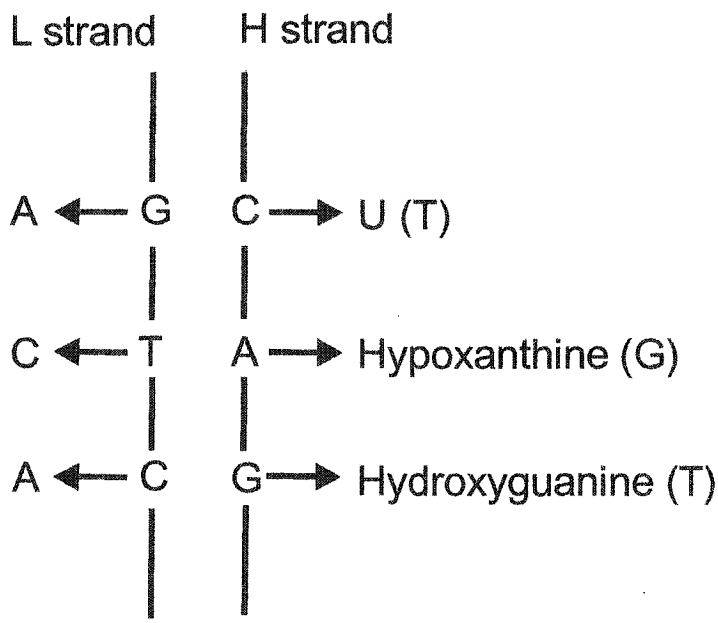


Figure 5.12. Diagram illustrates the proposed deamination and oxidation processes in mussel mtDNA. Nucleotides in brackets will increase in the H strand through these processes.

Chapter 6

Conclusions and Future Directions

In this study, I have employed two different approaches to investigate the mechanisms involved with the DUI system in mussels. The first approach is using genetic and cytological methods to examine behaviors of sperm mitochondria in developing embryos and sex ratios of crosses. Both projects involve selected pedigreed animals, known to produce different proportions of male and female offspring, hence a correlation between the observations and gender can be made. The second approach is using molecular techniques to analyze and compare the paternal and maternal mtDNA genome sequence features of DUI species and identify, if any, gender-specific features. The conclusions are summarized as follows.

1. In the mussel *M. edulis* paternal mitochondria enter all eggs at fertilization, regardless of the sex of the embryo. There is neither sperm mitochondrial destruction nor a replication advantage for paternal mtDNA occurring in early embryogenesis. Soon after fertilization, sperm mitochondria start to exhibit two distinct distribution patterns among zygotes. In embryos from females that produce only daughters, sperm mitochondria are randomly dispersed among blastomeres. In most embryos from females that produce a high proportion of sons, sperm mitochondria tend to aggregate and end up in one blastomere at the 2- and 4-cell stages. In the 2-cell embryos this blastomere is

known to be the progenitor of the primordial germ cells. Both patterns persist to late trochophore stage, at least in the few embryos where observations could be made late in development. Given the rare detection of tiny amounts of paternal mtDNA in various tissues of adult female mussels and the common presence of paternal mtDNA almost exclusively in the male gonad (Garrido-Ramos et al., 1998), the observations above suggest a mechanism of mtDNA transmission in mussels. The hypothesis is that in female mussels paternal mitochondria behave like maternal mitochondria in terms of random distribution and replication, hence paternal mtDNA is lost in most but retained in few females at very low amounts by random drift. In male mussels paternal mitochondria are delivered to germ cells through a regimental pathway, where paternal mtDNA takes the dominance probably through preferential replication of paternal mtDNA and/or eliminating of maternal mtDNA. This also leaves the soma of male mussels to be dominated by the maternal genome.

This is the first evidence suggesting that DUI may be achieved through differential segregation of sperm mitochondria in different genders as an initial step. The hypothesis also explains the phenomenon of DUI leakage observed occasionally in mussels. Because the different segregation patterns appear very early after fertilization, it implies that which pattern the sperm mitochondria will follow in a particular egg seems to be determined by the sex to which the egg will develop. As the sex of mussels is under the control of their female parents (see below for details), the maintenance of DUI is likely dependent on mothers.

2. Sex ratio of crosses in *M. edulis* varies extremely from 0 sons to more than 90% sons, and it is a characteristic of the female regardless of the male parent. This observation is consistent with the findings previously reported in *M. galloprovincialis*, and suggests that female-dependent sex ratio bias may be the property of all *Mytilus* species.

The bias is about the same among crosses sharing the same female parent but made at different years or different time in the season, hence it makes environmental factors unlikely to be the major influence on sex ratio bias. The novel information from this study is that the sonless/son-bearing trait is transmitted maternally for four consecutive generations, which implies that this property of females is heritable. Given that sonless mothers produce both sonless and son-bearing daughters and male parents have no impact on the sex ratio of progeny, the probability that this trait may be controlled by mother's mtDNA or father's genetic material or any symbiont transmitted through parents is ruled out. This leaves the mother's nuclear genotype as the most likely alternative responsible for the control of sex ratio in mussels. A one-locus two-allele model appears to explain the sex ratio bias in mussels well, in which the sex of the offspring is proposed to be determined by the nuclear genotype of the mother rather than the offspring itself. A homozygous female parent produces either all sons or all daughters depending on which allele she possess, whereas a heterozygous female parent produces both sexes in a ratio depending on the degree of dominance of the two alleles. Empirical data fit this model reasonably well. Although the male parent has no direct influence on the sex of its offspring,

it appears to have impact on the sex of its grandchildren through contributing its nuclear genotype to its daughters.

3. The large unassigned region (LUR) of F and M mtDNA from *M. edulis* and *M. galloprovincialis* can be divided into three distinct domains (VD1, CD and VD2) in terms of sequence feature and conservation. VD1 is highly divergent between F and M genomes and contains a number of deletions mainly in M molecules, while the CD region is extremely conserved and lacking any deletions. VD2 is intermediate on these two aspects but rich in purines itself. Comparison of the LUR with the control region of human mtDNA shows that the LUR consists of similar motifs and secondary structures present in human mtDNA at corresponding locations. These suggest that the LUR is very likely the control region of mtDNA in mussels. In addition, two other interesting findings may have important implications. One is that M molecules may hold extra structure and elements in VD1. Given the high sequence divergence displayed in this region, VD1 may play a sex-specific role in mtDNA transmission. The other is that CD is extremely conserved despite relaxed selective constraint on mussel mtDNA genes. This implies that CD may be assigned to a crucial role such as interacting with nuclear gene coded products, hence under a strong co-evolutionary constraint between nuclear and mtDNA sequences.

4. Sequence comparisons of a 3.8 kb fragment including various genes from three *Mytilus* species suggest that the M genome is likely smaller than their counterpart F genome in species with DUI. The M molecule has a faster evolution rate over the whole region examined than the F molecule. There is no

sex-specific feature at the molecular level between F and M in terms of rRNA and tRNA secondary structures, base composition, codon usage and local sequence motifs. Based on the nucleotide sequence, GTG is used as the start codon of the *Cytb* gene for *M. trossulus* M molecule, while ATG is for *M. trossulus* F molecule and others. The 119bp non-coding region between the *ND3* and *COI* genes is identified as the possible O_L. For the genome as a whole, *M. edulis* F- type mtDNA experiences an asymmetrical directional mutation pressure involving in transcription and/or replication processes, which affects its nucleotide base composition and codon usage. These are the properties observed in other metazoans, some bacteria and chloroplasts. Secondary structures of rRNAs and tRNAs from *Mytilus* mtDNA are structurally compatible with their closely related species. Therefore, mussel mtDNA genomes resemble typical animal mtDNAs. In addition, the overall sequence feature similarity shared by F and M mtDNA explains the feasibility of F and M genome recombination and the invasion of F into M route as these events will not cause any disruptions in the gene structures or incompatibility for the codon usage.

Future directions

The results presented in this study shed an important new light on understanding the properties of DUI, meanwhile they also raise some interesting topics that deserve more attention. First, microscopic observation suggests that sperm mitochondria may be directed into the primordial germ cells in male mussels. Given the extreme bias against paternal mtDNA in the embryo's mtDNA

pool, the M genome can only gain dominance in germ cells through either active destruction of maternal mtDNA or preferential replication of paternal mtDNA. The analysis of the F and M mtDNA control region (LUR) supports the latter as motifs probably playing a sex-specific role have been identified in M molecules. To prove this hypothesis, the newly masculinized F molecule is a good candidate for examination. The result could uncover how the F molecule transforms into M. I have obtained desired sequences from such individuals. Preliminary analysis shows that a large piece of M LUR has been incorporated into F LUR and consists of a larger mosaic control region. The significance of this will be further investigated.

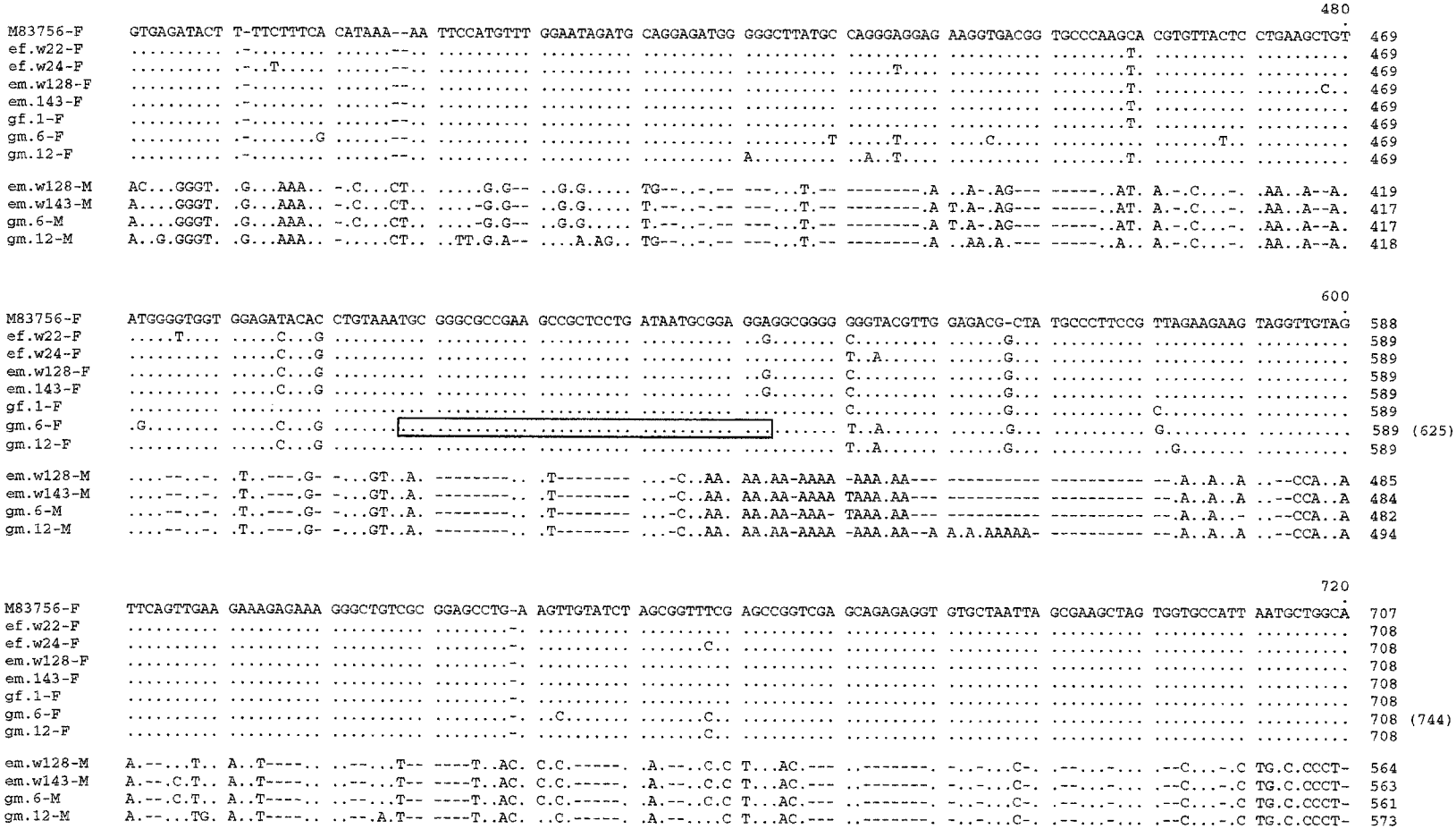
Second, in *M. galloprovincialis* all triploids were identified as males, whereas diploids had a 1:1 sex ratio (Kiyomoto et al., 1996). The authors proposed that a Z:W or Z:O sex determination model may operate in this species. However, their results were not obtained from selected crosses, and the mortalities in triploid and diploid groups were not estimated either. Hence the possibility that the male bias in their triploid mussels was caused by high mortality in female triploid mussels exists. This uncertainty can be solved by using pedigreed female mussel parents, which are known to produce all daughters or mostly sons. It would be of interest to examine whether sonless mothers will produce triploid sons; if they do, whether their son's gonads are dominanted by the M genome. The same can be done for the son-bearing mothers. Meanwhile, microscopic observation of sperm mitochondrial distribution in triploid embryos is needed. This package of data will elaborate our

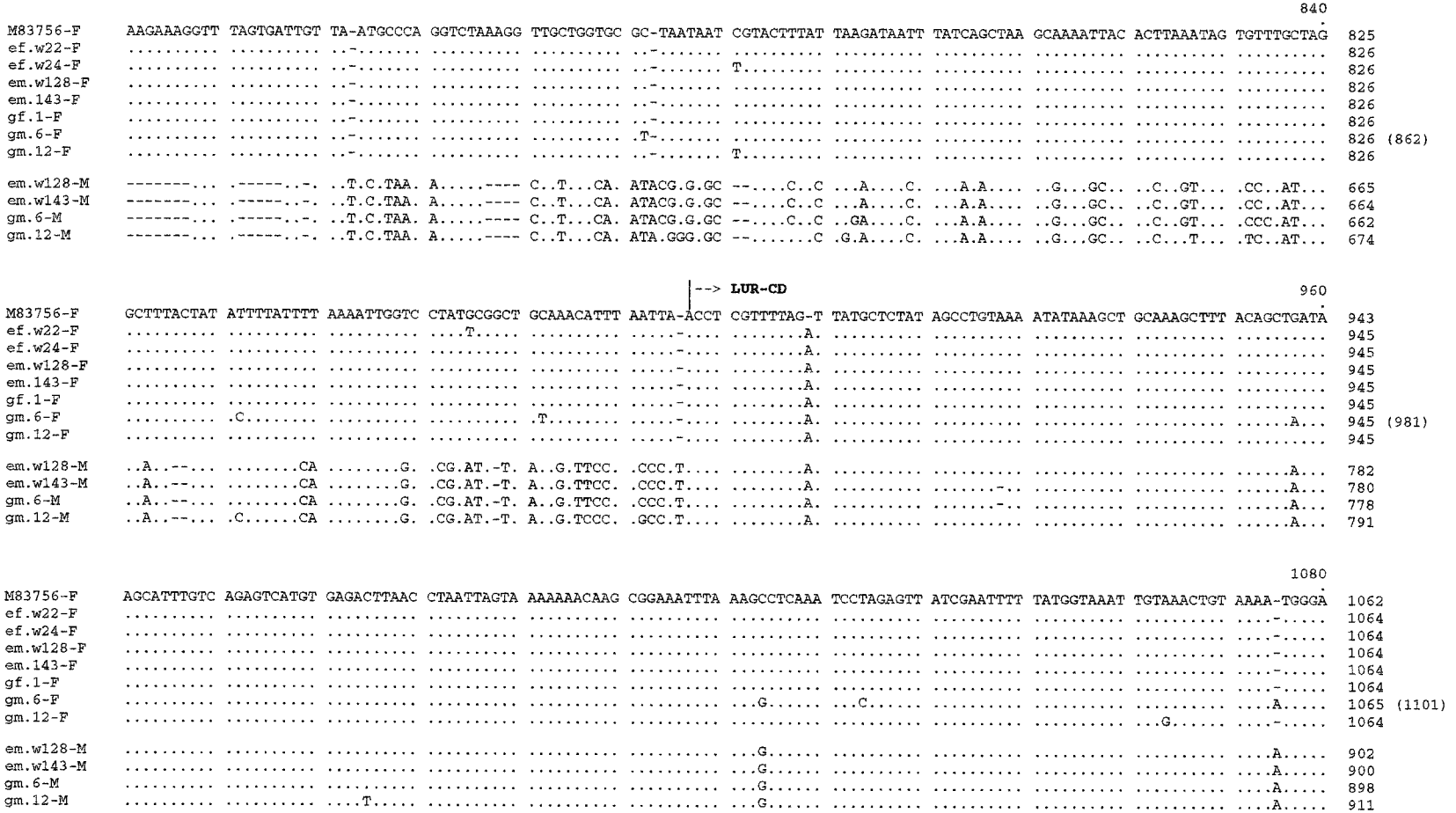
understanding for sex determination in mussels. I have had some preliminary data in terms of microscopic observation on triploid embryos and the sex ratio of mature triploid mussels. Analysis will be undertaken and further experiments may be designed accordingly.

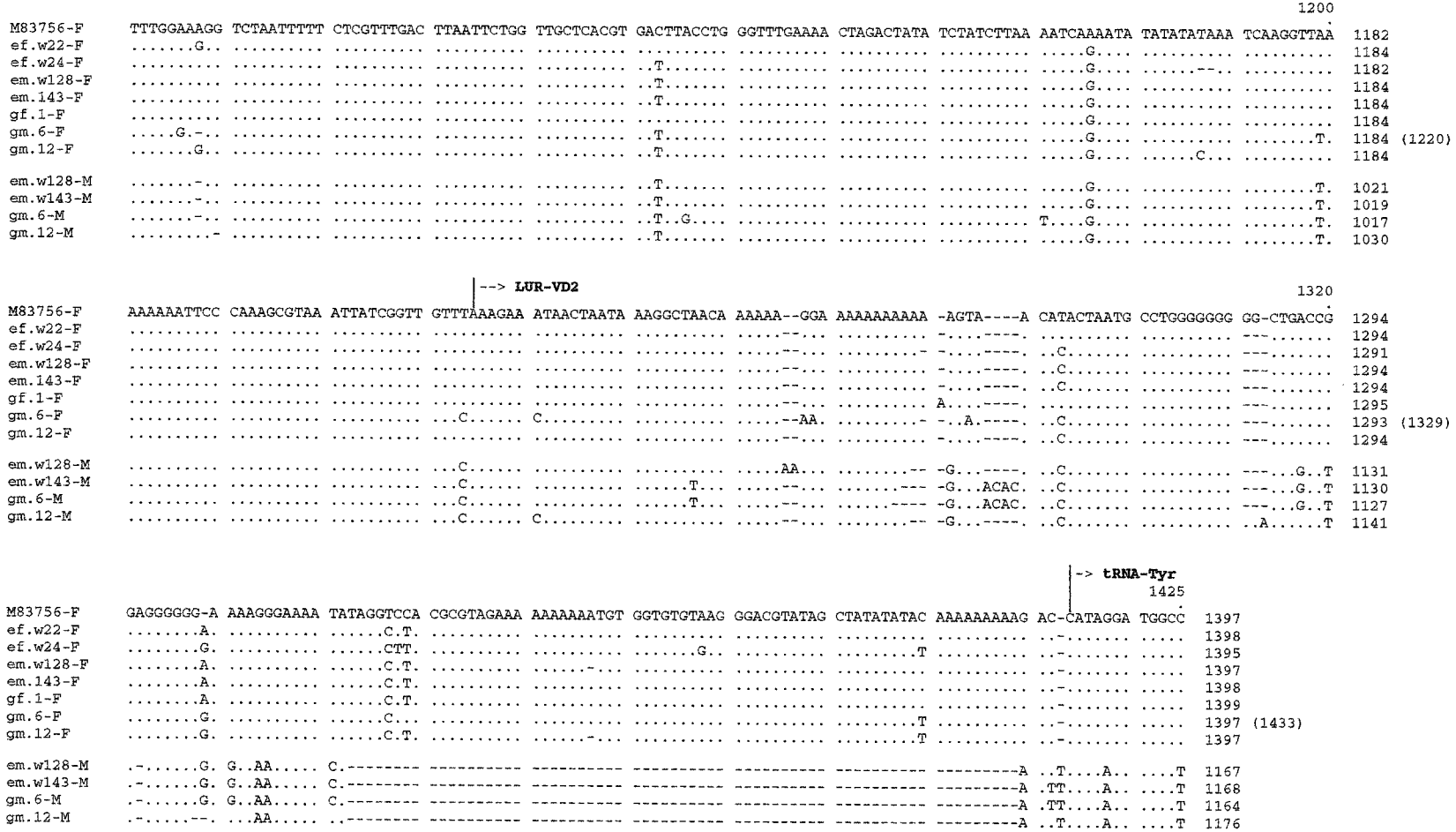
Appendix A

Multiple alignment of the twelve nucleotide sequences used in this study. The boxed sequence in gm.6-F is found in tandem repeat with two changes (T instead of A in the last position, and T instead of C in the seventh position from the end). Numbers in parenthesis (far right column) indicate the actual length of sequence gm.6-F.

```
             ...--> 16S-rRNA                                                                                              120
M83756-F     TAGATATCCT AGAGGCGGAG AAGCTTCTGA TGGTGGGTCT GTTCGCCCTT TAAAATCTAA CATGAGCTGA GTTCAGAACG GCGTGAGCTA GTTCAGTTTC TATCCTCTTT TTGAAA-TGA  119
ef.w22-F     .......... .......T.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ......-...  119
ef.w24-F     .......... .......T.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ......-...  119
em.w128-F    .......... .......T.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ......-...  119
em.143-F     .......... .......T.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ......-...  119
gf.1-F       .......... .......T.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ......-...  119
gm.6-F       ......C... .......T.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ..A...-...  119
gm.12-F      .......... .......T.. .......... .......... .......... .......... .......... .......... .......... .......... .......... ......-...  119

em.w128-M    ...G.G.... .....T.T.. ....C...A. G...A..... .......... .......... .......... .......... .......... .......... .......... .AA...A.A.  120
em.w143-M    ...G.G.... .....T.T.. ....C...A. G...A..... .......... .......... .......... .......... .......... .......... .......... .AA...-.A.  119
gm.6-M       ...G.G.... .....T.T.. ....C...A. G...A..... .......... .......... .......... .......... .......... .......... .......... .AA...-.A.  119
gm.12-M      ...G.G.... .....T.T.. ....C...A. G...A..... .......... .......... .......... .......... .......... .......... .......... .AA...-.A.  119

                                                                                                                 |--> LUR-VD1
M83756-F     GCTAATTTTG TACGAAAGGA CTTTTTCGCT AAAGCAATGC TTTGTCACAA ATCTGTGGTT GCATAAATAA TAGGGTATAA TAGGTTGATT AGCCTATAGG AAGGGTTACC TTTTATGTAA  239
ef.w22-F     .......... .......... .......... .......... .......... .......... .......... ...T...... ....C..... .......... .......... ..........  239
ef.w24-F     .......... .......... .......... .......... .......... .......T.. ....G..... ...T..G... .......... ...T.....A .......... ..........  239
em.w128-F    .......... .......... .......... .......... .......... .......... .......... ...T...... .......... .......... .......... ..........  239
em.143-F     .......... .......... .......... .......... .......... .......... .......... ...T...... .......... .......... .......... ..........  239
gf.1-F       .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... .......... ..........  239
gm.6-F       .......... .......... .......... .......... .......... .......... .......... ...T...... C......... .......... .......... ..........  239
gm.12-F      .......... .......... .......... .......... .......... .......T.. ....G..... ...T...... .......... .........A .....C.... ..........  239

em.w128-M    .......... .......... .......... ....T..... ....G.C..G CC..A.AA.. A..C.....T ..ATAC..G. CGA.C....C .A.TC...AA .....C.G.. ..........  240
em.w143-M    .......... .......... .......... ....T..... ....G.C..G CC..A.AA.. A..C.....T ..ATAC..G. CGA.C....C .A.TC...AA .....C.G.. ..........  239
gm.6-M       .......... .......... .......... ....T..... ....G.C..G CC..A.AA.. A..C.....T ..ATAC..G. CGA.C....C .A.TC...AA .....C.G.. ..........  239
gm.12-M      .......... .......... .......... ....T..... ....G.C..G CC..A.AA.. A..C.....T .G.TAC..G. CG..C....C .A.TC...AA .....C.G.. ....G.....  239

                                                                                                                          360
M83756-F     ATGAGCATAT TGCTTGGTGA TAGGTTGTTA AGTGTGGTAG ATT-TTA--- -GAGAAGTCC TATGTAGCTG GTTTAAAGCA GGCTTTTTAG -TAAAAAAAG ACCTGCTGT- TGAGGGGAGT  352
ef.w22-F     .......... ..T....... .......... .......... ...-...--- -......... .......... .......... .......... -......... ....A....- ..........  352
ef.w24-F     .......... ..T....... .......... .......... ...-...--- -......... .......... .......... .......... -......... .........- .....A...C  352
em.w128-F    .......... ..T....... ..T....... .......... ...-...--- -......... .......... .......... .......... -......... .........- ..........  352
em.143-F     .......... ..T....... ..T....... .......... ...-...--- -......... .......... .......... .......... -......... .........- ..........  352
gf.1-F       .......... ..T....... .......... .......... ...-...--- -......... .......... .......... .......... -......... .........- ..........  352
gm.6-F       ....A..... ..T....... .......... ..C....... ...-...--- -...G..... .......... .......... .......... -......... .........- .....A.G..  352
gm.12-F      .......... ..T....... .......... .......... ...-...--- -......... .......... .......... .......... -......... .........- .....A....  352

em.w128-M    ..-..----- -.T.------ --..C.A.-. .....TC... ...AC..CAG G...GG.C.. -----..T.A A..AG...T. A--.CC.C.C C.TG....G. -....G.C.T ..G.TA---.  333
em.w143-M    ..-..----- -.T.------ --..C.A.-. .....TC... ...AC.GCAG G...GG.... -----..T.A A..AG...T. A--..C.C.C C.TG....G. -.T..T.T.T ..G.TA---.  332
gm.6-M       ..-..----- -.T.------ --..C.A.-. .....TC... ...AC.GCAG G...GG.... -----..T.A A..AG...T. A--..C.C.C C.TG....G. -.T..T.T.T ..G.TA---.  332
gm.12-M      .-..A----- -.T.------ --.TC.A.-. .....TC... .C.TC..TAG G...GG.CT. -----..T.A A..AG...T. A--CCC-C.C C.TG....G. -...ATCT.T ....TA---.  331
```


480
M83756-F GTGAGATACT T-TTCTTTCA CATAAA--AA TTCCATGTTT GGAATAGATG CAGGAGATGG GGGCTTATGC CAGGGAGGAG AAGGTGACGG TGCCCAAGCA CGTGTTACTC CTGAAGCTGT 469
ef.w22-F 469
ef.w24-F 469
em.w128-F 469
em.143-F 469
gf.1-F 469
gm.6-F 469
gm.12-F 469
em.w128-M AC...GGGT. .G...AAA.. -.C...CT.. .....G.G-- ..G.G..... TG--..-.-- ---...T.-- --------.A ..A-.AG--- ------..AT. A.-.C...-. .AA..A--A. 419
em.w143-M A....GGGT. .G...AAA.. -.C...CT.. ....-G.G-- ..G.G..... T.--..-.-- ------.T.-- --------.A T.A-.AG--- ------..AT. A.-.C...-. .AA..A--A. 417
gm.6-M A....GGGT. .G...AAA.. -.C...CT.. ....-G.G-- ..G.G..... T.--..-.-- ---...T.-- --------.A T.A-.AG--- ------..AT. A.-.C...-. .AA..A--A. 417
gm.12-M A..G.GGGT. .G...AAA.. -.....CT.. ..TT.G.A-- ....A.AG.. TG--..-.-- ---...T.-- --------.A ..AA.A.--- ------..A.. A.-.C...-. .AA..A--A. 418
600
M83756-F ATGGGGTGGT GGAGATACAC CTGTAAATGC GGGCGCCGAA GCCGCTCCTG ATAATGCGGA GGAGGCGGGG GGGTACGTTG GAGACG-CTA TGCCCTTCCG TTAGAAGAAG TAGGTTGTAG 588
ef.w22-F 589
ef.w24-F 589
em.w128-F 589
em.143-F 589
gf.1-F 589
gm.6-F 589 (625)
gm.12-F 589
em.w128-M 485
em.w143-M 484
gm.6-M 482
gm.12-M 494
720
M83756-F TTCAGTTGAA GAAAGAGAAA GGGCTGTCGC GGAGCCTG-A AGTTGTATCT AGCGGTTTCG AGCCGGTCGA GCAGAGAGGT GTGCTAATTA GCGAAGCTAG TGGTGCCATT AATGCTGGCA 707
ef.w22-F 708
ef.w24-F 708
em.w128-F 708
em.143-F 708
gf.1-F 708
gm.6-F 708 (744)
gm.12-F 708
em.w128-M A.--...T.. A..T----.. ..--...T-- ----T..AC. C.C.-----. .A.--..C.C T...AC.--- ..-------. -..-...C-. .--...-... .--C...-.C TG.C.CCCT- 564
em.w143-M A.--.C.T.. A..T----.. ..--...T-- ----T..AC. C.C.-----. .A.--..C.C T...AC.--- ..-------. -..-...C-. .--...-... .--C...-.C TG.C.CCCT- 563
gm.6-M A.--.C.T.. A..T----.. ..--...T-- ----T..AC. C.C.-----. .A.--..C.C T...AC.--- ..-------. -..-...C-. .--...-... .--C...-.C TG.C.CCCT- 561
gm.12-M A.--...TG. A..T----.. ..--.A.T-- ----T..AC. ..C.-----. .A.--....C T...AC.--- ..-------. -..-...C-. .--...-... .--C...-.C TG.C.CCCT- 573


840
M83756-F AAGAAAGGTT TAGTGATTGT TA-ATGCCCA GGTCTAAAGG TTGCTGGTGC GC-TAATAAT CGTACTTTAT TAAGATAATT TATCAGCTAA GCAAAATTAC ACTTAAATAG TGTTTGCTAG 825
ef.w22-F 826
ef.w24-F 826
em.w128-F 826
em.143-F 826
gf.1-F 826
gm.6-F 826 (862)
gm.12-F 826
em.w128-M 665
em.w143-M 664
gm.6-M 662
gm.12-M 674
--> LUR-CD
960
M83756-F GCTTTACTAT ATTTTATTTT AAAATTGGTC CTATGCGGCT GCAAACATTT AATTA-ACCT CGTTTTAG-T TATGCTCTAT AGCCTGTAAA ATATAAAGCT GCAAAGCTTT ACAGCTGATA 943
ef.w22-F 945
ef.w24-F 945
em.w128-F 945
em.143-F 945
gf.1-F 945
gm.6-F 945 (981)
gm.12-F 945
em.w128-M 782
em.w143-M 780
gm.6-M 778
gm.12-M 791
1080
M83756-F AGCATTTGTC AGAGTCATGT GAGACTTAAC CTAATTAGTA AAAAAACAAG CGGAAATTTA AAGCCTCAAA TCCTAGAGTT ATCGAATTTT TATGGTAAAT TGTAAACTGT AAAA-TGGGA 1062
ef.w22-F 1064
ef.w24-F 1064
em.w128-F 1064
em.143-F 1064
gf.1-F 1064
gm.6-F 1065 (1101)
gm.12-F 1064
em.w128-M 902
em.w143-M 900
gm.6-M 898
gm.12-M 911


1200
M83756-F TTTGGAAAGG TCTAATTTTT CTCGTTTGAC TTAATTCTGG TTGCTCACGT GACTTACCTG GGTTTGAAAA CTAGACTATA TCTATCTTAA AATCAAAATA TATATATAAA TCAAGGTTAA 1182
ef.w22-F .....G.. ....G.... 1184
ef.w24-F ....T..... ....G.... .....--.. 1182
em.w128-F ....T..... ....G.... 1184
em.143-F ....T..... ....G.... 1184
gf.1-F ....G.... 1184
gm.6-F .....G.-.. ....T..... ....G.... ....T. 1184 (1220)
gm.12-F .....G.. ....T..... ....G.... .....C... 1184
em.w128-M .....-.. ....T..... ....G.... ....T. 1021
em.w143-M .....-.. ....T..... ....G.... ....T. 1019
gm.6-M .....-.. ....T..G.... .... T....G.... ....T. 1017
gm.12-M .....- ....T..... ....G.... ....T. 1030
--> LUR-VD2
1320
M83756-F AAAAAATTCC CAAAGCGTAA ATTATCGGTT GTTTAAAGAA ATAACTAATA AAGGCTAACA AAAAA--GGA AAAAAAAAAA -AGTA----A CATACTAATG CCTGGGGGGG GG-CTGACCG 1294
ef.w22-F ....--... ..... -.....-. .... ---..... 1294
ef.w24-F ....--... ..... -....---. ..C..... ..... ---..... 1291
em.w128-F ....--... ..... -....---. ..C..... ..... ---..... 1294
em.143-F ....--... ..... -....---. ..C..... ..... ---..... 1294
gf.1-F ....--... ..... A....---. .... ---..... 1295
gm.6-F ....C..... C..... ....--AA. ..... -..A.---. ..C..... ..... ---..... 1293 (1329)
gm.12-F ....--... ..... -....---. ..C..... ..... ---..... 1294
em.w128-M ....C..... ....AA... .....-- -G..---. ..C..... ..... ---...G..T 1131
em.w143-M ....C..... ....T... .....--... .....--- -G...ACAC. ..C..... ..... ---...G..T 1130
gm.6-M ....C..... ....T... .....--... .....---- -G...ACAC. ..C..... ..... ---...G..T 1127
gm.12-M ....C..... C..... ....--... .....-- -G..---. ..C..... ....A.....T 1141
-> tRNA-Tyr
1425
M83756-F GAGGGGGG-A AAAGGGAAAA TATAGGTCCA CGCGTAGAAA AAAAAAATGT GGTGTGTAAG GGACGTATAG CTATATATAC AAAAAAAAAG AC-CATAGGA TGGCC 1397
ef.w22-F .....A. ....C.T. ....-..... 1398
ef.w24-F .....G. ....CTT. ....G.. ....T ....-..... 1395
em.w128-F .....A. ....C.T. ....-... ....-..... 1397
em.143-F .....A. ....C.T. ....-..... 1398
gf.1-F .....A. ....C.T. ....-..... 1399
gm.6-F .....G. ....C... ....T ....-..... ..... 1397 (1433)
gm.12-F .....G. ....C.T. ....-... ....T ....-..... 1397
em.w128-M .-.....G. G..AA..... C.-------- ---------- ---------- ---------- ---------- ---------- ---------A ..T....A.. ....T 1167
em.w143-M .-.....G. G..AA..... C.-------- ---------- ---------- ---------- ---------- ---------- ---------A .TT....A.. ....T 1168
gm.6-M .-.....G. G..AA..... C.-------- ---------- ---------- ---------- ---------- ---------- ---------A .TT....A.. ....T 1164
gm.12-M .-.....--. ...AA..... ..-------- ---------- ---------- ---------- ---------- ---------- ---------A ..T....A.. ....T 1176

Appendix B

PCR amplification information.

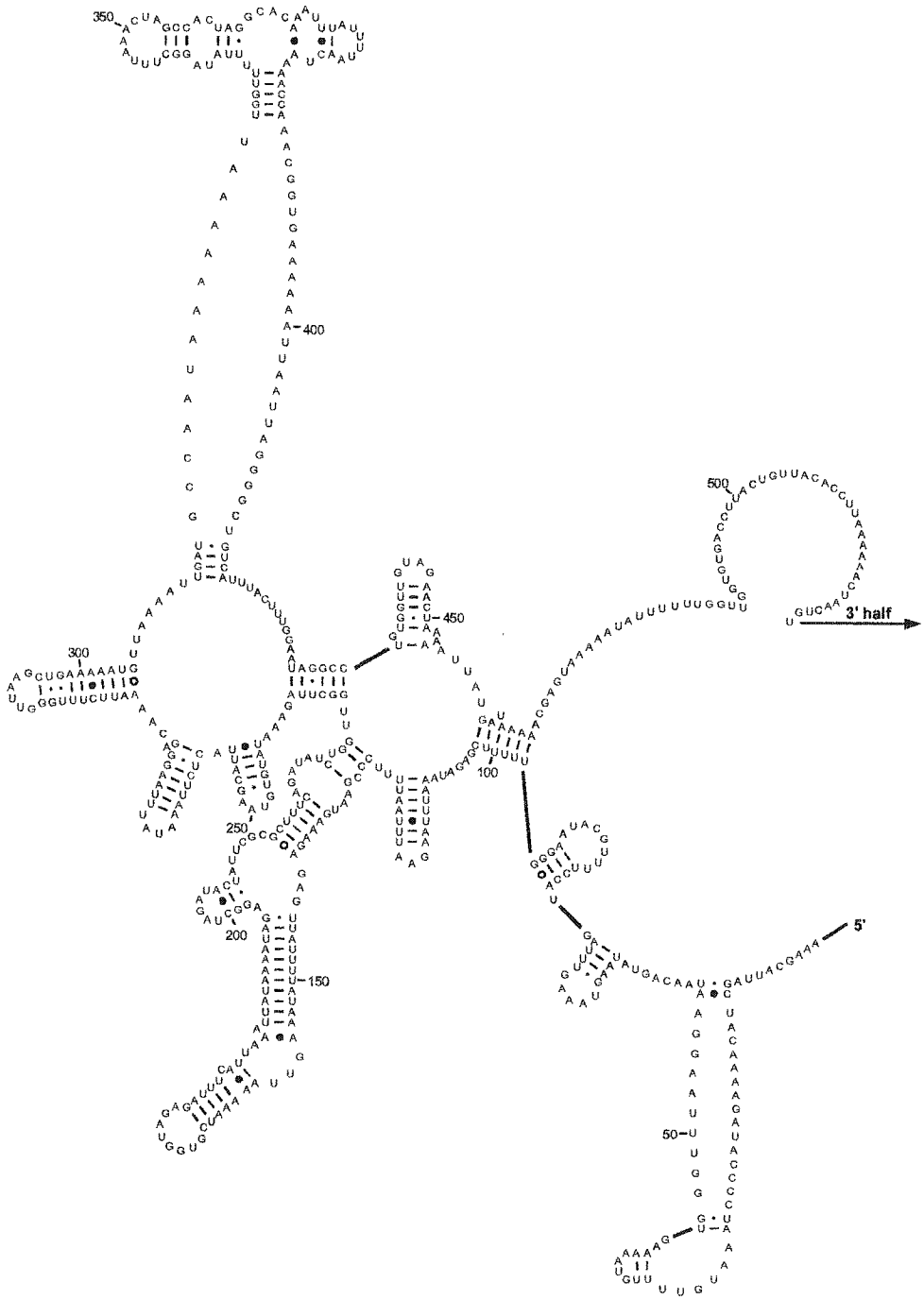
			Cycles[a]							
	Initial denature		Denature		Annealing		Extension		Final extension	
Primer set	Tm (°C)	T (min.)	Tm (°C)	T (sec.)	Tm (°C)	T (sec.)	Tm (°C)	T (sec.)	Tm (°C)	T (min.)
F1-16SBR	94	3	94	30	51	45	72	150	72	6
F2-R2	94	3	94	30	53	45	72	150	72	6
F4-Ucytb270R	94	3	94	30	51	45	72	180	72	6
F4-R4	94	3	94	60	53	90	72	90	72	6
F5-R5	94	3	94	30	51	45	72	90	72	6
Ucytb144F-Ucytb272I	94	3	94	30	46	45	72	45	72	4
F3-R3	94	3	94	30	51	45	72	180	72	6
16SAR-16SBR	94	3	94	20	51	20	72	45	72	4
F3-R3b	94	3	94	30	51	45	72	150	72	6

[a]: 40 cycles for all reactions.

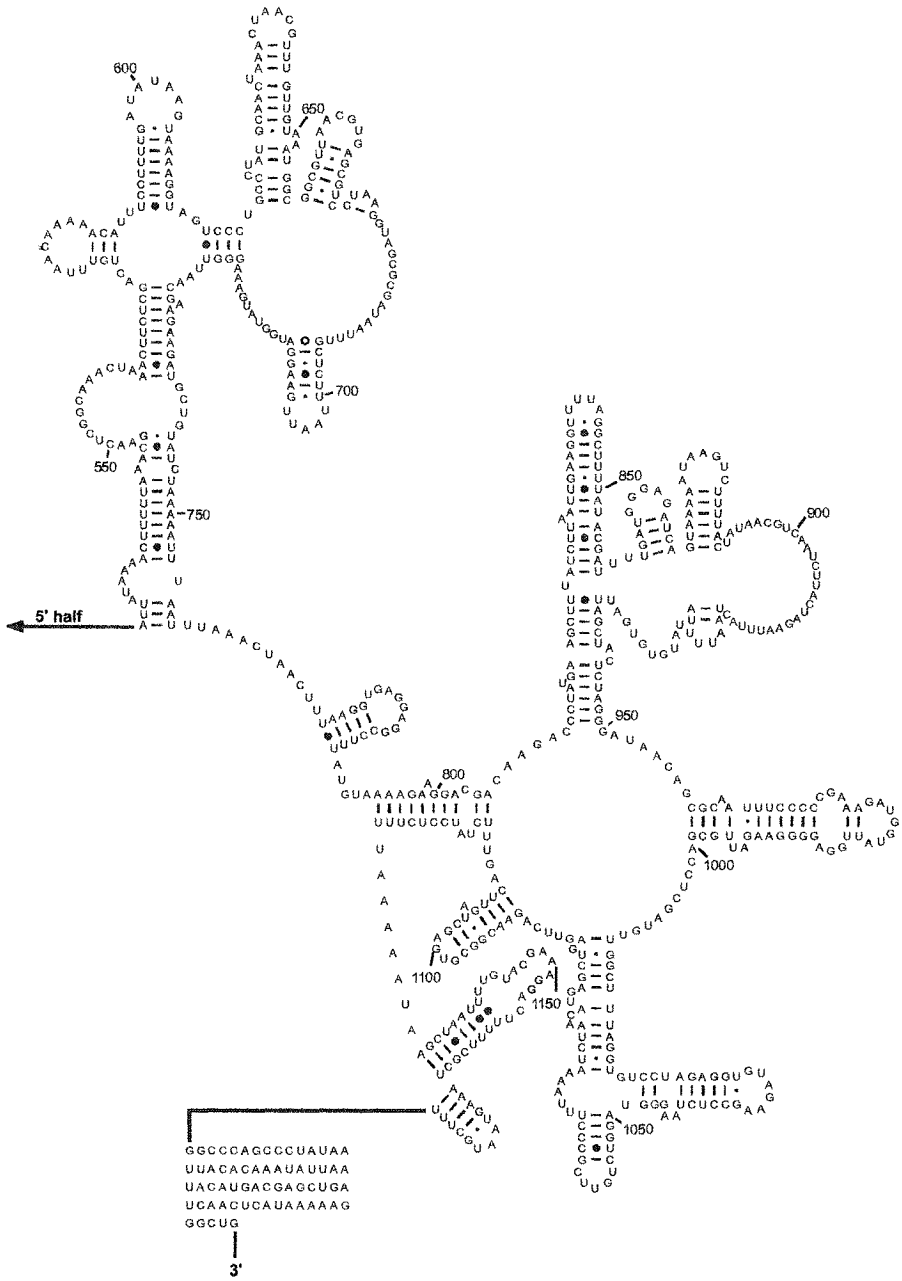
Appendix C

Secondary structure models of *Mytilus* 16S mitochondrial rRNA . *M. edulis* M-type 5' half (A1) and 3' half (A2); *M. trossulus* F-type 5' half (B1) and 3' half (B2); *M. trossulus* M-type 5' half (C1) and 3' half (C2); *M. californianus* F-type 5' half (D1) and 3' half (D2); *M. californianus* M-type 5' half (E1) and 3' half (E2).

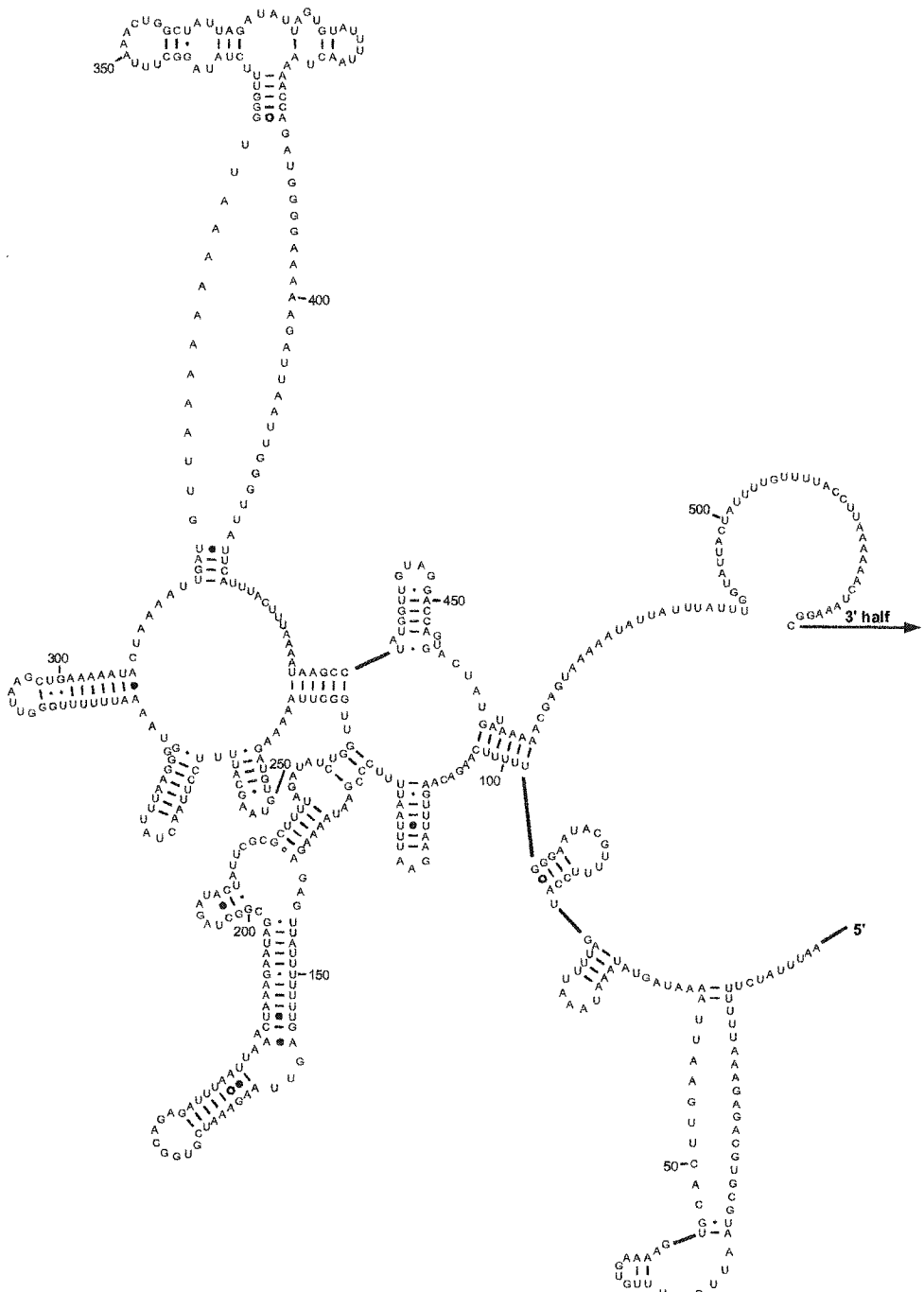
A1



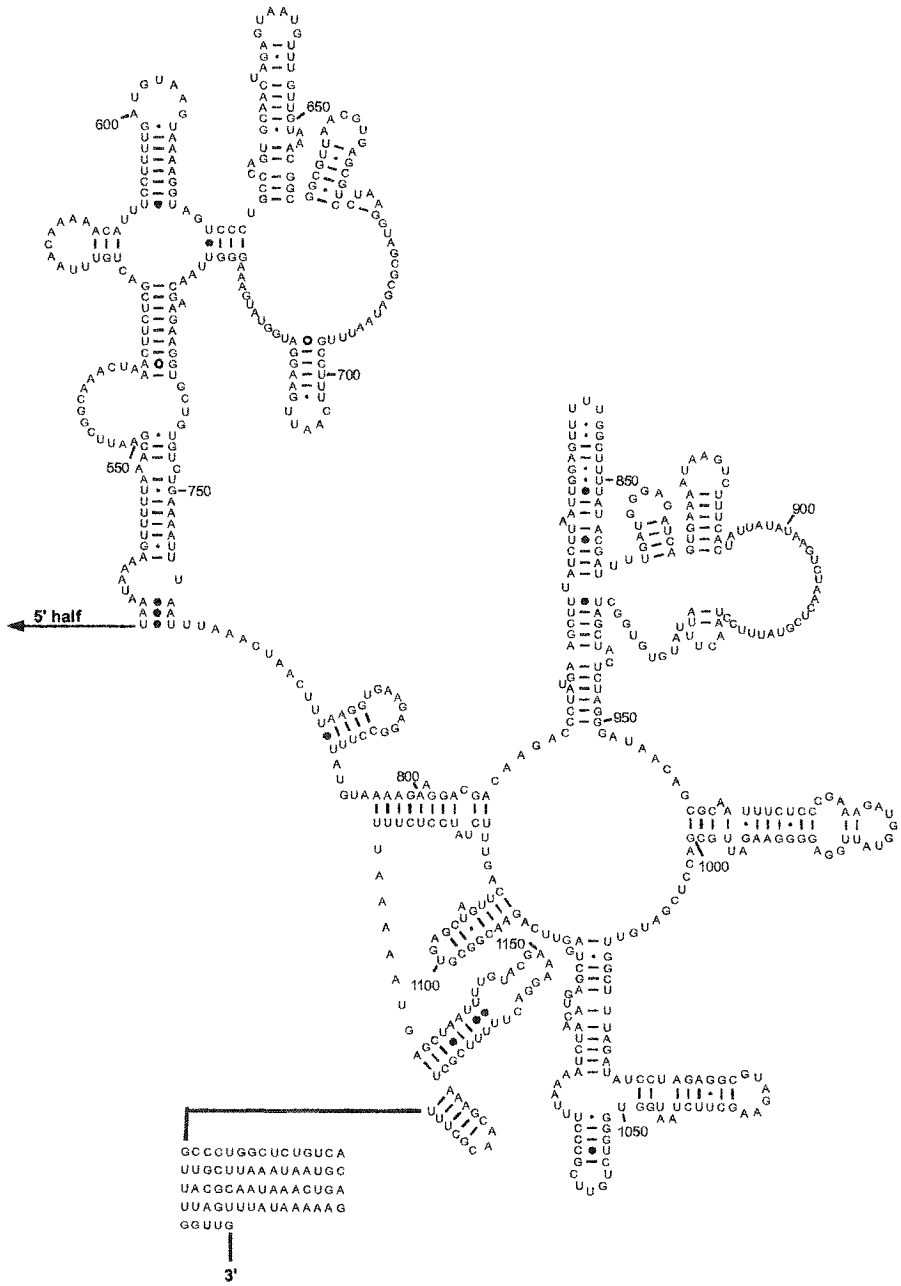
A2


600
650
700
550
750
5' half
800
850
900
950
1000
1100
1150
1050
GGCCCAGCCCUAUAA
UUACACAAAUAUUAA
UACAUGACGAGCUGA
UCAACUCAUAAAAAG
GGCUG
3'

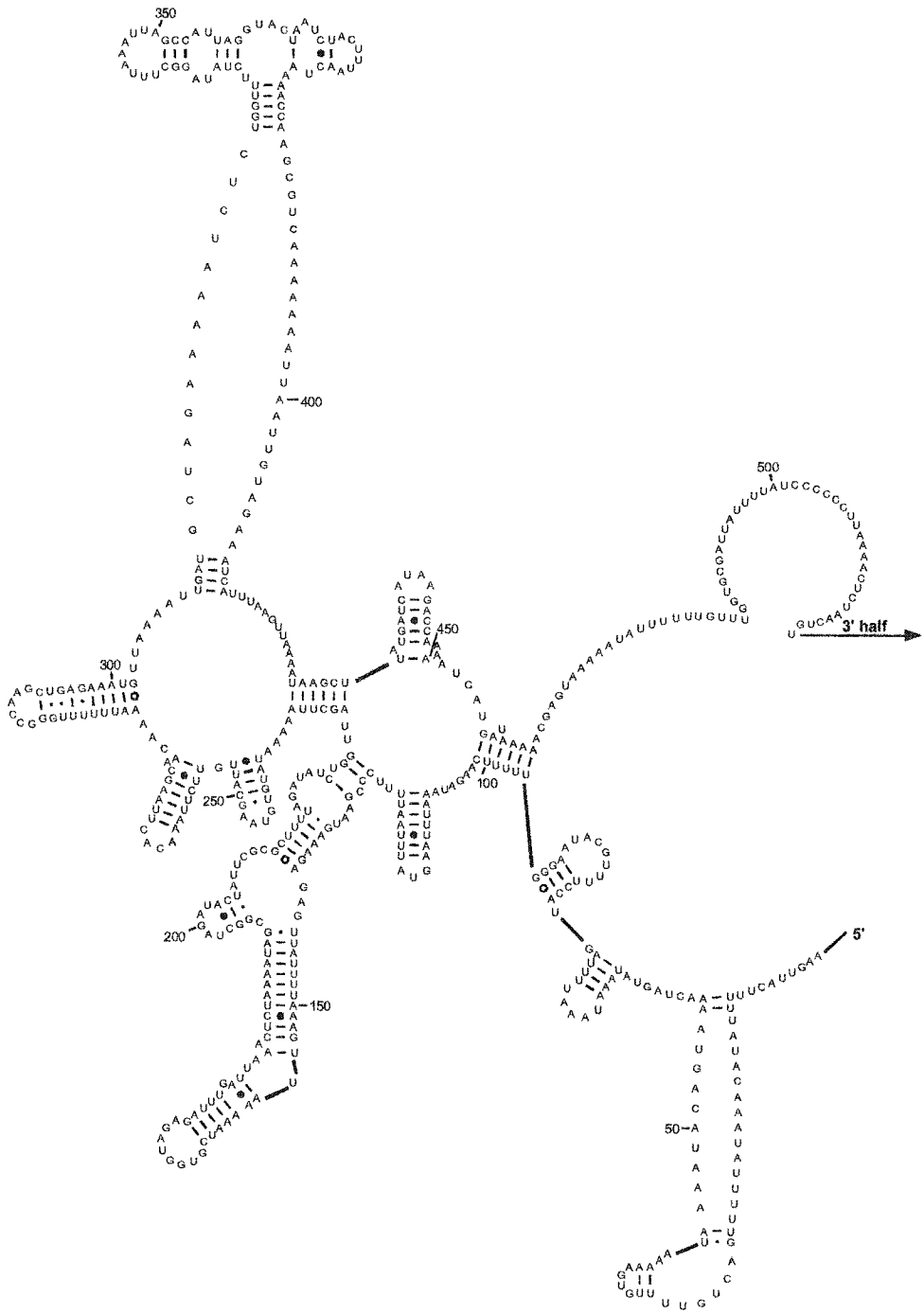
B1



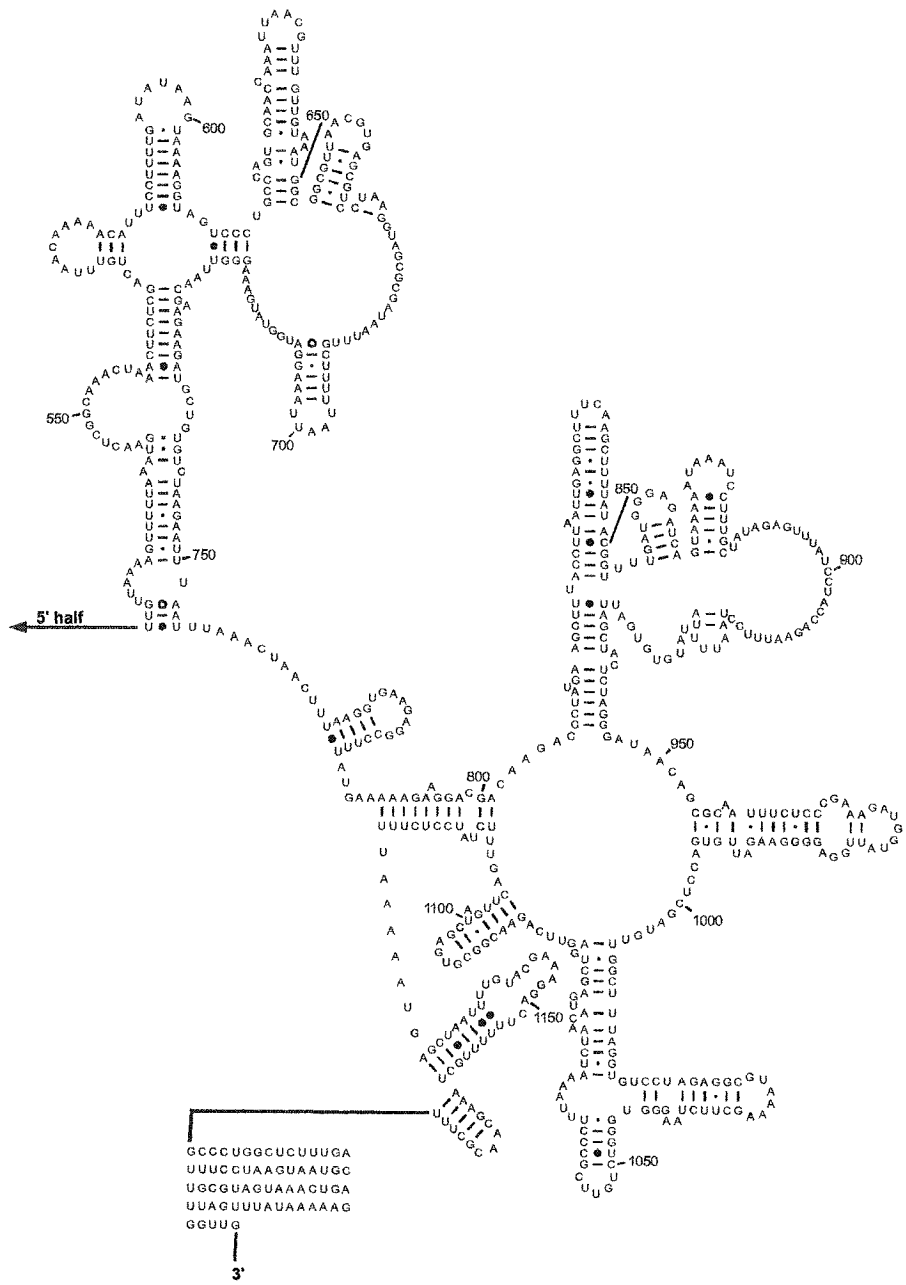
B2


5' half
600
650
700
750
800
850
900
950
1000
1050
1100
1150
GCCCUGGCUCUGUCA
UUGCUUAAAUAAUGC
UACGCAAUAAACUGA
UUAGUUUAUAAAAAG
GGUUG
3'

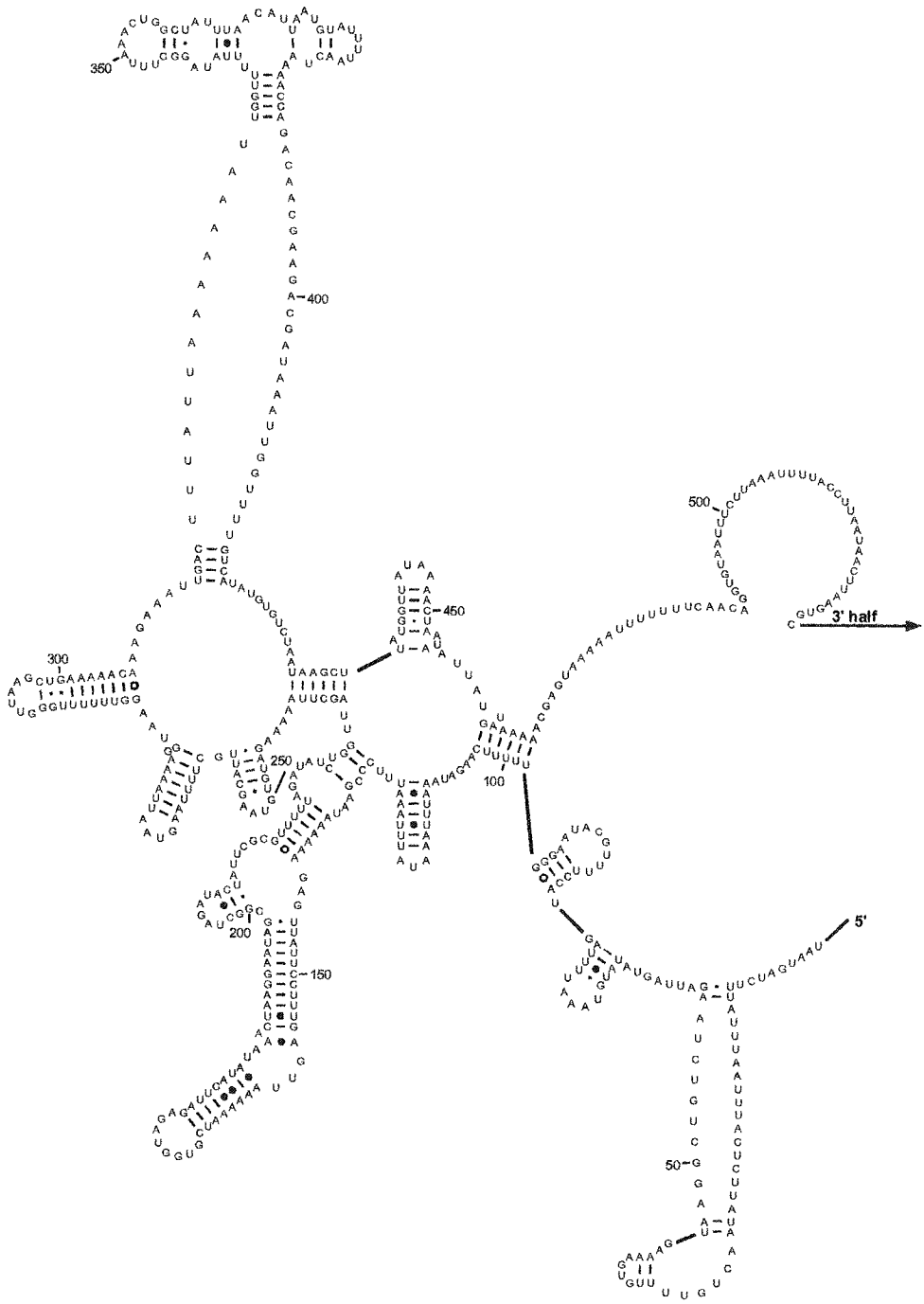
C1



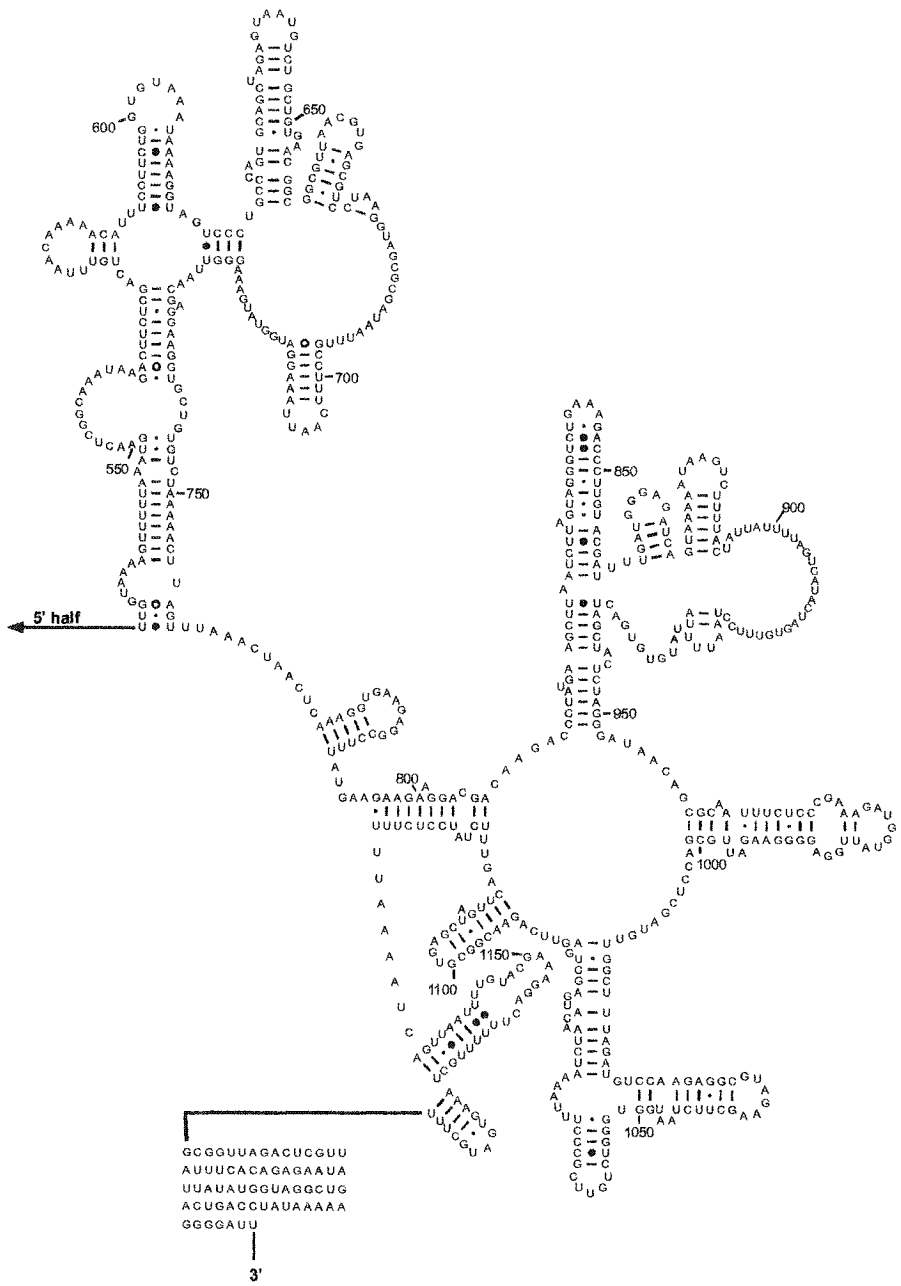
C2


600
650
700
550
750
5' half
800
850
900
950
1000
1100
1150
1050
GCCCUGGCUCUUUGA
UUUCCUAAGUAAUGC
UGCGUAGUAAACUGA
UUAGUUUAUAAAAAG
GGUUG
3'

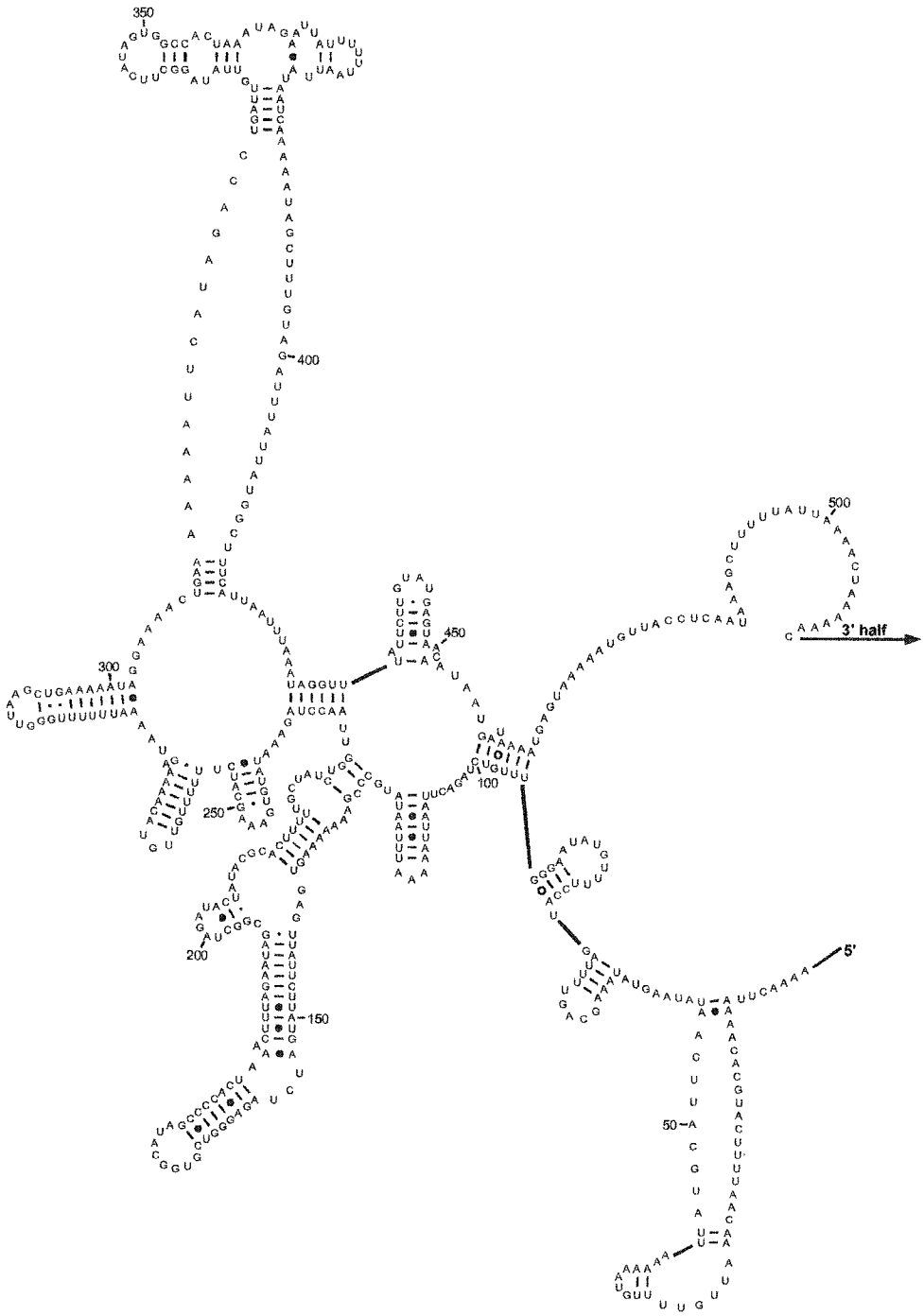
D1



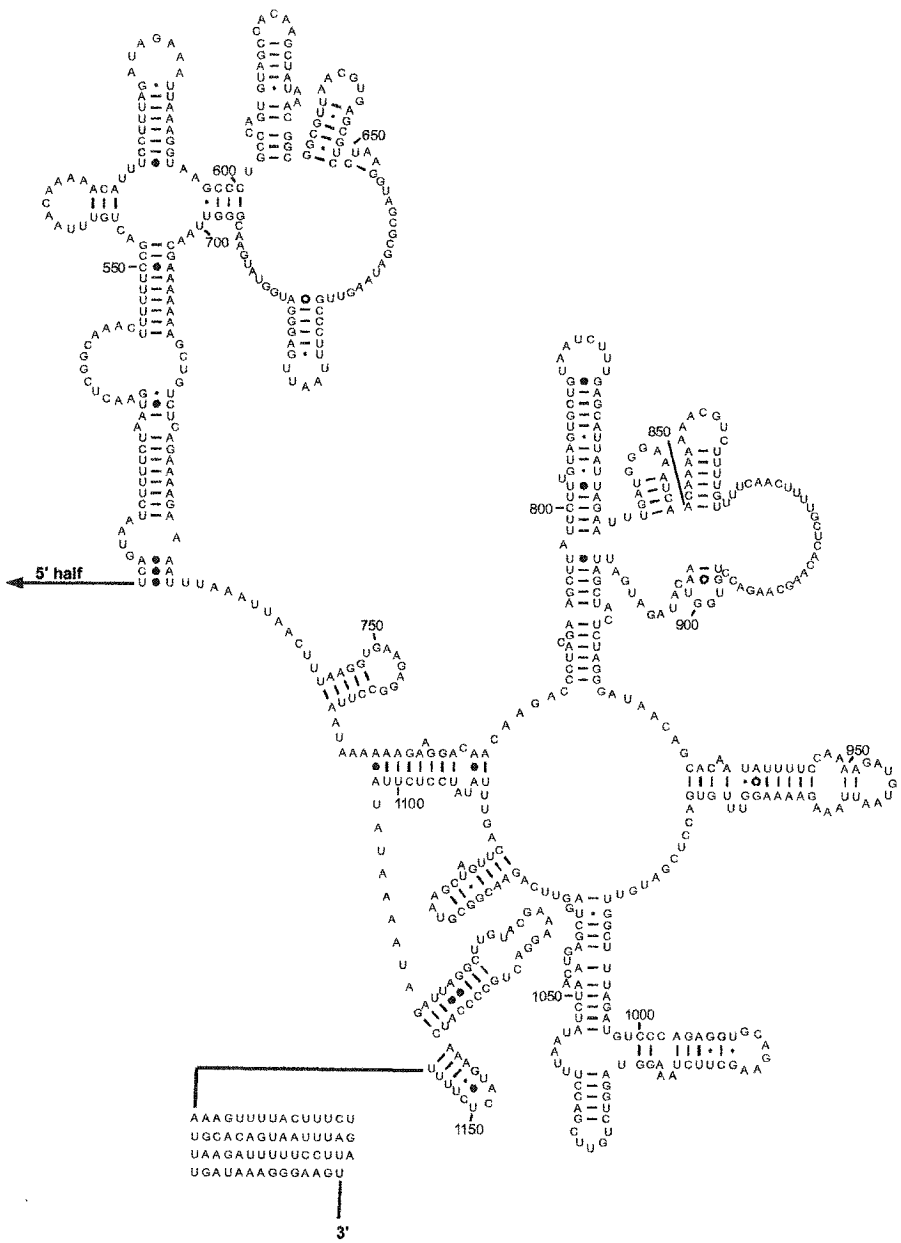
D2


5' half
600
650
700
750
800
850
900
950
1000
1050
1100
1150
GCGGUUAGACUCGUU
AUUUCACAGAGAAUA
UUAUAUGGUAGGCUG
ACUGACCUAUAAAAA
GGGGAUU
3'

E1



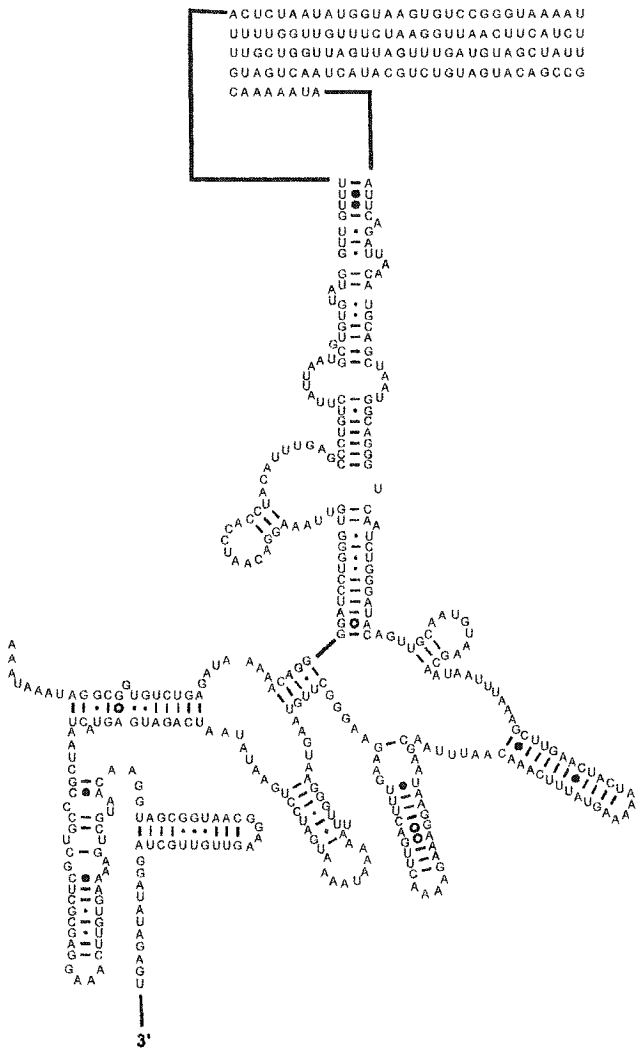
E2


5' half
600
650
700
550
750
800
850
900
950
1000
1050
1100
1150
AAAGUUUUACUUUCU
UGCACAGUAAUUUAG
UAAGAUUUUUCCUUA
UGAUAAAGGGAAGU
3'

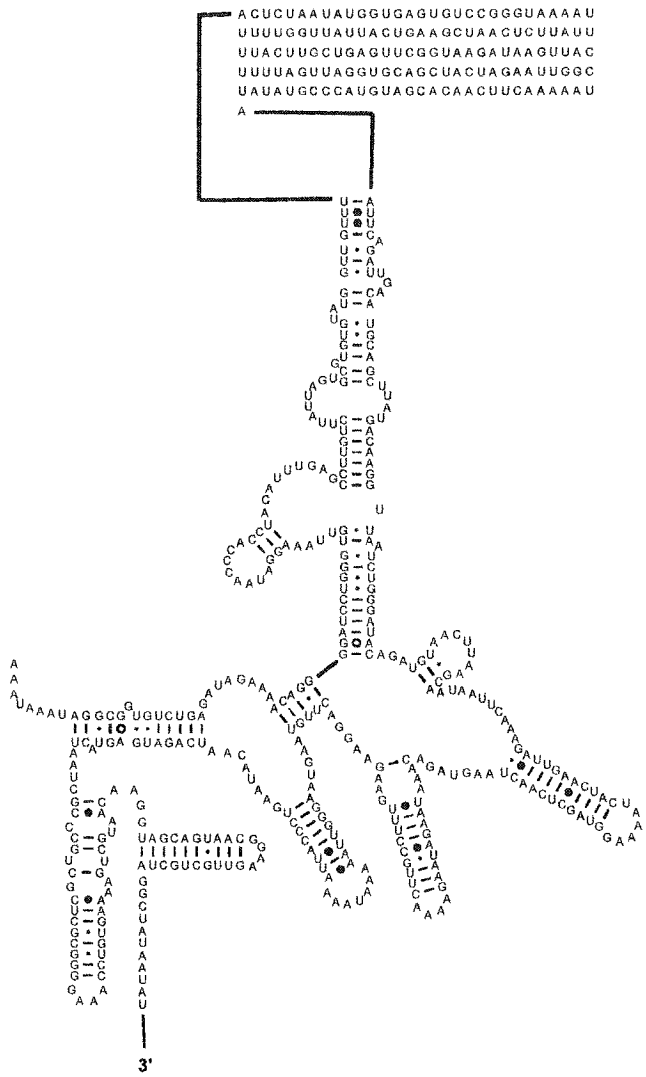
Appendix D

Secondary structure models of the third domain of *Mytilus* 12S mitochondrial rRNA. (A) *M. edulis* M-type; (B) *M. trossulus* F-type; (C) *M. trossulus* M-type; (D) *M. californianus* F-type; (E) *M. californianus* M-type.

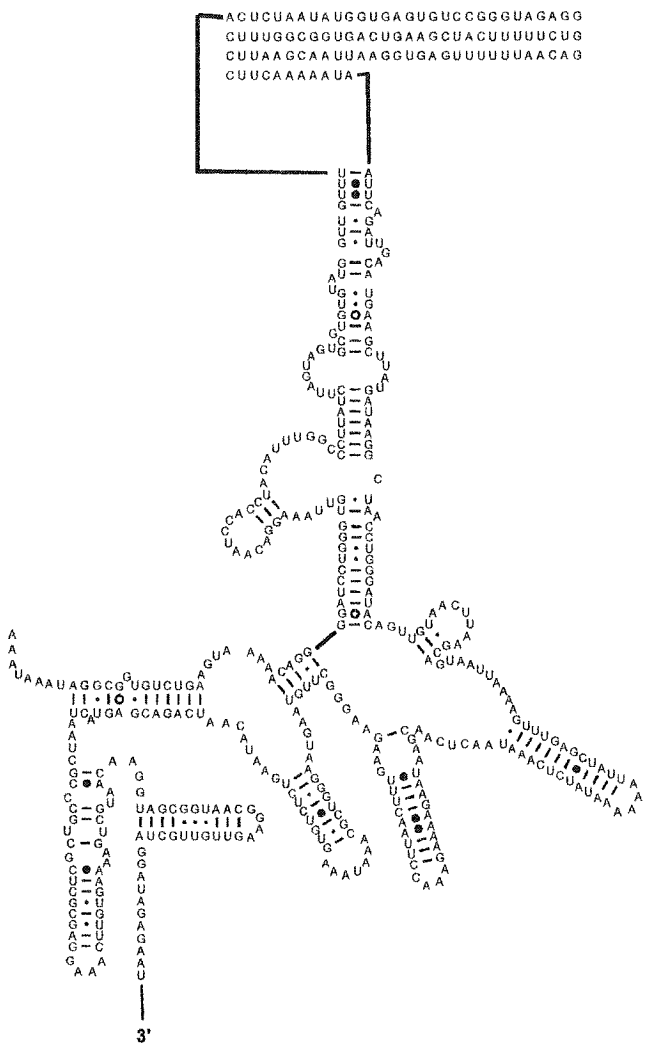
A



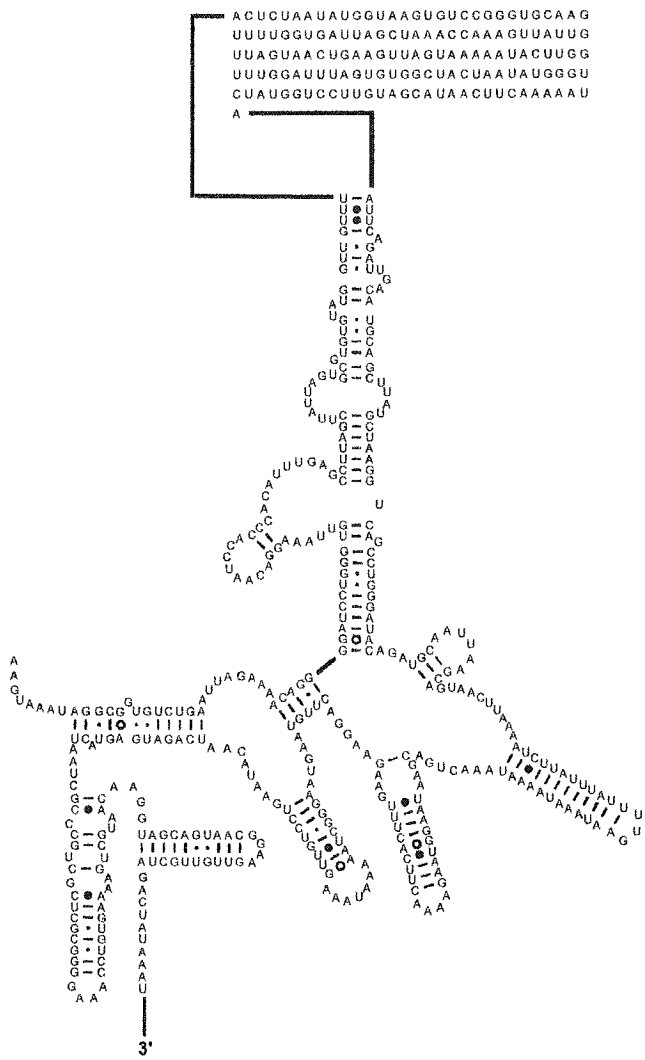
B


ACUCUAAUAUGGUGAGUGUCCGGGUAAAAU
UUUUGGUUAUUACUGAAGCUAACUCUUAUU
UUACUUGCUGAGUUCGGUAAGAUAAGUUAC
UUUUAGUUAGGUGCAGCUACUAGAAUUGGC
UAUAUGCCCAUGUAGCACAACUUCAAAAAU
A
3'

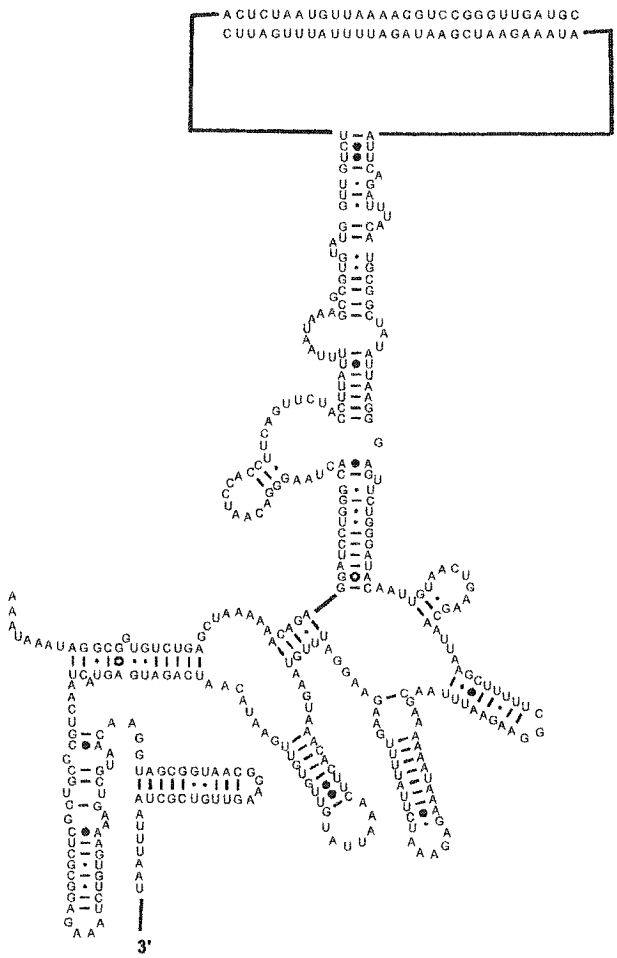
C


ACUCUAAUAUGGUGAGUGUCCGGGUAGAGG
CUUUGGCGGUGACUGAAGCUACUUUUUCUG
CUUAAGCAAUUAAGGUGAGUUUUUUAACAG
CUUCAAAAAUA
3'

D


ACUCUAAUAUGGUAAGUGUCCGGGUGCAAG
UUUUGGUGAUUAGCUAAACCAAAGUUAUUG
UUAGUAACUGAAGUUAGUAAAAAUACUUGG
UUUGGAUUUAGUGUGGCUACUAAUAUGGGU
CUAUGGUCCUUGUAGCAUAACUUCAAAAAU
A
3'

E

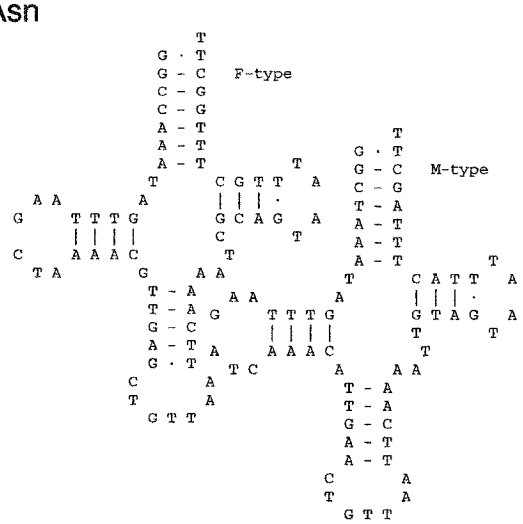

ACUCUAAUGUUAAAAGGUCCGGGUUGAUGC
CUUAGUUUAUUUUAGAUAAGCUAAGAAAUA
3'

Appendix E

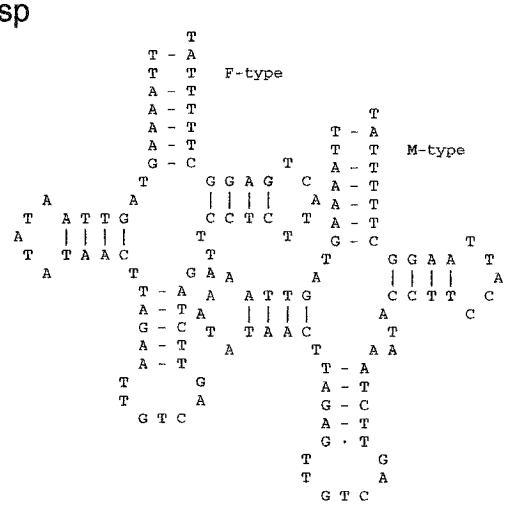
Secondary structures of 8 F-type and M-type tRNA genes from (A) *M. trossulus*; (B) *M. californianus.*

A

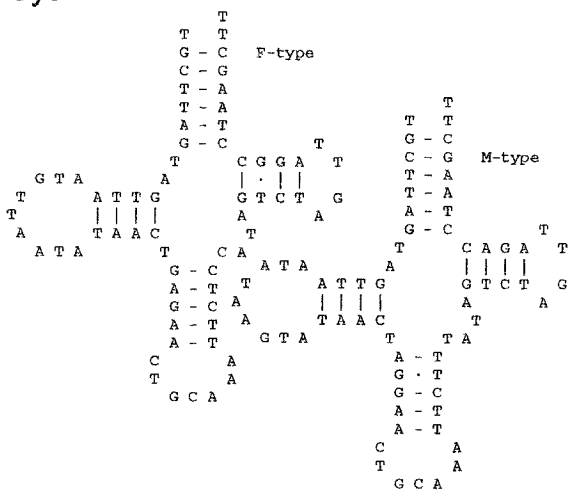
Asn



Asp

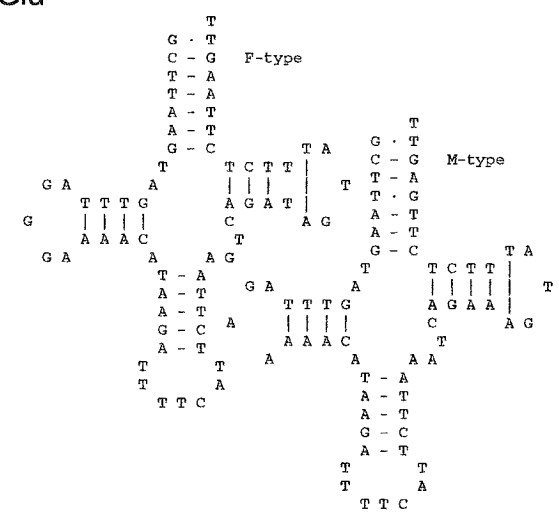


Cys

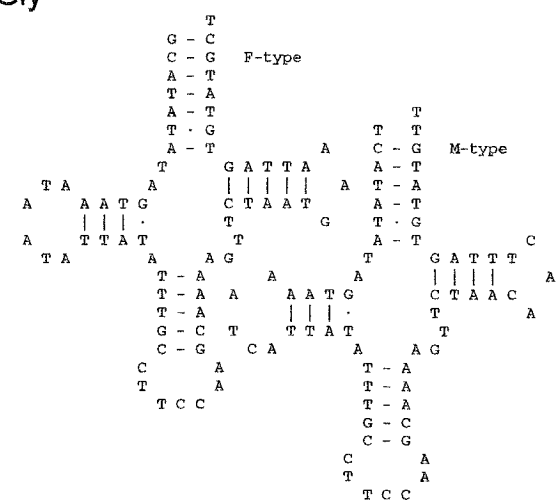


Gln

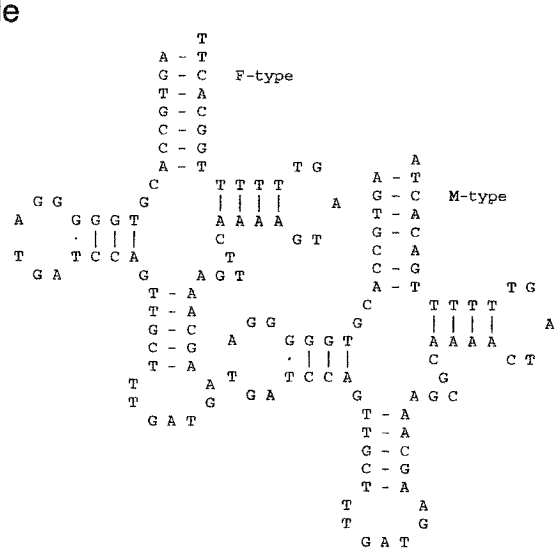
Glu



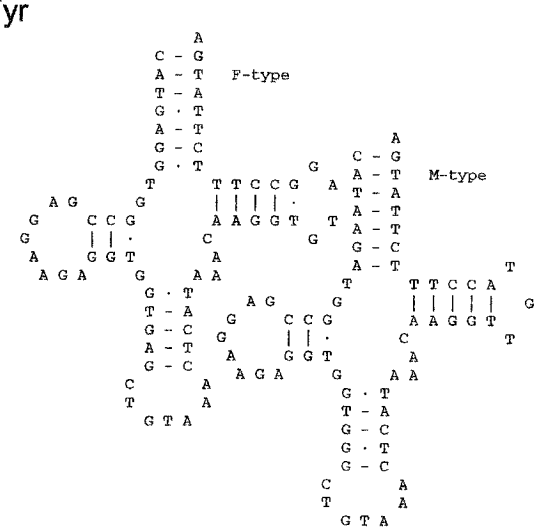
Gly



Ile

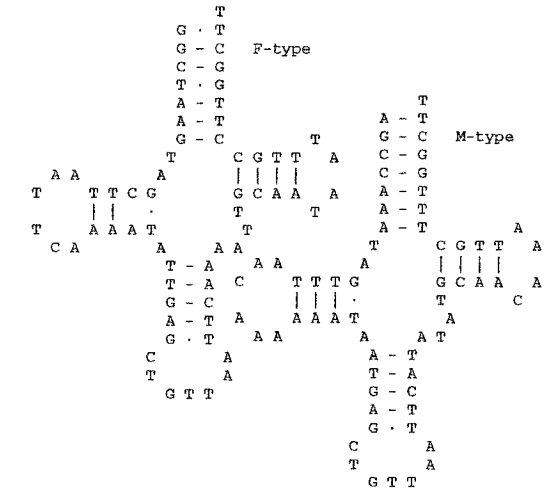


Tyr

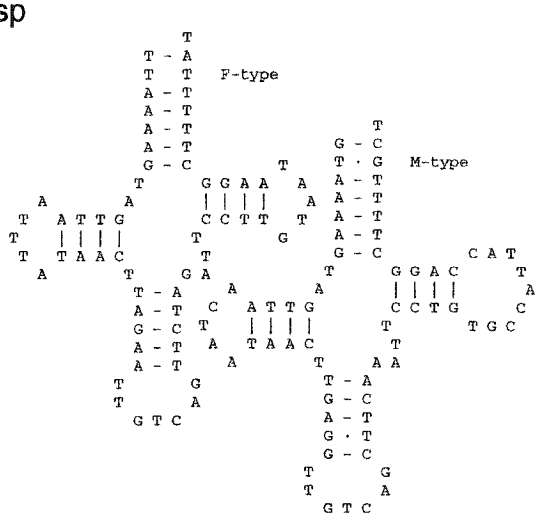


B

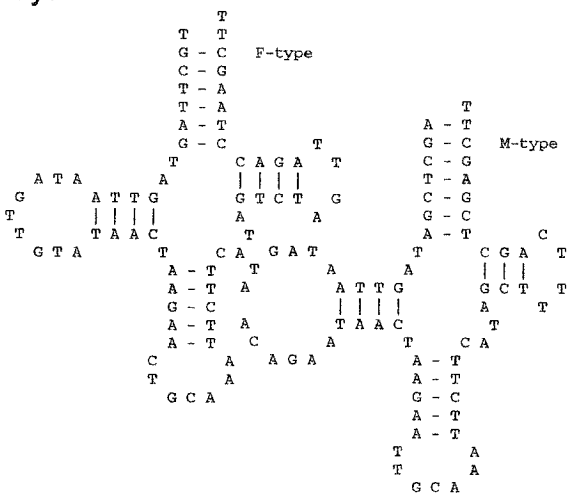
Asn



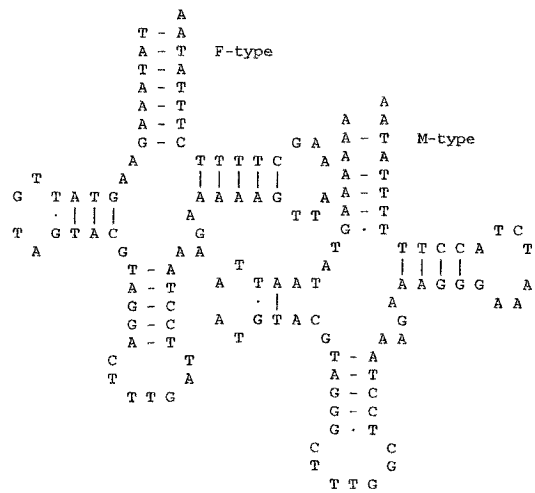
Asp



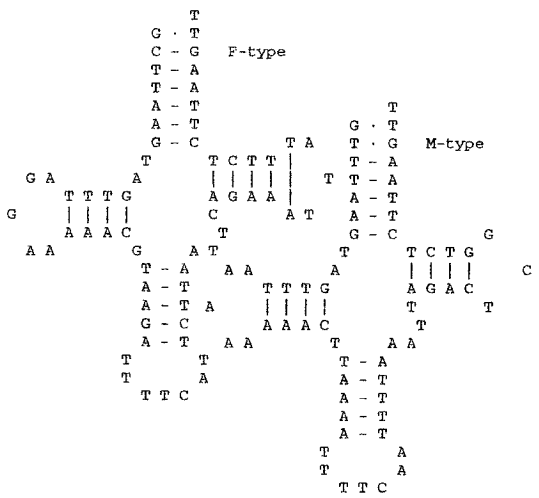
Cys



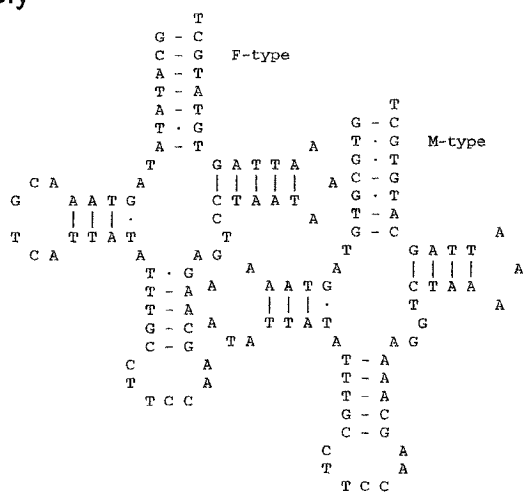
Gln



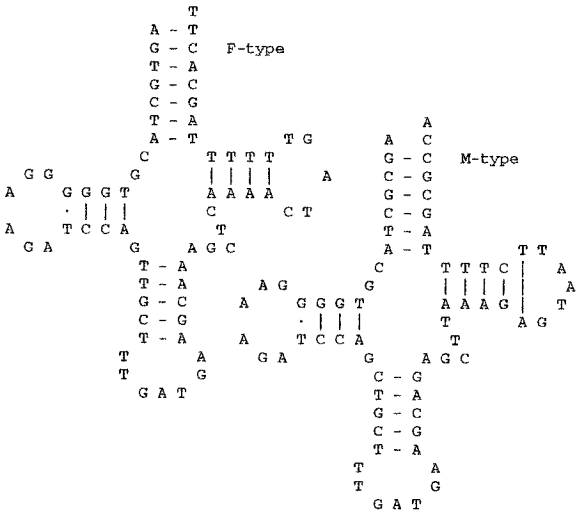
Glu



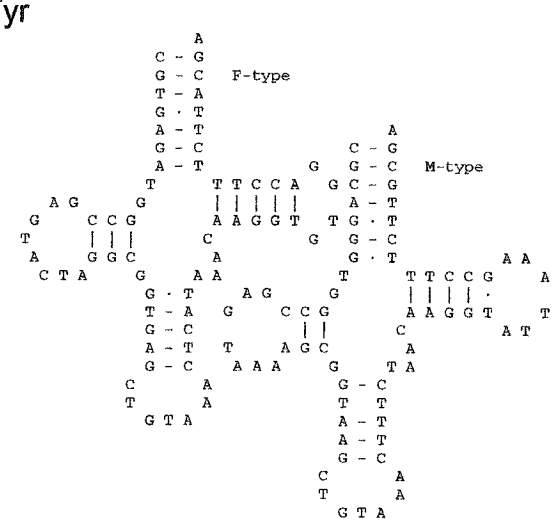
Gly



Ile



Tyr



References

Allen, S.K. (1986). Abnormal gametogenesis and sex ratio in triploid soft-shell clams (*Mya arenaria*). Biol. Bull. *170*, 198-220.

Anderson, W.A. (1968). Structure and fate of the paternal mitochondrion during early embryogenesis of *Paracentrotus lividus*. J. Ultrastruct. Res. *24*, 311-321.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M., and Howell,N. (1999). Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nat. Genet. *23*, 147.

Armstrong, M.R., Blok, V.C., and Phillips, M.S. (2000). A multipartite mitochondrial genome in the potato cyst nematode *Globodera pallida*. Genetics *154*, 181-192.

Åkesson, B. (1977). Crossbreeding and geographic races: experiments with the polychaete genus *Ophryotrocha*. Mikrofauna Meeresboden *61*, 11-18.

Backert, S., Neilson, B.L., and Borner, T. (1997). The mystery of the rings: structure and replication of mitochondrial genomes from higher plants. Trends Plant Sci. *2*, 477-483.

Bakeeva, L.E., Chentsov, Y.S., and Skulachev, V.P. (1981). Ontogenesis of mitochondrial reticulum in rat diaphragm muscle. Eur. J. Cell Biol. *25*, 175-181.

Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. Science *289*, 905-920.

Barsotti, G., and Meluzzi, C. (1968). Osservazioni su *Mytilus edulis* L. e *Mytilus galloprovincialis* Lamarck. Conchiglie Milan *4*, 50-58.

Beagley, C.T., Okimoto, R., and Wolstenholme, D.R. (1999). *Mytilus* mitochondrial DNA contains a functional gene for a tRNASer(UCN) with a dihydrouridine arm-replacement loop and a pseudo-tRNASer(UCN) gene. Genetics *152*, 641-652.

Beagley, C.T., Taylor, K.A., and Wolstenholme, D.R. (1997). Gender-associated diverse mitochondrial DNA molecules of the mussel *Mytilus californianus*. Curr. Genet *31*, 318-324.

Berg, O.G., and Kurland, C.G. (2000). Why mitochondrial genes are most often found in nuclei. Mol. Biol. Evol. *17*, 951-961.

Bernard, F.R. (1983). Catalogue of the living Bivalvia of the eastern Pacific Ocean: Bering Strait to Cape Horn. Can. Spec. Publ. Fish. Aquat. Sci. *61*, 1-102.

Birky, C.W.Jr. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proc. Natl. Acad. Sci. U. S. A *92*, 11331-11338.

Birky, C.W.Jr. (2001). The inheritance of genes in mitochondria and chloroplasts: laws, mechanisms, and models. Annu. Rev. Genet *35*, 125-148.

Boore, J.L., and Brown, W.M. (1994). Complete DNA sequence of the mitochondrial genome of the black chiton, *Katharina tunicata*. Genetics *138*, 423-443.

Bridge, D., Cunningham, C.W., Schierwater, B., DeSalle, R., and Buss, L.W. (1992). Class-level relationships in the phylum Cnidaria: evidence from mitochondrial genome structure. Proc. Natl. Acad. Sci. U. S. A. *89*, 8750-8753.

Brown, G.G., Gadaleta, G., Pepe, G., Saccone, C., and Sbisà, E. (1986). Structural conservation and variation in the D-loop-containing region of vertebrate mitochondrial DNA. J. Mol. Biol. *192*, 503-511.

Buckland-Nicks, J., Koss, R., and Chia, F.S. (1988). Fertilization in a chiton: acrosome-mediated sperm-egg fusion. Gamete Res. *21*, 199-212.

Bull, J.J. (1983). Evolution of sex determining mechanism (Benjamim-Cummings Publishing Co., Menlo Park, CA.).

Burzyński, A., Zbawicka, M., Skibinski, D.O., and Wenne, R. (2003). Evidence for recombination of mtDNA in the marine mussel *Mytilus trossulus* from the Baltic. Mol. Biol. Evol. *20*, 388-392.

Camasamudram, V., Fang, J.K., and Avadhani, N.G. (2003). Transcription termination at the mouse mitochondrial H-strand promoter distal site requires an A/T rich sequence motif and sequence specific DNA binding proteins. Eur. J. Biochem. *270*, 1128-1140.

Cannone, J.J., Subramanian, S., Schnare, M.N., Collett, J.R., D'Souza, L.M., Du, Y., Feng, B., Lin, N., Madabusi, L.V., Muller, K.M., Pande, N., Shang, Z., Yu, N., and Gutell, R.R. (2002). The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs: Correction. BMC. Bioinformatics. *3*, 15.

Cantatore, P., Roberti, M., Loguercio, P.P., Mustich, A., and Gadaleta, M.N. (1990). Mapping and characterization of *Paracentrotus lividus* mitochondrial transcripts: multiple and overlapping transcription units. Curr. Genet. *17*, 235-245.

Chang, D.D., and Clayton, D.A. (1984). Precise identification of individual promoters for transcription of each strand of human mitochondrial DNA. Cell *36*, 635-643.

Chang, D.D., and Clayton, D.A. (1985). Priming of human mitochondrial DNA replication occurs at the light-strand promoter. Proc. Natl. Acad. Sci. U. S. A. *82*, 351-355.

Chase, J.W., and Dawid, I.B. (1972). Biogenesis of mitochondria during *Xenopus laevis* development. Dev. Biol. *27*, 471-482.

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

Clary, D.O., and Wolstenholme, D.R. (1987). *Drosophila* mitochondrial DNA: conserved sequences in the A + T-rich region and supporting evidence for a secondary structure model of the small ribosomal RNA. J. Mol. Evol. *25*, 116-125.

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

Clayton, D.A. (2000). Transcription and replication of mitochondrial DNA. Hum. Reprod. *15 Suppl. 2*, 11-17.

Conklin, E.G. (1897). The embryology of *Crepidula*. J. Morphol. *13*, 1-226.

Cummins, J.M., Wakayama, T., and Yanagimachi, R. (1997). Fate of microinjected sperm components in the mouse oocyte and embryo. Zygote *5*, 301-308.

Curole, J.P., and Kocher, T.D. (2002). Ancient sex-specific extension of the cytochrome c oxidase II gene in bivalves and the fidelity of doubly-uniparental inheritance. Mol. Biol. Evol. *19*, 1323-1328.

D'Cotta, H., Fostier, A., Guiguen, Y., Govoroun, M., and Baroiller, J.F. (2001). Search for genes involved in the temperature-induced gonadal sex differentiation in the tilapia, *Oreochromis niloticus*. J. Exp. Zool. *290*, 574-585.

Dalziel, A.C., and Stewart, D.T. (2002). Tissue-specific expression of male-transmitted mitochondrial DNA and its implications for rates of molecular evolution in *Mytilus* mussels (Bivalvia: Mytilidae). Genome *45*, 348-355.

Delport, W., Ferguson, J.W., and Bloomer, P. (2002). Characterization and evolution of the mitochondrial DNA control region in hornbills (Bucerotiformes). J. Mol. Evol. *54*, 794-806.

Diaz, F., Bayona-Bafaluy, M.P., Rana, M., Mora, M., Hao, H., and Moraes, C.T. (2002). Human mitochondrial DNA with large deletions repopulates organelles faster than full-length genomes under relaxed copy number control. Nucleic Acids Res. *30*, 4626-4633.

Doda, J.N., Wright, C.T., and Clayton, D.A. (1981). Elongation of displacement-loop strands in human and mouse mitochondrial DNA is arrested near specific template sequences. Proc. Natl. Acad. Sci. U. S. A. *78*, 6116-6120.

Doolittle, W.F. (1998). You are what you eat: a gene transfer ratchet could account for bacterial genes in eukaryotic nuclear genomes. Trends Genet. *14*, 307-311.

Elliott, D.J., and Jacobs, H.T. (1989). Mutually exclusive synthetic pathways for sea urchin mitochondrial rRNA and mRNA. Mol. Cell Biol. *9*, 1069-1082.

Fisher, C., and Skibinski, D.O.F. (1990). Sex-biased mitochondrial DNA heteroplasmy in the marine mussel *Mytilus*. Proc. R. Soc. Lond. Ser B. Biol. Sci. *242*, 149-156.

Fisher, R.P., Topper, J.N., and Clayton, D.A. (1987). Promoter selection in human mitochondria involves binding of a transcription factor to orientation-independent upstream regulatory elements. Cell *50*, 247-258.

Forget, L., Ustinova, J., Wang, Z., Huss, V.A., and Lang, B.F. (2002). Hyaloraphidium curvatum: a linear mitochondrial genome, tRNA editing, and an evolutionary link to lower fungi. Mol. Biol. Evol. *19*, 310-319.

Fraser, C.M., Gocayne, J.D., White, O., Adams, M.D., Clayton, R.A., Fleischmann, R.D., Bult, C.J., Kerlavage, A.R., Sutton, G., Kelley, J.M., Fritchman, J.L., Weidman, J.F., Small, K.V., Sandusky, M., Furhmann, J., Nguyen, D., Utterback, T.R., Saudek, D.M., Phillips, C.A., Merrick, J.M., Tomb, J.-F., Dougherty, B.A., Bott, K.F., Hu, P.-C., Lucier, T.S. Peterson, S.N., Smith, H.O., Hutchison, C.A., and Venter, C.Jr. (1995). The minimal gene complement of *Mycoplasma genitalium*. Science *270*, 397-403.

Garrido-Ramos, M.A., Stewart, D.T., Sutherland, B.W., and Zouros, E. (1998). The distribution of male-transmitted and female-transmitted mitochondrial DNA types in somatic tissues of blue mussels: implications for the operation of doubly uniparental inheritance of mitochondrial DNA. Genome *41*, 818-824.

Geller, J.B. (1994). Sex-specific mitochondrial DNA haplotypes and heteroplasmy in *Mytilus trossulus* and *Mytilus galloprovincialis* populations. Mol. Mar. Biol. Biotech. *3*, 334-337.

Goddard, J.M., and Wolstenholme, D.R. (1978). Origin and direction of replication in mitochondrial DNA molecules from *Drosophila melanogaster*. Proc. Natl. Acad. Sci. U. S. A. *75*, 3886-3890.

Goddard, J.M., and Wolstenholme, D.R. (1980). Origin and direction of replication in mitochondrial DNA molecules from the genus *Drosophila*. Nucleic Acids Res. *8*, 741-757.

Gray, M.W. (1992). The endosymbiont hypothesis revisited. Int. Rev. Cytol. *141*, 233-357.

Gray, M.W., Burger, G., and Lang, B.F. (1999). Mitochondrial evolution. Science *283*, 1476-1481.

Gray, M.W., Lang, B.F., Cedergren, R., Golding, G.B., Lemieux, C., Sankoff, D., Turmel, M., Brossard, N., Delage, E., Littlejohn, T.G., Plante, I., Rioux, P., Saint-Louis, D., Zhu, Y., and Burger, G. (1998). Genome structure and gene content in protist mitochondrial DNAs. Nucleic Acids Res. *26*, 865-878.

Grun, P. (1996). Cytoplasmic Genetics and Evolution (Columbia University Press, New York).

Guo, X., Hedgecock, D., Hersherberger, W.K., Cooper, K., and Allen, S.K. (1998). Genetic component of protandric sex in a *Crassostrea* oyster. Evolution *53*, 394-402.

Guo, X., and Allen, S.K., Jr. (1994). Sex determination and polyploid gigantism in the dwarf surfclam (Mulinia lateralis Say). Genetics *138*, 1199-1206.

Gyllensten, U., Wharton, D., Josefsson, A., and Wilson, A.C. (1991). Paternal inheritance of mitochondrial DNA in mice. Nature *352*, 255-257.

Hales, K.G., and Fuller, M.T. (1997). Developmentally regulated mitochondrial fusion mediated by a conserved, novel, predicted GTPase. Cell *90*, 121-129.

Hatzoglou, E., Rodakis, G.C., and Lecanidou, R. (1995). Complete sequence and gene organization of the mitochondrial genome of the land snail *Albinaria coerulea*. Genetics *140*, 1353-1366.

Haugland, R.P. (1996). Handbook of fluorescent probes and research chemicals (Molecular Probes, Inc., Eugene, OR.).

Hayashi, J.I., Yonekawa, H., Gotoh, O., Watanabe, J., and Tagashira, Y. (1978). Strictly maternal inheritance of rat mitochondrial DNA. Biochem. Biophys. Res. Commun. *83*, 1032-1038.

Heath, D.D., Rawson, P.D., and Hilbish, T.J. (1995). PCR-based nuclear markers identify alien blue mussel (*Mytilus* spp.) genotypes on the west coast of Canada. Can. J. Fish. Aquat. Sci. *52*, 2621-2627.

Helm, M., Brule, H., Friede, D., Giege, R., Putz, D., and Florentz, C. (2000). Search for characteristic structural features of mammalian mitochondrial tRNAs. RNA *6*, 1356-1379.

Hermann, G.J., Thatcher, J.W., Mills, J.P., Hales, K.G., Fuller, M.T., Nunnari, J., and Shaw, J.M. (1998). Mitochondrial fusion in yeast requires the transmembrane GTPase Fzo1p. J. Cell Biol. *143*, 359-373.

Hickson, R.E., Simon, C., Cooper, A., Spicer, G.S., Sullivan, J., and Penny, D. (1996). Conserved sequence motifs, alignment, and secondary structure for the third domain of animal 12S rRNA. Mol. Biol. Evol. *13*, 150-169.

Hiraiwa-Hasegawa, M. (1993). Skewed birth sex ratios in primates: should high-ranking mother have daughters or sons? Trends Ecol. Evol. *8*, 395-400.

Hiraoka, J., and Hirao, Y. (1988). Fate of sperm tail components after incorporation into the hamster egg. Gamete Res. *19*, 369-380.

Hoeh, W.R., Stewart, D.T., Saavedra, C., Sutherland, B.W., and Zouros, E. (1997). Phylogenetic evidence for role-reversals of gender-associated mitochondrial DNA in *Mytilus* (Bivalvia: Mytilidae). Mol. Biol. Evol. *14*, 959-967.

Hoeh, W.R., Stewart, D.T., Sutherland, B.W., and Zouros, E. (1996a). Multiple origins of gender associated mitochondrial DNA lineages in bivalves (Mollusca: Bivalvia). Evolution *50*, 2276-2286.

Hoeh, W.R., Stewart, D.T., Sutherland, B.W., and Zouros, E. (1996b). Cytochrome c oxidase sequence comparisons suggest an unusually high rate of mitochondrial DNA evolution in *Mytilus* (Mollusca: Bivalvia). Mol. Biol. Evol. *13*, 418-421.

Hoepfner, D., van den, B.M., Philippsen, P., Tabak, H.F., and Hettema, E.H. (2001). A role for Vps1p, actin, and the Myo2p motor in peroxisome abundance and inheritance in *Saccharomyces cerevisiae*. J. Cell Biol. *155*, 979-990.

Hoffmann, R.J., Boore, J.L., and Brown, W.M. (1992). A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. Genetics *131*, 397-412.

Humphreys, W.J. (1962). Electron microscope studies on eggs of *Mytilus edulis*. J. Ultrastruct. Res. *7*, 467-487.

Hurst, G.D.D., Jiggins, F.M., Schulerlenburg, J.H.G., Bertrand, D., West, S.A., Goriacheva, I.I., Zakharov, I.A., Werren, J.H., Stouthamer, R., and Majerus, M.E.N. (1999). Male-killing *Wolbachia* in two species of insect. Proc. R. Soc. London Ser. B *366*, 735-740.

Ingman, M., Kaessmann, H., Pääbo, S., and Gyllensten, U. (2000). Mitochondrial genome variation and the origin of modern humans. Nature *408*, 708-713.

Jacobs, H.T., Herbert, E.R., and Rankine, J. (1989). Sea urchin egg mitochondrial DNA contains a short displacement loop (D-loop) in the replication origin region. Nucleic Acids Res. *17*, 8949-8965.

Jensen, R.E., Hobbs, A.E., Cerveny, K.L., and Sesaki, H. (2000). Yeast mitochondrial dynamics: fusion, division, segregation, and shape. Microsc. Res. Tech. *51*, 573-583.

Jenuth, J.P., Peterson, A.C., Fu, K., and Shoubridge, E.A. (1996). Random genetic drift in the female germline explains the rapid segregation of mammalian mitochondrial DNA. Nat. Genet. *14*, 146-151.

Jiggins, F.M., Hurst, G.D.D., and Majerus, M.E.N. (1998). Sex ratio distortion in *Acraea encedon* is caused by a male-killing bacterium. Heredity *81*, 87-91.

Kageyama, D., Hoshizaki, S., and Ishikawa, Y. (1998). Female biased sex ratio in the Asian corn borer, *Ostrinial furnacalist*: evidence for the occurrence of feminizing bacterial in an insect. Heredity *81*, 311-316.

Kaneda, H., Hayashi, J.H., Takahama, S., Taya, C., Lindahl, K.F., and Yonekawa, H. (1995). Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. Proc. Natl. Acad. Sci. U. S. A. *92*, 4542-4546.

Keddie, E.M., Higazi, T., and Unnasch, T.R. (1998). The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. Mol. Biochem. Parasitol. *95*, 111-127.

Kenchington, E., Landry, D., and Bird, C.J. (1995). Comparison of taxa of the mussel *Mytilus* (Bivalvia) by analysis of the nuclear small-subunit rRNA gene sequence. Can. J. Fish. Aquat. Sci. *52*, 2613-2620.

Kenchington, E., MacDonald, B., Cao, L., Tsagkarakis, D., and Zouros, E. (2002). Genetics of mother-dependent sex ratio in blue mussels (*Mytilus* spp.) and implications for doubly uniparental inheritance of mitochondrial DNA. Genetics *161*, 1579-1588.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. *16*, 111-120.

Kiyomoto, M., Komaru, A., Scarpa, J., Wada, K.T., and Danton, E. (1996). Abnormal gametogenesis, male dominant sex ratio, and Sertoli cell morphology in induced triploid mussels, *Mytilus galloprovincialis*. Zool. Sci. *13*, 393-402.

Koehler, C.M., Lindberg, G.L., Brown, D.R., Beitz, D.C., Freeman, A.E., Mayfield,J.E., and Myers,A.M. (1991). Replacement of bovine mitochondrial DNA by a sequence variant within one generation. Genetics *129*, 247-255.

Kondo, R., Satta, Y., Matsuura, E.T., Ishiwa, H., Takahata, N., and Chigusa, S.I. (1990). Incomplete maternal transmission of mitochondrial DNA in *Drosophila*. Genetics *126*, 657-663.

Kubo, T., Nishizawa, S., Sugawara, A., Itchoda, N., Estiati, A., and Mikami, T. (2000). The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta vulgaris L.*) reveals a novel gene for tRNA(Cys)(GCA). Nucleic Acids Res. *28*, 2571-2576.

Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. Bioinformatics *17*, 1244-1245.

Ladoukakis, E.D., Saavedra, C., Magoulas, A., and Zouros, E. (2002). Mitochondrial DNA variation in a species with two mitochondrial genomes: the case of *Mytilus galloprovincialis* from the Atlantic, the Mediterranean and the Black Sea. Mol. Ecol. *11*, 755-769.

Ladoukakis, E.D., and Zouros, E. (2001). Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. Mol. Biol. Evol. *18*, 1168-1175.

Lambert, C.C., and Battaglia, D.E. (1993). Loss of the paternal mitochondrion during fertilization. Zool. Sci. *10*, 31-37.

Lang, B.F., Gray, M.W., and Burger, G. (1999). Mitochondrial genome evolution and the origin of eukaryotes. Annu. Rev. Genet. *33* , 351-397.

Lango-Reynoso, F., Devauchelle, N., Le Pennec, M., and Hatt, P.J. (1999). Elements of reproductive strategy in oysters, *Crassostrea gigas*, from the "Rade de Brest", France. Invert. Reprod. Dev. *36*, 141-144.

Leutert, R. (1975). Sex-determination in *Bonellia*. In Interseuality in the animal kingdom, R.Reinboth, ed. (Springer, Berlin), pp. 84-90.

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

Lewis, D.L., Farr, C.L., Farquhar, A.L., and Kaguni, L.S. (1994). Sequence, organization, and evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA. Mol. Biol. Evol. *11*, 523-538.

Lewis, D.L., Farr, C.L., and Kaguni, L.S. (1995). *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. Insect Mol. Biol. *4*, 263-278.

Lindahl, T. (1993). Instability and decay of the primary structure of DNA. Nature *362*, 709-715.

Liu, H.P., Mitton, J.B., and Wu, S.K. (1996). Paternal mitochondrial DNA differentiation far exceeds maternal mitochondrial DNA and allozyme differentaition in the freshwater mussel, *Anodonta grandis grandis*. Evolution *50*, 952-957.

Longo, F.J., and Anderson, E. (1969). 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*, 97-119.

Longo, F.J. and Dornfeld, E.J. (1967). The fine structure of spermatid differentiation in the mussel, *Mytilus edulis*. J. Ultrastruct. Res. *20*, 462-480.

Lui, H.P., Mitton, J.B., and Wu, S.K. (1996). Paternal mitochondrial DNA differentiation far exceeds maternal DNA and allozyme differentiation in the freshwater mussel, *Anodonta grandis grandis*. Evolution *50*, 952-957.

Lydeard, C., Holznagel, W.E., Schnare, M.N., and Gutell, R.R. (2000). Phylogenetic analysis of molluscan mitochondrial LSU rDNA sequences and secondary structures. Mol. Phylogenet. Evol. *15*, 83-102.

Magoulas, A., and Zouros, E. (1993). Restriction-site heteroplasmy in anchovy (*Engraulis encrasicolus*) indicates incidental biparental inheritance of mitochondrial DNA. Mol. Biol. Evol. *10*, 319-325.

Manly, B.F.J. (1997). Randomization, Bootstrap and Monte Carlo Methods in Biology. (Chapman & Hall, London).

Marshall, H.D., and Baker, A.J. (1997). Structural conservation and variation in the mitochondrial control region of fringilline finches (*Fringilla* spp.) and the greenfinch (*Carduelis chloris*). Mol. Biol. Evol. *14*, 173-184.

Martin, W., and Herrmann, R.G. (1998). Gene transfer from organelles to the nucleus: how much, what happens, and Why? Plant Physiol. *118*, 9-17.

Martin, W., and Müller, M. (1998). The hydrogen hypothesis for the first eukaryote. Nature *392*, 37-41.

Mayhook, A.G., Rinaldi, A.M., and Jacobs, H.T. (1992). Replication origins and pause sites in sea urchin mitochondrial DNA. Proc. R. Soc. Lond B Biol. Sci. *248*, 85-94.

McInerney, J.O. (1998). Replicational and transcriptional selection on codon usage in *Borrelia burgdorferi*. Proc. Natl. Acad. Sci. U. S. A. *95*, 10698-10703.

Miller, S.A., Dykes, D.D., and Polesky, H.F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. *16*, 1215.

Morton, B.R. (1999). Strand asymmetry and codon usage bias in the chloroplast genome of *Euglena gracilis*. Proc. Natl. Acad. Sci. U. S. A. *96*, 5123-5128.

Neefs, J.M., Van de, P.Y., De Rijk, P., Chapelle, S., and De Wachter, R. (1993). Compilation of small ribosomal subunit RNA structures. Nucleic Acids Res. *21*, 3025-3049.

Nierman, W.C., Feldblyum, T.V., Laub, M.T., Paulsen, I.T., Nelson, K.E., Eisen, J.A., Heidelberg, J.F., Alley, M.R., Ohta, N., Maddock, J.R., Potocka, I., Nelson, W.C., Newton, A., Stephens, C., Phadke, N.D., Ely, B., DeBoy, R.T., Dodson, R.J., Durkin, A.S., Gwinn, M.L., Haft, D.H., Kolonay, J.F., Smit, J., Craven, M.B., Khouri, H., Shetty, J., Berry, K., Utterback, T., Tran, K., Wolf, A., Vamathevan, J., Ermolaeva, M., White, O., Salzberg, S.L., Venter, J.C., Shapiro, L., Fraser, C.M., and Eisen, J. (2001). Complete genome sequence of *Caulobacter crescentus*. Proc. Natl. Acad. Sci. U. S. A. *98*, 4136-4141.

Nunnari, J., Marshall, W.F., Straight, A., Murray, A., Sedat, J.W., and Walter, P. (1997). Mitochondrial transmission during mating in *Saccharomyces cerevisiae* is determined by mitochondrial fusion and fission and the intramitochondrial segregation of mitochondrial DNA. Mol. Biol. Cell *8*, 1233-1242.

Oldenburg, D.J., and Bendich, A.J. (2001). Mitochondrial DNA from the liverwort *Marchantia polymorpha*: circularly permuted linear molecules, head-to-tail concatemers, and a 5' protein. J. Mol. Biol. *310*, 549-562.

Passamonti, M., and Scali, V. (2001). Gender-associated mitochondrial DNA heteroplasmy in the venerid clam Tapes *philippinarum* (Mollusca Bivalvia). Curr. Genet. *39*, 117-124.

Peterson, J.J. (1972). Factors affection sex ratios of a mermithid parasite of mosquitoes. J. Nematology *4*, 83-87.

Peterson, J.J. (1977). Effects of host size and parasite burden on sex ratio in the mosquito parasite *Octomymermis muspratti.* J. Nemat. *9*, 343-346.

Pfannenstiel, H.D. (1984). Sex determination and intersexuality in polychaetes. Fortschr. -Zool. *29*, 81-98.

Pikó, L. and Taylor, K.D. (1987). Amounts of mitochondrial DNA and abundance of some mitochondrial gene transcripts in early mouse embryos. Dev. Biol. *123*, 364-374.

Pitnick, S., and Karr, T.L. (1998). Paternal products and by-products in *Drosophila* development. Proc. R. Soc. Lond B Biol. Sci. *265*, 821-826.

Pojeta, J.P., Runnegar, Jr.B., Peel, J .S., Gordon, Jr.,M., and Rowell, A.J. (1987). Phylum Mollusca. In Fossil Invertebrates, R.S.Boardman and A.H.Cheetham, eds. (Blackwell Scentific Publications, Oxford), pp. 270-435.

Quesada, H., Warren, M., and Skibinski, D.O. (1998). Nonneutral evolution and differential mutation rate of gender-associated mitochondrial DNA lineages in the marine mussel *Mytilus*. Genetics *149*, 1511-1526.

Quesada, H., Wenne, R., and Skibinski, D.O. (1999). Interspecies transfer of female mitochondrial DNA is coupled with role-reversals and departure from neutrality in the mussel *Mytilus trossulus*. Mol. Biol. Evol. *16*, 655-665.

Randi, E., and Lucchini, V. (1998). Organization and evolution of the mitochondrial DNA control region in the avian genus *Alectoris*. J. Mol. Evol. *47*, 449-462.

Rawson, P.D., and Hilbish, T.J. (1995). Evolutionary relationships among the male and female mitochondrial DNA lineages in the *Mytilus edulis* species complex. Mol. Biol. Evol. *12*, 893-901.

Rawson, P.D., and Hilbish, T.J. (1998). Asymmetric introgression of female and male lineage mitochondrial DNA (mtDNA) haplotypes within the European hybrid zone between *Mytilus edulis* and *M. galloprovincialis*. Evolution *53*, 100-108.

Rawson, P.D., Joyner, K.L., Meetze, K., and Hilbish, T.J. (1996a). Evidence for intragenic recombination within a novel genetic marker that distinguishes mussels in the *Mytilus edulis* species complex. Heredity *77*, 599-607.

Rawson, P.D., Secor, C.L., and Hilbish, T.J. (1996b). The effects of natural hybridization on the regulation of doubly uniparental mtDNA inheritance in blue mussels (*Mytilus* spp.). Genetics *144*, 241-248.

Ray, D.A., and Densmore, L. (2002). The crocodilian mitochondrial control region: general structure, conserved sequences, and evolutionary implications. J. Exp. Zool. *294*, 334-345.

Reyes, A., Gissi, C., Pesole, G., and Saavedra, C. (1998). Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol. Biol. Evol. *15*, 957-966.

Rigaud, T., Souty-grosset, C., Raimond, R., and Mocquard, J. (1991). Feminizing endocytobiosis in the terrestrial crustacean *Armadillidium vulgare* Latr. (Isopoda): recent acquisitions. Endocytobiosis Cell. Res. *7*, 259-273.

Riginos, C., Sukhdeo, K., and Cunningham, C.W. (2002). Evidence for selection at multiple allozyme loci across a mussel hybrid zone. Mol. Biol. Evol. *19*, 347-351.

Rousset, F., Bouchon, D., Pintureau, B., Juchault, P., and Solignac, M. (1992). *Wolbachia* endosymbionts respondible for various alterations of sexuality in arthropods. Proc. R. Soc. London Ser. B *250*, 91-98.

Rozas, J., and Rozas, R. (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. Bioinformatics *15*, 174-175.

Rubenstein, J.L., Brutlag, D., and Clayton, D.A. (1977). The mitochondrial DNA of *Drosophila melanogaster* exists in two distinct and stable superhelical forms. Cell *12*, 364-374.

Saavedra, C., Reyero, M.I., and Zouros, E. (1997). Male-dependent doubly uniparental inheritance of mitochondrial DNA and female-dependent sex-ratio in the mussel *Mytilus galloprovincialis*. Genetics *145*, 1073-1082.

Saavedra, C., Stewart, D.T., Stanwood, R.R., and Zouros, E. (1996). Species-specific segregation of gender-associated mitochondrial DNA types in an area where two mussel species (*Mytilus edulis* and *M. trossulus*) hybridize. Genetics *143*, 1359-1367.

Saccone, C., Attimonelli, M., and Sbisà, E. (1987). Structural elements highly preserved during the evolution of the D-loop-containing region in vertebrate mitochondrial DNA. J. Mol. Evol. *26*, 205-211.

Saccone, C., De Giorgi, C., Gissi, C., Pesole, G., and Reyes, A. (1999). Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene *238*, 195-209.

Saccone, C., Gissi, C., Lanave, C., Larizza, A., Pesole, G., and Reyes, A. (2000). Evolution of the mitochondrial genetic system: an overview. Gene *261*, 153-159.

Saccone, C., Gissi, C., Reyes, A., Larizza, A., Sbisà, E., and Pesole, G. (2002). Mitochondrial DNA in metazoa: degree of freedom in a frozen event. Gene *286*, 3-12.

Saccone, C., Pesole, G., and Sbisà, E. (1991). The main regulatory region of mammalian mitochondrial DNA: structure-function model and evolutionary pattern. J. Mol. Evol. *33*, 83-91.

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. *4*, 406-425.

Sastry, A.N. (1979). Pelecypoda (excluding Ostyeidae). In Reproduction of Marine Invertebrates, A.C.Giese and J.S.Peasre, ed. (Academic Press, New York), pp. 113-292.

Satta, Y., Ishiwa, H., and Chigusa, S.I. (1987). Analysis of nucleotide substitutions of mitochondrial DNAs in *Drosophila melanogaster* and its sibling species. Mol. Biol. Evol. *4*, 638-650.

Sbisà, E., Tanzariello, F., Reyes, A., Pesole, G., and Saccone, C. (1997). Mammalian mitochondrial D-loop region structural analysis: identification of new conserved sequences and their functional and evolutionary implications. Gene *205*, 125-140.

Selosse, M., Albert, B., and Godelle, B. (2001). Reducing the genome size of organelles favours gene transfer to the nucleus. Trends Ecol. Evol. *16*, 135-141.

Shadel, G.S. and Clayton, D.A. (1997). Mitochondrial DNA maintenance in vertebrates. Annu. Rev. Biochem. *66*, 409-435.

Shalgi, R., Magnus, A., Jones, R., and Phillips, D.M. (1994). Fate of sperm organelles during early embryogenesis in the rat. Mol. Reprod. Dev. *37*, 264-271.

Shitara, H., Hayashi, J.I., Takahama, S., Kaneda, H., and Yonekawa, H. (1998). Maternal inheritance of mouse mtDNA in interspecific hybrids: segregation of the leaked paternal mtDNA followed by the prevention of subsequent paternal leakage. Genetics *148*, 851-857.

Shitara, H., Kaneda, H., Sato, A., Inoue, K., Ogura, A., Yonekawa, H., and Hayashi, J.I. (2000). Selective and continuous elimination of mitochondria microinjected into mouse eggs from spermatids, but not from liver cells, occurs throughout embryogenesis. Genetics *156*, 1277-1284.

Simon, V.R., Karmon, S.L., and Pon, L.A. (1997). Mitochondrial inheritance: cell cycle and actin cable dependence of polarized mitochondrial movements in *Saccharomyces cerevisiae*. Cell Motil. Cytoskeleton *37*, 199-210.

Skibinski, D.O., Gallagher, C., and Beynon, C.M. (1994a). Mitochondrial DNA inheritance. Nature *368*, 817-818.

Skibinski, D.O., Gallagher, C., and Beynon, C.M. (1994b). Sex-limited mitochondrial DNA transmission in the marine mussel *Mytilus edulis*. Genetics *138*, 801-809.

Soot-Ryen, T. (1955). A report on the family Mytilidae Pelecypoda. Allan Hancock Pacific Expeditions, University of Southern California *20*, 1-153.

Stewart, D.T., Kenchington, E.R., Singh, R.K., and Zouros, E. (1996). Degree of selective constraint as an explanation of the different rates of evolution of gender-specific mitochondrial DNA lineages in the mussel *Mytilus*. Genetics *143*, 1349-1357.

Stewart, D.T., Saavedra, C., Stanwood, R.R., Ball, A.O., and Zouros, E. (1995). Male and female mitochondrial DNA lineages in the blue mussel (*Mytilus edulis*) species group. Mol. Biol. Evol. *12*, 735-747.

Stien, A., Halvorsen, O., and Leinaas, H.P. (1996). Density-dependent sex ratio in *Echinomermella matsi* (Nematoda), a parasite of the sea urchin *Strongylocentrotus droebachiensis*. Parasitology *112*, 105-112.

Stoneking, M. (2000). Hypervariable sites in the mtDNA control region are mutational hotspots. Am. J. Hum. Genet. *67*, 1029-1032.

Stouthamer, R., and Luck, R.F. (1994). Cytogenetics of microbe-associated parthenogenesis and its consequences for gene flow in *Trichogramma* wasps. Heredity *73*, 317-327.

Sutherland, B., Stewart, D., Kenchington, E.R., and Zouros, E. (1998). The fate of paternal mitochondrial DNA in developing female mussels, *Mytilus edulis*: implications for the mechanism of doubly uniparental inheritance of mitochondrial DNA. Genetics *148*, 341-347.

Sutovsky, P., Moreno, R., Ramalho-Santos, J., Dominko, T., Thompson, W.E., and Schatten, G. (2001). A putative, ubiquitin-dependent mechanism for the recognition and elimination of defective spermatozoa in the mammalian epididymis. J. Cell Sci. *114*, 1665-1675.

Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C., and Schatten, G. (1999). Ubiquitin tag for sperm mitochondria. Nature *402*, 371-372.

Sutovsky, P., Moreno, R.D., Ramalho-Santos, J., Dominko, T., Simerly, C., and Schatten, G. (2000). Ubiquitinated sperm mitochondria, selective proteolysis, and the regulation of mitochondrial inheritance in mammalian embryos. Biol. Reprod. *63*, 582-590.

Sutovsky, P., Navara, C.S., and Schatten, G. (1996). Fate of the sperm mitochondria, and the incorporation, conversion, and disassembly of the sperm tail structures during bovine fertilization. Biol. Reprod. *55*, 1195-1205.

Suzuki, H., Hosokawa, Y., Nishikimi, M., and Ozawa, T. (1991). Existence of common homologous elements in the transcriptional regulatory regions of human nuclear genes and mitochondrial gene for the oxidative phosphorylation system. J. Biol. Chem. *266*, 2333-2338.

Szollosi, D. (1965). The fate of sperm middle-piece mitochondria in the rat egg. J. Exp. Zool. *159*, 367-378.

Taanman, J.W. (1999). The mitochondrial genome: structure, transcription, translation and replication. Biochim. Biophys. Acta *1410*, 103-123.

Tavarè, S., Marshall, C.R., Will, O., Soligo, C., and Martin, R.D. (2002). Using the fossil record to estimate the age of the last common ancestor of extant primates. Nature *416*, 726-729.

Terry, R.S., Dunn, A.M., and Smith, J.E. (1997). Cellular distribution of a feminizing microsporidian parasite: a strategy for transovarial transmission. Parasitology *115*, 157-163.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., and Higgins, D.G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. *25*, 4876-4882.

Thompson, P.E., and Bowen, J.S. (1972). Interactions of differentiated primary sex factors in *Chironomus tentans*. Genetics *70*, 491-493.

Thorsness, P.E., and Weber, E.R. (1996). Escape and migration of nucleic acids between chloroplasts, mitochondria, and the nucleus. Int. Rev. Cytol. *165*, 207-234.

Tomita, K., Yokobori, S., Oshima,T., Ueda, T., and Watanabe, K. (2002). The cephalopod *Loligo bleekeri* mitochondrial genome: multiplied noncoding regions and transposition of tRNA genes. J. Mol. Evol. *54*, 486-500.

Turner, D.H., and Sugimoto, N. (1988). RNA structure prediction. Annu. Rev. Biophys. Biophys. Chem. *17*, 167-192.

Ursprung, H., and Schabtach, E. (1965). Fertilization in tunicates: loss of the paternal mitochondrion prior to sperm entry. J. Exp. Zool. *159*, 379-383.

van den Biggelaar, J.A.M., Dictus, W.J.A.G., and Serras, F. (1994). Molluscs. In Embryos, Color Atlas of Development, J.B.L.Bard, ed. (Wolfe Publishing, London), pp. 77-91.

Varvio, S.L., Koehn, R.K., and Väinöiä, R. (1988). Evolutionary genetics of the *Mytilus edulis* complex in the North Atlantic region. Mar. Biol. *98*, 51-60.

Vawter, L., and Brown, W.M. (1993). Rates and patterns of base change in the small subunit ribosomal RNA gene. Genetics *134*, 597-608.

Verdonk, N.H., and van den Biggelaar, J.A.M. (1983). Early development and the formation of the germ layers. In The Mollusca, Vol 3, Development, (Academic press, NY), pp. 91-122.

Vermeij, G.J. (1992). Trans-equatorial connections between biotas in the temperate eastern Atlantic. Mar. Biol. *112*, 343-348.

Wallace, D.C., and Lott, M.T. (2003). MITOMAP: A Human Mitochondrial Genome Database. http://www.mitomap.org.

Watanabe, K.I., Bessho, Y., Kawasaki, M., and Hori, H. (1999). Mitochondrial genes are found on minicircle DNA molecules in the mesozoan animal Dicyema. J. Mol. Biol. *286*, 645-650.

White, M.J.D. (1973). Animal Cytology and Evolution. (Cambridge University Press, Cambridge, U.K.).

Wilson, R.J., and Williamson, D.H. (1997). Extrachromosomal DNA in the apicomplexa. Mol. Biol. Rev. *61*, 1-16.

Wolstenholme, D.R. (1992). Animal mitochondrial DNA: structure and evolution. Int. Rev. Cytol. *141*, 173-216.

Wright, D.D., Ryser, J.J., and Kiltie, R.A. (1995). First -cohort advantage hypothesis: a new twist on facultative sex ratio adjustment. Am. Nat. *145*, 133-145.

Wright, F. (1990). The 'effective number of codons' used in a gene. Gene *87*, 23-29.

Zbawicka, M., Skibinski, D., and Wenne, R. (2003). Doubly uniparental transmission of mitochondrial DNA length variants in the mussel *Mytilus trossulus*. Mar. Biol. *142*, 455-460.

Zouros, E. (2000). The exceptional mitochondrial DNA system of the mussel family Mytilidae. Genes Genet. Syst. *75*, 313-318.

Zouros, E., Ball, A.O., Saavedra, C., and Freeman, K.R. (1994a). Mitochondrial DNA inheritance. Nature *368*, 818.

Zouros, E., Ball, A.O., Saavedra, C., and Freeman, K.R. (1994b). A new type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. Proc. Natl. Acad. Sci. U. S. A. *91*, 7463-7467.

Zouros, E., Freeman, K.R., Ball, A.O., and Pogson, G.H. (1992). Direct evidence for extensive paternal mitochondrial DNA inheritance in the marine mussel *Mytilus*. Nature *359*, 412-414.

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res. *31*, 3406-3415.