Winter 2002

Evolution of gender associated mitochondrial DNA in bivalves

Jason Paul Curole
University of New Hampshire, Durham

INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

Evolution of Gender Associated Mitochondrial DNA in Bivalves

By

Jason P. Curole

B.S., Louisiana State University, 1994

M.S., Louisiana State University, 1997

Dissertation

Submitted to the University of New Hampshire
in Partial Fulfillment of
the Requirements for the Degree of

Doctor of Philosophy

in

Genetics

December, 2002

UMI Number: 3070972

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This dissertation has been examined and approved.

Dissertation Director, Thomas D. Kocher, Professor of Zoology

John Collins, Associate Professor of Biochemistry and Molecular Biology

Marianne Litvaitis, Associate Professor of Zoology

Anita Klein, Associate Professor of Biochemistry and Molecular Biology

Kelley Thomas, Hubbard Chair in Genomics

Jeffrey Todd Streelman, Postdoctoral Fellow in Genomics and Genetic Mechanisms

9/3/2002

Date

For Dad.

"It is those we live with and love and should know who elude us."

Norman MacLean, *A River Runs Through It*

ACKNOWLEDGEMENTS

I must begin by thanking my advisor, Thomas Kocher, for providing a challenging environment that facilitated my development as an independent scientist. I would like to thank the members of my committee for their help and guidance; in particular, I would especially like to thank John Collins, Anita Klein and Marianne Litvaitis for their encouragement during difficult times. Kelley Thomas provided much needed and appreciated support and career advisement during the writing phase. I would also like to thank Todd Streelman for helpful scientific and career advisement that often provoked me to approach problems in a constructive manner. Pat Danley was often available to listen to my thoughts on molecular evolution and was an outstanding friend during out tenure in the Kocher lab. Additionally, I would like to thank the past and present members of the Kocher lab who provided comments and criticisms during my numerous group meeting presentations. This task would not have come to completion without the guidance of Steve Safryn whose counseling also provoked me to approach problems in a constructive manner. Lastly, I must thank my wife Sarah who sailed with me through the many tempests during my tenure at UNH.

FOREWORD

The work described herein involves the unique system of mitochondrial DNA inheritance in bivalves discovered by the labs of David Skibinski and Eleftherios Zouros. In the published papers describing this mode of inheritance Skibinski and colleagues deemed the mode of inheritance, sex-limited mitochondrial DNA inheritance; Zouros and colleagues deemed it doubly-uniparental inheritance. Throughout this dissertation when referring to the mode of inheritance in acronymic form I have used DUI. This is not intended to indicate a preference for either term, but resulting from the observation that DUI is not easily confused with other evolutionary or biological acronyms.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

EVOLUTION OF GENDER ASSOCIATED MITOCHONDRIAL DNA IN BIVALVES
by
Jason P. Curole

University of New Hampshire, December, 2002

Bivalves of the families Unionidae and Mytilidae possess a unique mode of mitochondrial DNA inheritance dubbed sex-limited mitochondrial DNA (mtDNA) inheritance or doubly-uniparental inheritance (DUI). In contrast to the matrilineally restricted inheritance of mtDNA in other metazoans, these bivalves have both matrilineally and patrilineally inherited mtDNA lineages. Although the matrilineally inherited mtDNA (or F type) is present in the somatic tissues of both sexes and the female gonad, the patrilineally inherited mtDNA (or M type) is limited to the male gonad. In *Mytilus* several observations have been made concerning the evolution of DUI. Firstly, the fidelity of DUI is not perfect; many times in the history of these bivalves the F type mtDNA has invaded the M lineage, resulting in a masculinization event. After invasion the two mtDNA lineages begin divergence *de novo*. Secondly, the *Mytilus* M type mtDNA has been shown to evolve more rapidly than the F type and it is hypothesized that this is due to relaxed selection on the M genome. Lastly, the *Mytilus* F type mtDNA has been shown to evolve more rapidly than other metazoan mtDNA and this is also hypothesized to be due to relaxed selection on the F genome. I examined the evolution of DUI in the freshwater family Unionidae and these results indicate that the mechanisms of DUI are different between the two systems. In the Unionidae, masculinizations are not observed resulting in lineages whose divergence dates to over

450 million years ago. This lack of masculinization correlates with an approximately 600 base pair extension of the cytochrome *c* oxidase II gene. As with the mytilids the M type mtDNA evolves more rapidly than the F type mtDNA in the unionids; however, this increase in rate of evolution appears to be largely due to male driven evolution and not relaxed selection. Lastly, the increase in rate of the F type *Mytilus* mtDNA relative to other metazoans is not seen in the unionid lineage; in contrast, the F type mtDNA has a rate of evolution similar to that of other gastropods. This suggests that the rapid rate of evolution of the *Mytilus* F type mtDNA is not due to relaxed selection, but may be due to other mechanistic forces. The disparity of the unionid system relative to the mytilid system and possible future directions for understanding these differences are discussed.

CHAPTER I

SEX-LIMITED MITOCHONDRIAL DNA INHERITANCE IN BIVALVES: A REVIEW

Sex-biased mitochondrial DNA heteroplasmy in *Mytilus*

Mitochondrial DNA (mtDNA) heteroplasmy has been reported for several natural populations (Solignac, Monnerot, and Mounolou 1983; Densmore, Wright, and Brown 1985; Harrison, Rand, and Wheeler 1985; Bermingham, Lamb, and Avise 1986; Wallis 1987); however, heteroplasmy is thought to be a rare aberration of standard maternal inheritance. The possible causes of mtDNA heteroplasmy include balancing selection, mutation and paternal leakage. The conditions necessary to maintain heteroplasmy through balancing selection are very restrictive and the significant reduction in mitochondrial population size prior to gametogenesis make it difficult for rare variants to be temporally maintained (Reviewed in Rand 2001).

On the contrary, breeding studies of hybridization of mice (Gyllensten *et al.* 1991; Kaneda *et al.* 1995; Shitara *et al.* 1998) and *Drosophila* (Kondo *et al.* 1990), as well as studies of hybrid zones (Magoulas and Zouros 1993) have provided evidence for paternal leakage of mtDNA. This phenomenon appears to be more common during hybridization events between moderately divergent species. Despite the introduction of two divergent genomes into a single individual, segregation appears to occur quickly, and maintenance of heteroplasmy beyond a single generation is limited (Shitara *et al.* 1998).

Population studies of the marine mussel *Mytilus* revealed a high level of heteroplasmy with some peculiar characteristics. In a hybrid zone between *Mytilus edulis* and *Mytilus galloprovincialis*, individuals were identified as heteroplasmic for two highly

divergent mitochondrial genomes (Fisher and Skibinski 1990). The more common genome, labeled F, was present in all individuals and gave a very strong signal when hybridized with a *Mytilus edulis* mitochondrial genome probe. The second genome, labeled M, gave a weaker signal when hybridized and was only present in about 20% of the population. In particular, this genome was significantly sex biased, being present in only 5% of females, whereas it was present in 37% of males. In addition, no individuals were homoplasmic for the M genome. Fisher and Skibinski (1990) hypothesized that these genomes might be the result of separate lines of descent; the F genome being passed via the female line of descent and the M genome passed via the male line of descent. Statistical analysis of the data rejected the hypothesis of paternal transmission, as biparental inheritance would result in random combinations of haplotypes, not the sex biased association observed. They concluded that the results were more consistent with maternal heteroplasmy but strongly argued that additional experiments were necessary to test for evidence of paternal transmission (Fisher and Skibinski 1990).

Biparental inheritance was given support with two additional studies of heteroplasmy in *Mytilus*. Sampling 150 individuals from 16 populations of *Mytilus edulis*, Hoeh, Blakely and Brown (Hoeh, Blakley, and Brown 1991) observed a high frequency of heteroplasmy. In this study, 57% of individuals surveyed were heteroplasmic and similar to the previous study the two genomes were highly divergent (estimated in this case to be as high as 20%); however, here the authors observed homoplasmic individuals of all mitotypes. These observations suggested that heteroplasmy is mating dependent, as only highly divergent mitotypes occurred within heteroplasmic individuals. The possibility that such heteroplasmy might be present in

other bivalve taxa was also raised with the evidence that similar mtDNA heteroplasmy was observed in the freshwater bivalve *Anodonta fragilis* (Palaeoheterodonta: Unionidae) (Hoeh, Blakley, and Brown 1991).

Following up on Fisher and Skibinki's (1990) suggestion, Zouros *et al.* (1992) set up a series of pair crosses between *Mytilus edulis* and *Mytilus trossulus*. In homospecific and heterospecific crosses, individuals with a significant paternal contribution of mtDNA were observed. In the 11 families whose offspring exhibited heteroplasmy, the overall rate of paternal contribution was 40% of the total progeny. Paternal leakage was not observed in five families, two of which were heterospecific crosses. Thus, this study suggested that there was a large variance in paternal leakage that did not appear to be significantly correlated with any biological factors, with the exception that amongst the parents the majority of heteroplasmic individuals were male. Unfortunately, Zouros *et al.* (1992) did not sex the progeny and, therefore, could not examine correlation of heteroplasmy and sex in the F_1. This work rejected the hypothesis that this heteroplasmy is female based, as female individuals were mated to different males and in concordance with the paternal transmission hypothesis, heteroplasmic offspring carried both the maternal and respective paternal mitotypes.

Sex limited transmission or doubly-uniparental inheritance

The mystery of sex-biased heteroplasmy in *Mytilus* was solved independently by the Zouros and Skibinski groups (Skibinski, Gallagher, and Beynon 1994a, b; Zouros *et al.* 1994a, b). The Skibinski group (Skibinski, Gallagher and Beynon 1994a, b) used a combination of Southern hybridization and PCR to determine the distribution of the M and F mitotypes within individuals. Slot blots and PCR of female gill and egg DNAs produced strong signals for the F genome, but indicated that the M genome was completely absent. In contrast, slot blots of male gill tissue showed signal for both the F and M genomes, whereas testes showed signal for only the M genome. These results were supported by PCR analysis, with the exception that testes produced weak signal for the F genome.

The Zouros group (Zouros *et al.* 1994a, b) continued with the successful methodology of examining pair crosses using Southern hybridization and PCR, except that for this study the progeny were scored for sex. Indeed, in homospecific crosses all male progeny carried both their mothers and fathers mitotype; whereas, female progeny only carried their mother's mitotype. Thus, conclusive evidence was now available that *Mytilus* exhibits a unique mode of mitochondrial DNA inheritance where sons receive mtDNA from their mothers and from their fathers. The maternal contribution is not transmitted by males; however, the paternal contribution is transmitted by males to their sons. Daughters only receive the maternal contribution and transmit this genome to all of their progeny. Skibinski, Gallagher and Beynon (1994b) labeled this "sex limited transmission"; Zouros *et al.* (1994b) titled it "doubly-uniparental inheritance" (DUI) to emphasize the fact that inheritance was uniparental but involved both parents.

These studies also produced additional observations about DUI. Homospecific crosses were faithful to the maintenance of DUI, whereas in heterospecific crosses, DUI frequently failed. In the *M. edulis* X *M. trossulus* crosses some males completely lacked a paternal contribution, and in one cross a single female was heteroplasmic for paternal and maternal genomes (Zouros *et al.* 1994b). Homoplasmic males and heteroplasmic females were previously observed in natural hybrid zone (Fisher and Skibinski 1990), but these results strongly suggested that breakdown of DUI is a phenomenon specific to hybridization. Families that only produced homoplasmic progeny were also observed in the previous cross (Zouros *et al.* 1992). The newly determined gender phenotypes of the progeny indicate that these families likely consisted strictly of daughters (Zouros *et al.* 1994b). In particular, certain females were shown to produce families that were completely biased towards daughters regardless of the male parent, suggesting that sex ratio in *Mytilus* is controlled by the mother's genotype. At the molecular level, the M genome showed greater genotype diversity than the F genome, indicating that the paternal contribution may be evolving more rapidly than the maternal genome (Skibinski, Gallagher and Beynon 1994b). Lastly, PCR analysis indicated the possibility of some leakage of maternal mtDNA into the paternal transmission route, as the F genome was present in some sperm DNA preparations.

It was then clear that *Mytilus edulis*, *Mytilus trossulus* and *Mytilus galloprovincialis* exhibited this unique mode of mitochondrial DNA inheritance, but the broader phylogenetic distribution of DUI was in question. Previous work indicated that heteroplasmy was also present in the palaeoheterodont *Anodonta fragilis*, indicating that DUI may be phylogenetically widespread and ancient (the Palaeoheterodonta diverged

from other bivalve lineages in the middle Cambrian; see below) (Hoeh, Blakley, and Brown 1991). A re-interpretation of a report of highly divergent haplotypes in a previous study of *M. trossulus and M. galloprovincialis* gave support to the distribution of DUI in these species (Geller 1994). A second report of sex-biased heteroplasmy in *M. galloprovincialis* extended presence of DUI to Atlantic populations of this mussel and offered the first evidence that the origin of DUI might predate the *Mytilus* taxa (Quesada, Skibinski, and Skibinski 1996). A later report extended the phylogenetic distribution to *Mytilus californianus*, and by this time it was clear that DUI was widespread within the mytilid taxa if not within the bivalves.

The first report to confirm the presence of highly divergent sex associated lineages in taxa outside of the Mytilidae came from studies of the palaeoheterodont, *Pyganodon grandis* (Palaeoheterodonta: Unionidae) (Liu, Mitton, and Wu 1996). Females from two populations carried mtDNA haplotypes that were only 0.4% divergent (using whole genome RFLP Southern hybridization). Allozyme analysis at 12 loci supported this level of genetic divergence with estimates of Nei's D equal to 0.055. In contrast, males from each population carried highly divergent haplotypes that were both 6.8% divergent from the female haplotypes of their respective populations and 11.5% divergent from one another. In a second study of unionids, *Geukensia demissa* (Pteriomorpha: Mytilidae) was also shown to carry divergent sex-associated lineages as were two additional species of unionids (*Pyganodon fragilis* and *Fusconia flava*) (Hoeh *et al.* 1996b). More recently the presence of DUI has been extended to a third subclass of bivalves, the Heterodonta, with the discovery of divergent sex-associated lineages in *Tapes* (Passamonti and Scali 2001). The absence of divergent sex-associated lineages

has not been reported for any bivalve taxa, although any bias against reporting negative results may be a hindrance. Also, if breakdown of DUI is frequent, as it may be in organisms such as *C. virginica* where protandry may amalgamate the paternal and maternal genomes, divergence between sex-associated lineages would be retarded.

Functional mechanisms of DUI: a model based on failure

Previous work showed that although DUI is faithfully maintained in homospecific *Mytilus edulis* crosses, the paternal transmission route is unstable in heterospecific crosses between *M. edulis* and *M. trossulus* (Zouros *et al.* 1994b). Additionally, analysis of a hybrid zone in southwest England revealed individuals whose mitotypes did not conform with those predicted by DUI (Fisher and Skibinski 1990). Thus, it is clear that matings between individuals from divergent populations could lead to a breakdown of DUI, as similar matings caused a breakdown of standard maternal inheritance (SMI) in some taxa.

In concordance with previous work, individuals from a *M. galloprovincialis/M. trossulus* hybrid zone in San Francisco Bay exhibited extensive breakdown of DUI. Most individuals homozygous for nuclear paternal genotypes (as assessed by PCR at three loci) carried mitotypes expected under DUI, but mussels with hybrid nuclear genotypes did not exhibit gender-associated mitotypes. In particular, several males were homoplasmic for an F mitotype and there was a high frequency of females heteroplasmic for both F and M mitotypes. Males homoplasmic for the F mitotype and heteroplasmic females were also observed among individuals with paternal genotypes, suggesting that hybrid introgression is extensive and DUI may be disrupted in advanced hybrid individuals.

In contrast, analysis of a *M. edulis/M. galloprovincialis* hybrid zone in Whitsand Bay (UK) indicated a very limited level of hybridization-driven breakdown of DUI (Rawson, Secor, and Hilbish 1996). In this population there is a significant association between sexual phenotype and mtDNA haplotype, regardless of nuclear genotype. Similar results were observed for a *M. edulis/M. trossulus* hybrid zone near Nova Scotia,

indicating a very limited amount of breakdown of DUI between populations of high divergence (Nei's identity ranges from 0.0-0.236 for 5 allozyme loci) (Saavedra *et al.* 1997). Without exception, hybrid females (based on allozyme genotype) were homoplasmic and in a majority of cases carried a mitotype concordant with its nuclear genotype. The only heteroplasmic female was found among the 47 females that could be assigned as pure *M. edulis* based on allozyme genotype, and this female was heteroplasmic for an *edulis* M mitotype. More strikingly, all of the hybrid males were heteroplasmic, in contrast to the large number of homoplasmic males found in the hybrid zone in southwest England. Again, with hybrid males there was a significant correlation between mitotype and nuclear genotype. Four hybrid males carried heterospecific mitotypes (all were *edulis* F/*trossulus* M), but only one of these had an allozyme genotype compatible with an F_1. Analysis of individuals that were categorized as pure *edulis* or pure *trossulus* indicated that introgression of heterospecific mitotypes was non-existent. Thus, it is clear that in these populations introgression is blocked, and hybridization causes little disruption of the DUI mode of inheritance.

These observations led Rawson, Secor and Hilbish (1996) to propose a model for the regulation of DUI. Earlier cytological work established that in *M. edulis* the entire sperm penetrates and enters the egg upon fertilization (Longo and Anderson 1969); this finding and the presence of the M mitotype in some females supports the hypothesis that the sperm enters the egg and is retained regardless of the presumably predetermined sex of the egg. After ingression of the sperm, interactions between the sperm mitochondrial membrane and the egg cytoskeleton determine if the paternal mitochondria enter the germ line or are destroyed. After the male genome is established in testes, it is then

preferentially replicated, becoming the dominant mitotype in the male gonad. Presumably, populations between which hybridization leads to a breakdown of DUI (*e.g.* San Francisco Bay and southwest England) have diverged such that the egg factor(s) responsible for elimination/segregation of paternal mitochondria no longer recognize the associated sperm factor (Rawson, Secor and Hilbish 1996). Alternatively, the fidelity of DUI in some hybrid populations (*e.g.* Whitsand Bay and Nova Scotia) indicates that there is little differentiation at these loci; that geographically adjacent populations (Whitsand Bay and SW England) show such disparate levels of hybridization driven breakdown may related to the ambiguity of taxonomic distinction in the *M. edulis* species complex. This model is also consistent with observations in *Mus* (Kaneda *et al.* 1995; Shitara *et al.* 1998) and *Drosophila* (Pitnick and Karr 1998). In crosses between homospecific individuals, nuclear encoded factors recognize the sperm carcass and eliminate it from the developing embryo. Heterospecific crosses cause a breakdown of this recognition mechanism and allow the paternal mitochondrial contribution to survive in the developing progeny. Thus, breakdown of DUI and SMI follow a similar model, and this suggests that DUI may be a co-option of the SMI sperm mitochondria elimination mechanism.

The population approach to understanding DUI was then followed by a pair of breeding studies that examined paternal transmission in greater detail (Saavedra, Reyero and Zouros 1997; Sutherland *et al.* 1998). In the previous crosses between *M. edulis* and *M. trossulus*, Zouros *et al.* (1994b) observed significant variation between females for progeny sex ratio and little variation in maternally common families sired by different

fathers. Paying careful attention to the sex-ratio bias of a particular mother allowed for a qualitative control over the progeny sex-ratio.

Expanding on this study, Saavedra, Reyero and Zouros (1997) examined sex ratio and fidelity of DUI in pair crosses of *M. galloprovincialis*. Again, these crosses indicated significant inter-dam differences in sex ratio, with the exception of one dam (Saavedra, Reyero, and Zouros 1997). Accordingly, this dam was the only mother to show significant intra-dam sex ratio variance. Conveniently, females could be classified into four categories of sex ratio bias: a strong female bias, a weak female bias, equal sex ratios and a male bias, suggesting that sex ratio may be controlled by a small number of genes. These results also clearly indicated that in *Mytilus* sex ratio is under the control of the mother's genotype.

In a second set of cleverly constructed crosses, Sutherland *et al.* (1998) used this sex ratio bias to examine the survival of the paternal mitochondria in developing larvae. Dams with a strong female bias and no sex ratio bias were crossed with several sires. Larvae were then haplotyped for presence/absence of the M haplotype using a sensitive PCR assay. Larvae tested were then divided into three categories: strong M signal (positive for both EtBr staining and a chemiluminescent assay), weak M signal (only positive for the chemiluminescent assay) and absent M signal (negative for both detection methods). Progeny from families mothered by the dam with a female biased sex ratio showed predominately absent M mitotype patterns with some progeny showing a weak signal. Consistent with DUI, about 50% of progeny from the dam with no sex ratio bias were classified as having a strong M signal.

It was clear that the M genome had been eliminated in these progeny prior to the first sampling point (14 days); therefore, a second set of crosses was undertaken (Sutherland *et al.* 1998). At eighteen hours, all progeny tested were positive for the M mitotype indicating that the sperm mitochondria entered the egg regardless of sex. By 24 hours a significant number of progeny were negative for the M mitotype, indicating that destruction of paternal mitochondria occurs very early in the development of this species, consistent with an early elimination of paternal mitochondria in *Mus* and *Drosophila* (Kaneda *et al.* 1995; Pitnick and Karr 1998; Shitara *et al.* 1998).

Completely surprising and inconsistent with previous results was the male dependent breakdown of DUI in the first homospecific *M. galloprovincialis* cross (Saavedra, Reyero and Zouros 1997). A total of 50 male progeny (out of 239) failed to receive their father's mitotype. In particular, one male whose sperm was M positive produced 44 of these M negative sons, suggesting that DUI may be under the control of the father's genotype. This is supported by the observation that the failure of this sire to pass on his mitotype to his sons was independent of the progeny's mother. The six remaining M negative sons were sired by the only two males with sperm heteroplasmic for the M and F types. Given this coincidence, rare parent-independent M negative sons might be expected to be the result of M negative sperm. Examining the parents of these M negative sons, only one cross was between a sire and dam with different F mitotypes, and the two M negative sons showed no evidence of carrying their father's F mitotype. These results indicate that even in homospecific crosses DUI is subject to both sire-dependent and sire-independent failure.

These data spurred Saavedra, Reyero and Zouros (1997) to expand on the previously proposed model. In *Mytilus*, sex ratio is controlled by the mother's genotype; therefore, concentration of a female-encoded cytoplasmic egg factor (factor Z) is postulated to be responsible for sex determination (Saavedra, Reyero and Zouros 1997). Although sex ratio bias can be divided into roughly three categories, sex determination is not clearly under the control of a single locus. Epistatic or environmental factors may play a role, and much remains to be learned in this system.

Doubly-uniparental inheritance appears to be under the control of the father's genotype, and a male-encoded sperm mitochondrial factor (factor W) is postulated to provide the male mtDNA with a replicative advantage (Saavedra, Reyero and Zouros 1997). Although the data suggest paternal dependent breakdown of DUI, it should be noted that the majority of M negative progeny were from a single spawning of one male. This male was then sacrificed before additional spawning could be undertaken, leaving open the question of whether this breakdown of DUI was spawn specific. Additionally, sons of this sire were not crossed also leaving unanswered the question of whether this DUI negative phenotype is heritable.

Lastly, presence of a threshold concentration of paternal mitochondria (or an associated factor) is hypothesized to cause development of testis (Saavedra, Reyero and Zouros 1997). This is clearly the most contentious aspect of the model. The only evidence in support of this hypothesis is the incomplete correlation between the presence of the M haplotype and testes. Males who are M negative are explained as the product of sperm that were deficient in factor W (Saavedra, Reyero and Zouros 1997); however, this may require a very fine line between the requisite number of paternal mitochondria for

gonadal development and the minimum number that would be experimentally detectable in testis. This model also fails to explain the presence of M positive females, who presumably contain enough M type mitochondria for experimental detection but not enough for testis development.

Despite these inconsistencies, this model is in agreement with most of the observations of DUI in *Mytilus*. Satisfactorily, the model and observations of inheritance of the paternal contribution are also in agreement with results from other systems (Kaneda *et al.* 1995; Pitnick and Karr 1998; Shitara *et al.* 1998). What is lacking is an embryological framework for these observations. Although it is clear that the sperm penetrates the egg at fertilization, the mechanism of segregation of the paternal contribution is not understood. Identification of factor Z would clearly be helpful to this effort, as this factor may be directly responsible in the destruction of the paternal contribution in female embryos.

Evolutionary history of doubly-uniparental inheritance: masculinization but not feminization

While studies were investigating the inheritance patterns of DUI in crosses and hybrid zones, additional work was examining the evolutionary history of the gender-associated lineages. In particular, the phylogenetic distribution of DUI and the age of the male and female lineages were of interest. These studies proceeded primarily through sequence analysis of intragenic regions of cytochrome *c* oxidase I (COI), cytochrome *c* oxidase III (COIII) and the 16S ribosomal subunit (16S).

Phylogenetic analysis of *M. edulis* and *M. trossulus* COIII sequences indicated that the origin of the M and F mitotypes preceded the divergence of these two species (Stewart *et al.* 1995). Previous work had suggested that the M genome might be evolving more rapidly (Skibinski, Gallagher and Beynon 1994b) than the F genome, and the COIII analysis supported this observation (Stewart *et al.* 1995). Branches on the phylogenetic tree of the M lineage were longer than those of the F lineage and the M lineage showed a greater diversity of mitotypes. Separation of sites into synonymous and non-synonymous categories suggested that the M lineage may be under relaxed selection. Although the difference in synonymous distance was not significantly different, the difference in non-synonymous distance was highly significant. Simultaneously published work extended the divergence of the M and F lineages and supported the evidence for the faster rate of the M lineage (Rawson and Hilbish 1995). Again, *trossulus*, *edulis* and *galloprovincialis* M and F 16s mitotypes clustered by gender, with 8.3% divergence between gender associated clades. Thus, the origin of DUI clearly predated the divergence of these three species. By calibrating genetic distance with the divergence of the *M. edulis*/*M.*

galloprovincialis clade from the *M. trossulus* clade, it was estimated that divergence of the M and F mitotypes was approximately 5.3 MYA. Gender rates of evolution were also unequal, with the male lineage evolving about twice as fast as the female lineage.

The first study of DUI in organisms other than *Mytilus* examined the evolutionary history of DUI in *Mytilus*, *Geukensia demissa* and three species of the Unionidae (Hoeh *et al.* 1996a). Systematic analysis of the COI gene indicated multiple origins of the gender associated mitochondrial DNA lineages. The F and M mitotypes of *M. edulis* and *M. trossulus* clustered together by gender, in agreement with the previous COIII and 16s analyses. The *Geukensia* M and F mitotypes also clustered by gender, to the exclusion of heterogeneric M and F mitotypes. This was also the case for the unionid mitotypes, which clustered by gender to the exclusion of other species' mitotypes. This evolutionary history can only be explained by two hypotheses (Hoeh *et al.* 1996b). The first is that there have been multiple origins of the DUI mode of inheritance; however, as amazing as the assumed single co-option of the SMI mode is, multiple co-options are significantly less likely. An alternate hypothesis presented itself in the observation that DUI frequently failed. Under this model a male would fail to pass on his paternal mitotype leaving his sons with only their mother's mitotype to transmit. This newly captured genome would be subject to the wares and fate of genetic drift. A large majority of these captured genomes would become extinct; however, rarely one of these captured genomes would become fixed in the population. The newly captured genome would begin to differentiate from the maternal lineage, but it would still share its most recent ancestry with the female lineage. These events are called "masculinization events" to indicate the masculinizing of a female genome (Hoeh *et al.* 1996a). The capture of the male genome

by the female lineage, or "feminization", was also hypothesized although it had not been observed.

These masculinization events are clearly occurring in the marine mussels (*Mytilus* and *Geukensia*) as indicated by the evolutionary history of the M and F mitotypes, and the relatively low level of divergence between mitotypes. The *Geukensia* mitotypes are about as divergent as the *trossulus/edulis* mitotypes suggesting that these genomes are about the same age. This is not the case with the Unionidae. The two genera sampled (*Pyganodon* and *Fusconia*) diverged over 100 MYA, indicating that the unionid mitotypes are much older than the marine mussel mitotypes. This suggested the possibility of taxon specific difference in the fidelity of DUI (Hoeh *et al.* 1996a).

Conclusive evidence of masculinization was presented in a later study examining several mitotypes of the *Mytilus* group (Hoeh *et al.* 1997). In particular, one of the *M. trossulus* M mitotypes clustered with a female mitotype to the exclusion of the other M mitotype (which clustered with one of the *M. edulis* M mitotypes). This result was mirrored in *M. edulis* where an *edulis* M mitotype clustered with an F mitotype to the exclusion of the other M mitotype. The ancestral *M. edulis* and *M. trossulus* M mitotypes were the sister clade to a clade containing the F and newly masculinized M mitotypes. *Mytilus* populations are harboring both the ancestral and the newly masculinized mitotypes and either mitotype may become fixed in the population. An excellent example of a fixed masculinization event is in European *M. trossulus* (Quesada, Wenne, and Skibinski 1999). Although American *M. edulis* and *M. trossulus* F sequences are highly differentiated, European *M. edulis* and *M. trossulus* show little differentiation between themselves and between the pair and American *M. edulis*. The introgression of

the *M. edulis* F mitotype into European *M. trossulus* appears to have also led to masculinization of the newly introgressed F mitotype and fixation of the masculinized mitotype in European *M. trossulus*.

Examining the evolutionary history of the marine and unionid mitotypes, a total of six masculinization events were inferred, five of which have occurred within the marine mussels (Hoeh *et al.* 1997). Two of these masculinization events are also male lineage captures, one in the ancestor of *G. demissa* and one in the ancestor of the *M. edulis*/*M. trossulus* lineage, not including the male lineage capture in European *M. trossulus* (Quesada, Wenne and Skibinski 1999). Thus, it is likely that the mechanisms of DUI are operating differently in the marine mussels and the Unionidae.

Rates of molecular evolution: faster and fastest

The third area of the study of doubly-uniparental inheritance focused on the earlier findings of different rates of evolution of the M and F genomes (Skibinski, Gallagher and Beynon 1994b, Rawson and Hilbish 1995). Ribosomal DNA (16s) sequences indicated a two-fold increase in rate between the M and F genomes (Rawson and Hilbish 1995), whereas distances based on genomic RFLP of unionid M and F genomes indicated a twenty-fold increase in rate (Liu, Mitton and Wu 1996). Examination of synonymous and non-synonymous substitutions in 321 bp of COIII led Stewart *et al.* (1995) to propose relaxed selection as a possible explanation for this rate increase.

A more thorough examination of this rate difference was undertaken for *M. edulis* and *M. trossulus* (Stewart *et al.* 1996). The COIII data set was extended to 813 bp and 118 bp was sequenced for a region of unassigned function. Analysis of the unassigned function sequence and COIII indicated a two- to nine-fold increase in rate of substitution in the M genome as compared with the F genome. Although the difference in rates at synonymous sites was nearly two-fold (2.105 vs. 1.237), this difference was not significant. On the other hand, the four-fold difference in rate at non-synonymous sites (0.056 vs. 0.014) was significant suggesting that synonymous rates are not different, but non-synonymous rates are. The proportion of fixed to polymorphic substitutions for synonymous and non-synonymous sites were also significantly different as tested with the McDonald-Kreitman test. Lastly, variable sites in the M genome had nearly four times as many amino-acid altering substitutions. Although the authors appeared to favor the hypothesis of relaxed selection, they pointed out that several hypotheses are

consistent with the results (Stewart *et al.* 1996). The M genome could have a smaller effective population size due to skewed sex-ratio, variance in reproductive success of males or the smaller number of mtDNA molecules carried by the sperm and in these cases neutral or nearly neutral mutations would be more likely to be fixed in the M genome.

In contrast, analysis of European *M. edulis*, *M. galloprovincialis* and American *M. galloprovincialis* showed significantly greater haplotype diversity for the M genome at synonymous sites but not at non-synonymous sites (Quesada, Warren, and Skibinski 1998). Also, synonymous and non-synonymous rates of substitution were significantly different for the two genomes in comparisons between European taxa and American *M. trossulus*. The non-synonymous to synonymous rate ratio was almost four times higher for the M genome consistent with the hypothesis of relaxed selection. Also, both Atlantic and Mediterranean populations of *M. galloprovincialis* showed a significant excess of replacement polymorphisms when analyzed with the McDonald-Kreitman test, indicating a departure from neutral evolution. Lastly, in comparisons between American *M. trossulus* and the pooled European lineages, the F genome showed a significant excess of synonymous substitutions, but the M genome appeared to be evolving under neutral expectations.

Thus, the authors argued that the forces influencing replacement and silent substitutions are different depending on the divergence time of the genomes. The findings of a significantly greater synonymous rate of evolution between European taxa and American *M. trossulus* and the greater haplotype diversity of the male genome at synonymous but not non-synonymous sites in some European and American populations

argue in favor of an increased mutation rate of the M genome. The authors point out that a smaller effective population size could also cause the difference in synonymous rates of substitution; however, this hypothesis also predicts a reduced diversity of the M genome relative to the F, which is the opposite of what is observed.

Regardless of the underlying mechanism, it is clear that the male genome is evolving more rapidly than the female genome. This increase in rate was extended to the female genome in a comparison with other metazoan mitochondrial DNA (Hoeh *et al.* 1996b). Analysis of 660 base pairs of the COI gene and 762 bp of the COIII gene indicated that the male and female genomes of *Mytilus edulis* are evolving about twice as fast as most other metazoan mtDNA's, including a gastropod, chiton, several arthropods and several vertebrates (Hoeh *et al.* 1996b). Rate of substitution of *Ascaris suum* mtDNA was nearly as great as *Mytilus*, but was consistently, though not significantly, smaller. Consistent with the relaxed selection hypothesis, the authors proposed a model where there is a "division of labor" (Hoeh *et al.* 1996b) between the lineages. The male genome is under relaxed selection **relative** to the female genome because it only functions in the male germ line; thus, the average twofold greater rate of the male genome. The female genome is also under relaxed selection **relative** to other metazoan mitochondrial DNA because it only functions in two of the three milieu, again leading to a relatively higher rate of evolution.

Limitations of executed work

Almost all of the work that has been done on DUI has been on species from the single genus *Mytilus*. Although this work has been extensive it is still uncertain whether DUI functions in the Unionidae in the same way as in *Mytilus*. This leaves many questions open. In particular, the preliminary results suggest that masculinization may occur with different frequencies in the Unionidae. Also, the available data suggest that the male and female genomes evolve at different rates in the Unionidae, similar to *Mytilus*, but the cause of this difference in the Unionidae is still unclear. Lastly, it is uncertain if the overall increase in rate of evolution for the male and female genome is particular to *Mytilus* or is a common result of doubly-uniparental inheritance.

CHAPTER II

ANCIENT SEX-SPECIFIC EXTENSION OF THE CYTOCHROME C OXIDASE II GENE IN BIVALVES AND THE FIDELITY OF DOUBLY-UNIPARENTAL INHERITANCE

Abstract

Bivalves of the families Mytilidae and Unionidae show a unique mode of mitochondrial DNA inheritance called doubly-uniparental inheritance. In addition to receiving the maternally transmitted mtDNA lineage, males receive a separate mtDNA genome from their fathers. This system is sometimes compromised in that female genomes are occasionally recruited into the male cycle of inheritance. These masculinization events are common in the Mytilidae but have not been reported in the Unionidae. In order to estimate the age of the male and female lineages in the Unionidae and to look for evidence of masculinization we sequenced the junction between the cytochrome *c* oxidase II gene and the cytochrome *c* oxidase I gene. The unionid male and female lineages diverged ~450 MYA. There is no evidence for masculinization during this period, suggesting that there are taxon specific differences in the rate of masculinization. Coincidentally, a 200 codon extension of the COII gene is present in the male genome of the Unionidae and may be responsible for the absence of masculinization.

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. Molecular Biology and Evolution 19:1323-1328.

Introduction

Bivalves of the marine family Mytilidae (Pteriomorpha) and the freshwater family Unionidae (Palaeoheterodonta) show a unique mode of mitochondrial DNA (mtDNA) inheritance called doubly-uniparental inheritance (DUI)(Fisher and Skibinski 1990; Skibinski *et al.* 1994a, b; Zouros *et al.* 1994a, b; Zouros *et al.* 1992; Liu, Mitton and Wu 1996). Mothers pass their mitochondrial genome, the maternal lineage, on to all of their offspring. Fathers pass a copy of their mitochondrial genome, the paternal lineage, on to their sons. These two lineages evolve independently, leading to two divergent mitochondrial genomes present within the same organism. Although the female genome is widespread throughout the somatic tissue of males, the male genome tends to be more concentrated in the gonad.

In mytilids, the fidelity of this system is sometimes compromised. Occasionally, males can be found which lack the paternal genome (Saavedra *et al.* 1997; Zouros *et al.* 1994b). Under these circumstances a maternal genome is recruited into the paternal lineage and the male begins passing this new genome on to his sons (dubbed masculinization to indicate the masculinizing of a female genome; see Hoeh *et al.* 1996a). After masculinization, divergence between the paternal and maternal lineages begins *de novo* (Hoeh *et al.* 1996a; Hoeh *et al.* 1997). Masculinization events are most obvious in *Mytilus*, where some male haplotypes are more closely related to female haplotypes than to other conspecific male haplotypes (Hoeh *et al.* 1997). Hoeh *et al.* (1997) found two cases of such masculinization, one in *M. edulis* and another in *M. trossulus*. Based on these results, any case where male haplotypes group with female haplotypes to the exclusion of other male haplotypes has been taken to indicate a

masculinization event. Masculinized genomes have become fixed at least six times in the evolutionary history of DUI: four within the Mytilidae, one before the origin of *Geukensia* and one before the origin of the Unionidae (Hoeh *et al.* 1997).

In contrast, mitotypes of the two unionid *genera* sampled cluster together by gender (Hoeh *et al.* 1996a). Segregation of the gender-specific mitotypes has apparently remained faithful since the divergence of these genera 100 million years ago (MYA). This observation led Hoeh *et al.* (1997) to raise the question of possible taxonomic differences in the fidelity of DUI.

To better estimate the divergence of the male and female genomes of the Unionidae and to sample for possible masculinization events, we sequenced the junction between the cytochrome *c* oxidase I (COI) gene and the cytochrome *c* oxidase II (COII) gene for both male and female lineages in several species of Unionacea. The COI gene is downstream of the COII gene, resulting in a product that includes 300 base pairs (bp) of the 3′ end of the COII gene and approximately 400 bp of the 5′ end of the COI gene. Our data indicate that this divergence is ancient; we propose a model to account for the absence of masculinization in the Unionidae.

Methods

Sequencing of sex-specific mtDNA.

In the case of paternal sequences, DNA was extracted from male gonads undergoing active spermatogenesis. Maternal sequence was derived from female gonads or from mantle tissue. A primer anchored in the COII gene was designed based on sequence from *Quadrula refulgens* (unpub. data). The primer CO2.2 (5' CAGTGGTATTGGAGGTATGAGTA 3') is designed to anneal to positions 2413-2434 of the *Katharina tunicata* mitochondrial genome. The Folmer HCO primer was used to prime synthesis on the COI gene (Folmer *et al.* 1994). Reactions consisted of 1 X PCR buffer, 2.5mM $MgCl_2$, 0.18mM each dNTP and 0.4 μm each primer. Reactions were cycled at 94° for 15 seconds, 45° for 30 seconds and 72° for 60 seconds for a total of 30 cycles. PCR using DNA isolated from male gonads produced two bands: one the same size as the maternal product and a band of a size unique to males. These unique paternal bands were excised from the gel, cleaned of contaminating agarose and cloned using the Promega TA cloning kit. Three clones were sequenced for each species (GenBank Accession # AF517636-AF517643). PCR products were sequenced using the Amersham cycle sequencing kit as per the manufacturer instructions. Sequencing reactions were visualized using an ABI 377 fluorescent sequencer.

Sequences for *Margaritifera hembeli*, *Quadrula refulgens*, *Quadrula quadrula* and *Anodonta implicata* were obtained using the above methods. Amplifications of *M. hembeli* male gonadal template DNA failed to produce a product different from the female sequence obtained from mantle tissue. This could be due to failure to prime *M.*

hembeli male sequence, rearrangements in the *M. hembeli* male genome disrupting the adjacency of the COII and COI genes or *M. hembeli* not exhibiting DUI. Additionally, amplifications from *A. implicata* mantle DNA failed (possibly due to poor quality of the tissue) and unlike the other species sequenced, amplifications from male gonadal tissue failed to produce both a male and female product. Again, failure to amplify this genome may be caused by failure of primer pairs or re-arrangements within the *A. implicata* female genome. Sequences for *Katharina tunicata* and *Crassostrea gigas* were obtained from GenBank (Acc.# U09810 and AF177226). Amino acid sequences were aligned using Clustal X (Thompson *et al.* 1997), and gaps were then inserted into the nucleotide sequences at the corresponding sites. This data matrix was imported into PAUP 4.0b10 (Swofford 2000) and used for phylogenetic analysis. Distances were estimated from the 1st and 2nd codon positions of the 285 nucleotides of the male/female homologous region of COII. The Neighbor-Joining algorithm and maximum likelihood estimation were used for tree construction. The Tamura-Nei model was used to correct for multiple substitutions for both distance and maximum likelihood estimation. The data set was bootstrapped (1000 replicates for NJ and 100 replicates for ML) to determine statistical support. Rates of synonymous and non-synonymous substitution were estimated using the modified Nei-Gojobori method of *MEGA* (Kumar *et al.* 2001).

Detection of COII mRNA

RNA was extracted from *A. implicata* male gonad using Trizol. The RT reaction was primed using a poly-T primer with an 18 bp unique sequence on the 5' end of the poly-T (Hisatomi *et al.* 1996). A separate control reaction was set up including all ingredients except the reverse transcriptase. The RT reaction was incubated for 1hr at

42°C and then at 70°C for 5 minutes to inactivate the enzyme (Frohman, Dush, and Martin 1988; Ohara, Dorit, and Gilbert 1989). PCR was performed using a primer that annealed to the unique sequence of the poly-T primer and a primer that was at the 3' end of the homologous region of COII (CO2.8, 5' AATCATTCGTTTATGCC 3'). Only the reverse transcriptase treatment produced a visible product when run out on an agarose gel. This band was excised, cleaned of agarose and cloned using the Promega TA cloning kit. Three clones were sequenced; all showed polyadenylation of the complete COII product.

Results

Phylogenetic analysis of the male/female homologous region of COII indicates an ancient divergence of these genomes (Fig. 2.1). Reciprocal monophyly of the male and female lineages is strongly supported by bootstrapping for both neighbor-joining and maximum likelihood analyses. This divergence was previously dated to 100 MYA (Hoeh *et al.* 1996a). To better estimate the age of the divergence we analyzed the female sequence from *Margaritifera hembeli* (Palaeoheterodonta: Margaritiferidae) and *Crassostrea gigas* (Pteriomorphia: Ostreidae). The inclusion of *M. hembeli* with the female lineage indicates that the male/female divergence preceded the divergence of the Margaritiferidae and Unionidae and places a minimum bound of 213 MYA on this divergence (Haas 1969) (Fig. 2.1). *C. gigas* branches off before the divergence of the male/female lineages and therefore places a maximum bound of approximately 500 MYA on the male/female divergence (The Palaeoheterodonta diverged from other bivalve lineages in the middle Cambrian) (Newell 1969). By calibrating divergence estimates with the fossil record, we estimate that this male/female divergence occurred more than 450 MYA.

PCR amplification of the COII-COI junction also revealed an approximately 600 bp insertion within the paternal genome that is absent from the maternal genome. We have found this sex-specific polymorphism in multiple individuals of the same species and in ten additional species (*Lampsilis claibornensis*, *Lampsilis radiata*, *Amblema plicata*, *Obliquaria reflexa*, *Plectomerus dombeyanus*, *Glebula rotundata*, *Pyganodon grandis*, *Villosa lineosa*, *Utterbackia imbecillis* and *Fusconaia flava)* where the male genome could be amplified by the PCR. There is the possibility that the sequences

identified could be highly divergent nuclear pseudogenes of the female sequence or nuclear pseudogenes of male sequence; however, male specific length products were only amplified from male gonadal tissue and were not present in female gonadal or mantle tissue. The extra male sequence is a 200 codon extension of the COII gene (Fig. 2.2). In the four taxa sequenced, the reading frame of the COII gene extends through this region and terminates with a stop codon just prior to the beginning of the COI gene. Comparison of the extension between the two most closely related male sequences reveals patterns of evolution consistent with the homologous protein coding region of COII. In both regions, the rate of substitution at 3rd codon positions is almost ten times that at 1st positions and over twenty times the rate at 2nd codon positions (Table 2.1a). The rate of synonymous substitutions is more approximately ten times greater than non-synonymous substitutions (Table 2.1a).

Alignments of the male specific extension indicate that the extension is evolving more rapidly than the male/female homologous region (Fig. 2.2). Distance estimates between male genomes for the male/female homologous region range from 1.6% to 15%; however, in the extension, these estimates range from 4%-108% (Table 2.1b). Blast analysis of the extension reveals no significant similarity with any sequences in GenBank.

To verify that this extension is expressed as a part of COII, we used RT-PCR to amplify polyadenylated COII mRNA from *Anodonta implicata*. The use of a poly-T primer was necessary to verify that this extension was processed as part of the COII gene because the mitochondrial genome is transcribed as a single strand. Using a primer anchored in the homologous region of COII and a unique primer sequence anchored to

the 5' end of the poly-T, RT-PCR produced a product of the expected size. Cloning and sequencing of the product revealed a polyadenylated sequence that was identical to the extended sequence and terminated with the termination codon (TAG) shown in figure 2.

Discussion

Previous studies of DUI in the Unionidae have failed to find masculinization events, which contrasts with the marine mussels where masculinization events occur with measurable frequency. In their survey of marine and freshwater bivalves, Hoeh *et al.* (1997) found five masculinization events within the marine bivalves, yet none in the freshwater bivalves. Our results bring the number of unionid genera sampled to five and the total number of species sampled to seven. This is over three times as many species as sampled from the single genus *Mytilus*, in which there have been numerous masculinization events. In concordance with this, the unionid mitotypes are two orders of magnitude older than the mytilid mitotypes. Hoeh *et al.* (1996a) set a minimum divergence for these mitotypes at 100 MYA. The two families represented in our study, the Unionidae and Margaritiferidae, diverged 213 MYA in the Triassic, establishing this as a minimum bound on the male/female divergence (Haas 1969). Calibrating this divergence with the fossil record and then applying that calibration to the male/female split, we arrive at an estimate of 450 MYA for the actual divergence of these two lineages. This places the divergence of these lineages close to the origin of the Palaeoheterodonta in the middle Cambrian; thus, it is likely that these lineages have remained separate throughout the evolution of the Unionacea. In stark contrast, the most common haplotypes in the *M. edulis* species complex, as well as *Geukensia*, are estimated to have diverged only 5.3 MYA (Hoeh *et al.* 1996a; Rawson and Hilbish 1995). The faithful maintenance of these separate lineages in the freshwater bivalves over 450 million years and the frequent breakdown of DUI in the marine bivalves leads us to conclude that the fidelity of DUI differs among taxa. We hypothesize that there are

mechanisms preventing masculinization in the Unionidae, rather than mechanisms promoting masculinization in the mytilids.

The absence of masculinization in the Unionidae correlates with the presence of a male-specific 200 codon extension of the COII gene. It is possible that this extension has no function and is simply extraneous DNA, but we reject this hypothesis on the following grounds. In all four taxa sequenced the extension is in frame with the COII gene, and stop codons are only present near the initiation codon of the COI gene. As is common in protein coding genes of the mitochondria, positional rates of substitution fall in the $3^{rd}>>1^{st}>2^{nd}$ pattern, and synonymous substitutions are an order of magnitude greater than non-synonymous substitutions. Lastly, we detected in male gonad a polyadenylated mRNA transcript of the COII gene that includes the extension. Based on these observations, we conclude that this extension is protein coding and functional.

The mitochondrial genome of metazoans is often very compact and devoid of any extraneous DNA. Although the genome can vary significantly in size, this is generally the result of duplications involving the D-loop (Brown *et al.* 1996; Gach and Brown 1997) rather than significant extensions of unique sequence. Alignments with *Katharina*, *Crassostrea*, and *Loligo* indicate that the length of the unionid female COII gene is similar to these COII genes, which excludes the possibility that there has been a contraction of the unionid female COII gene. Also, there has been ample opportunity for loss of the protein coding nature of the extension given the high rate of evolution in this region of the gene (Table 2.1). The conservation of this extension for over a hundred million years argues in favor of functionality of this extension.

Elucidation of the function(s) of the extension may help us to understand the evolution of DUI. Hypotheses about function can be divided into three broad categories: a role within the male mitochondrion itself (within mitochondria), a direct role in the survival of the male mitochondrion (within cells) and a function outside of the male mitochondrion (within organisms). These hypotheses and how each might lead to an absence of masculinization are outlined below.

The function of this extension could be within the sperm mitochondrion itself. This must be considered the current null hypothesis because of the normal localization of COII in the inner mitochondrial membrane. The possible functions of this extension in the mitochondrion are numerous, including the obvious function of mitochondrial respiration. However, the rate of evolution of this extension is substantially greater than the rate of evolution of the homologous region of the COII gene. If the function of the extension is limited to within the mitochondrion, the extension could be the direct cause of the absence of masculinization only if it provides a selective advantage over the female genome at some part of the life cycle. It is important to note that in this case the extension functions within the male mitochondrion, but selection at various levels for the presence of the extension could be maintaining the fidelity of DUI.

This extension might be directly involved in intracellular interactions determining the survival of the male mitochondrion. In other organisms, including bivalves, it has been shown that upon fertilization the sperm-derived mitochondria are targeted for elimination from the embryo (Kaneda *et al.* 1995; Pitnick and Karr 1998; Sutherland *et al.* 1998). In *Mytilus* this destruction is sex-specific; the sperm derived mitochondria survive in embryos destined to be males, whereas in females these mitochondria are

destroyed. Elimination of sperm derived mitochondria appears to involve nuclear encoded genes (Kaneda *et al.* 1995; Sutherland *et al.* 1998). The COII extension could be involved in blocking elimination to ensure survival of the male mitochondrion. Alternatively, the extension could play a role in the segregation of male mitochondria to the gonad. In either case, it should be possible to detect the protein product of the extension outside of the inner mitochondrial membrane. The protein product should also directly interact with the factors responsible for any of these processes. Again, a selective advantage is necessary for the extension to be the direct cause of the absence of masculinization.

Finally, the extension might function outside of the inner mitochondrial membrane in a way not directly related to survival of the mitochondrion. In this case, expression again might be observable outside of the inner mitochondrial membrane. However, the protein product could be involved in a number of other pathways, including sperm structure, species-specific recognition or gonadal development (Saavedra *et al.* 1997). Mitochondrially encoded genes are present in the cytoplasm of *Drosophila* (Amikura *et al.* 2001), *Xenopus* (Kashikawa, Amikura, and Kobayashi 2001), sea urchin (Ogawa *et al.* 1999), planarian (Sato *et al.* 2001) and ascidian (Oka *et al.* 1999) embryos. In *Drosophila*, mitochondrial ribosomes are present in the germ plasm and are necessary for proper germ line formation (Iida and Kobayashi 1998). Interestingly, if the extension plays a role in male gonadal development, fitness differences between genomes are not required to maintain the fidelity of DUI.

These hypotheses all require some type of functional inequality between the male and female genomes. Testing this hypothesis will require elucidation of the function of

this extension and quantification of male fitness with and without the male mitochondrial genome. We have not observed any male haplotype negative individuals in our surveys of natural populations; however, breakdown of DUI has been observed in pair crosses of *Mytilus* (Saavedra *et al.* 1997; Zouros *et al.* 1994b). Similar pair crosses within the Unionidae may be successful in generating the intermediate male phenotypes required for masculinization. Until these phenotypes are available, identification of the temporal and spatial expression of the extension will be essential.

Acknowledgements

We would like to thank Todd Streelman, Karen Carleton, Patrick Danley, Masahiro Okazaki and Eleftherios Zouros for comments on a precursor manuscript and A. Eyre-Walker and two anonymous referees for comments on this manuscript. JPC would like to thank Sarah Curole for support. This study was funded by a grant from the Conchologists of America.

Table 2.1. Genetic distances for male and female sequences.

(a)

	1st	2nd	3rd	K_S	K_A
Homologous	0.033	0.000	0.380	0.421	0.015
Extension	0.059	0.022	0.520	0.352	0.038

(b)

	L.t.M[a]	Q.q.M	Q.r.M	A.i.M
L.t.M		0.115	0.114	0.127
Q.q.M	0.691		0.016	0.154
Q.r.M	0.688	0.041		0.153
A.i.M	1.084	0.713	0.698	

(a) First (1st), second (2nd), third (3rd) codon position and synonymous (K_S) and non-synonymous (K_A) distances for *Q. refulgens* and *Q. quadrula* for the extension and the homologous region, (b) distances (1st and 2nd positions only) for male/female homologous region (above diagonal) and the male extension (below diagonal). A Tamura-Nei correction was applied to distance estimates except K_S & K_A where a Jukes-Cantor correction was applied.

[a]M-male: taxon names are abbreviated to generic/specific initials.

Figure 2.1. Neighbor-joining (a) and maximum likelihood (b) phylogenetic reconstruction of male and female lineages. Genetic distance for each branch is shown above the branches and bootstrap support is shown below the branches. Distances smaller than 1% are not shown. Both NJ and ML bootstrap support for the ancient separation of the male and female lineages is very strong and rejects the hypothesis of masculinization within these unionids. *Katharina tunicata* is used as an outgroup.

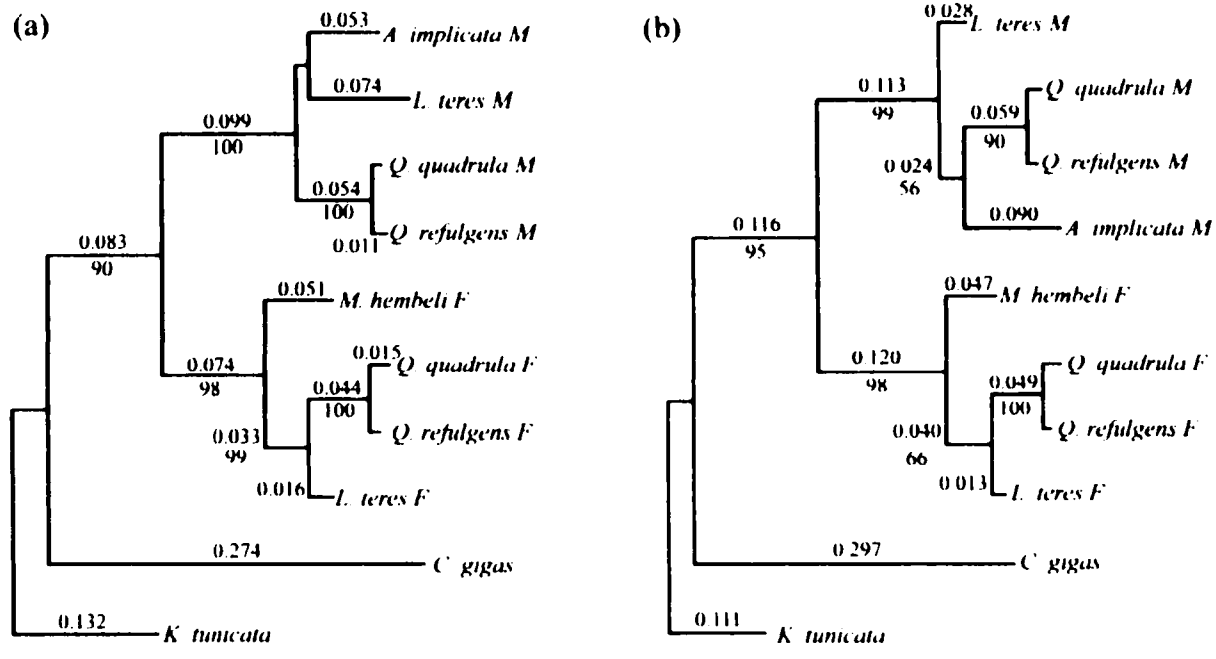


Figure 2.2. Alignment of the putative amino acid sequence of the 3' end of COII. M indicates male lineage and F indicates female lineage. Sequences end with the termination codon (symbolized by open circle). The extension sequences are highly divergent with a few areas of conserved amino acid identity. *, : and . indicate fully conserved, strong group conserved and weak group conserved sites respectively. After the female termination codon these symbols reflect male sequence identity only. Strong and weak groups are defined in (Thompson *et al.* 1997).

```
L. teres F      DGYRLLEVDNRCVLPYGVDSRVLVSSADVIHAWALPSIGVKVDAVPGRINQIGVHLMSSGVLYGQCSEICGVNHSFMPIGLESV
M. hembeli F    GGYRLLEVDNRCVLPYGVDSRVLVSSADVVHAWALPAVGVKADAVPGRINQLAIHLASSGVLYGQCSEICGVNHSFMPIGLEGV
Q. quadrula F   GGYRLLEVDNRCVIPYGVDSRILVSSADVIHAWALPSIGVKVDAIPGRINQLGVHLIGSGVMFGQCSEICGVNHSFMPVGLESV
Q. refulgens F  GGYRLLEVDNRCVVPYGVDSRVLISSADVIHAWALPSIGVKVDAIPGRINQLGVHLMGSGVMFGQCSEICGVNHSFMPIGLESV
Q. quadrula M   GYYRLLEVDNRCIVPSLLQMRGLVTSDDVIHSWAIPSSSVKVDGVPGRINQVGLCFIYSGVFYGQCSELCGVNHSFMPVCVEAV
Q. refulgens M  GHYRLLEVDNRCIVPSLLQMRGLVTSDDVIHSWAIPSSSIKVDGVPGRINQVGLCFIYSGVFYGQCSELCGVNHSFMPVCVEAV
L. teres M      GDYRLLEVDNRCVVSCLLQMRGLVTSDDVVHSWAIPSASIKADGVPGRINQVSLCFLYSGVFYGQCSELCGVNHSFMPVCVEAV
A. implicata M  GSYRLLEVNNRCVVAALLHMRALVTSADVVHSWAIPSACIKVDGIPGRINQVGLCFIYTGVFYGQCSELCGVNHSFMPVCVEAV
                . ******:***::.  :. * *::* **:*:**:*:  :*.*.:******:.: :  :**::*****:*********: :*.*

L. teres F      SPKVFYYWLISo
M. hembeli F    SPGVFYHWLMKo
Q. quadrula F   SPEVFYCWLVRo
Q. refulgens F  SPEVFYCWLVRo
Q. quadrula M   STKVFLNWIFENHSKDVNNS--GVVDSANS--FSLRG-FLMGVFKKIVKVLKMLGSLYIMWFYYVLYYGLYVPAKFAVFGGCDL
Q. refulgens M  STKVFLNWIFENHSKDVNNS--GVVDGVGG--FSLRG-FLMGVFKKVVKVLKMLGSLYVMWFYYVLYYGLYVPAKFAVFGGCDL
L. teres M      SSKVPSEWIMSNHDPNINAS--GSSKNKDRSYLSMIGDVICWVFGNLYSGTWYMIKLYLMWWCFFFKVAIYSPTKFLVESAFNL
A. implicata M  SVEVFTDWIIGNHKLNTNNNNNGWMDWFFSG VGLFTKYFLMGFGCLYEVMWKLGSMYVWWWEMPLHYGIYVPVKYTLSTSVDL
                *  **  *:. **. : * .  *  .       .:.    :  *  :       : .:*: *:  ..  .:* *.*: :  . :*

Q. quadrula M   IQWTLKSCLAIAEWMWWFLFSPVDASIFAFSYLVGKVSSGLWFVVTSPVTAVVWLAKGVWKGVCAIVWFP-    LTAFEAWFDS
Q. refulgens M  IQWALKSCLAVAEWMWWFLFSPVDASIFAFGYLVGKVSSGLWFVVTSPVTAFVWLVKGVWSGVCAIVWFP-    LTAFEAWFDS
L. teres M      SSWVLSVSYSVLYWLGWFVSDPFDASVSGIFWIGGKAFSIIRFCVSSPVTVVVWFTKQVWSVSCFVANIP-    FVIFNAWVDS
A. implicata M  IKWAIVKCVSIVSWVGWFMVSPVDASTYAVVYVSRQVCELTLFAITKPYEPSFWVVKSTYKVVCKTIYFVSQTAVFLFNTMMMS
                .*.:  . ::  *: **:  *.***  .. ::  :. .  * ::.*   .*..*  :.  *   :      .  *:: . *

Q. quadrula M   MSSFTDNDTKNLVVWHIYRNTKEFVWALMERYKD-o
Q. refulgens M  MSSFTDNDTKNMVVWHIYRNTKEFVWALMERYKD-o
L. teres M      VSSPSDNESKQWVVEQVARSSEVFYKAMMEYYSKKo
A. implicata M  MSSFTDDSFKTEVMERINQNTRKFLWIIENYYKGRo
                :***:*:. *  *: :: :.:. *   : : *.
```

CHAPTER III

EVIDENCE FOR MALE-DRIVEN EVOLUTION IN THE MITOCHONDRIAL GENOMES OF THE UNIONIDAE (BIVALVIA: PALAEOHETERODONTA: UNIONIDAE)

Abstract

Bivalves of the families Unionidae and Mytilidae exhibit a unique mode of mitochondrial DNA inheritance called doubly-uniparental inheritance (DUI). An interesting consequence of these two mtDNA lineages is their independence, in particular with respect to rates of evolution. In both families it has been observed that the M type genome (or paternally transmitted) evolves at a faster rate than the F type (or maternally transmitted). In *Mytilus* this difference in rate has been attributed to relaxed selection on the M genome. The cause of the rate difference in unionids is unknown; however, it is not unreasonable to consider that similar mechanisms may be responsible in both systems. We examined 360 bp of sequence from the 5' end of the cytochrome *c* oxidase I (COI) gene and 218 bp of sequence from the 3' end of the cytochrome *c* oxidase II (COII) gene from four species (two phylogenetically independent pairwise comparisons) using approximate and likelihood methods to estimate rates of synonymous (d_S) and non-synonymous (d_N) substitution. Approximate estimates of d_S for the COI locus were similar for the M and F type sequences, but were greater for the M type sequences at the COII locus. In contrast, approximate estimates of d_N were not significantly different for the M and F type sequences at the COII locus, but were significantly different at the COI locus. Likelihood estimates followed a similar pattern, but were only significant for the

COI d_N comparisons. We also examined intra-genic variation in the ratio of d_N/d_S (ω) using likelihood methods. Generally sites could be divided into two ω classes, a fixed or nearly fixed class ($\omega = 0$) and a variable class ($1 > \omega > 0$). The F type COI sequences had the largest proportion of sites in the fixed class and the other sequences had from 65-85% of sites in this class. The variable class exhibited varying ω rate ratios, ranging from 0.085 to 0.95. Finally, the fixed class for the M type COII sequences had a much higher ω rate ratio ($\omega = 0.0042$) than any of the other sequences ($\omega = 0.000001$). These results suggest that male driven evolution is responsible for the large differences in rates of evolution, but that selection levels may vary between the two gender type genomes in subtle ways.

Introduction

The relative contribution of forces affecting rates of evolution at the nucleotide and amino acid levels has been a contentious issue for many years (Li 1997. Neo-Darwinian theory predicts that differences in rate at the molecular level are driven by differing selective pressures on sequences. The neutral theory predicts that differences are due to neutral forces, such as effective metabolic rate or generation-time.

Past efforts to examine rate variation have focused on pairwise comparisons between taxa using the relative rate test (Li 1997). Observed differences in rate have been ascribed to a few different neutral mechanisms. Rate differences between organisms with different generation times are hypothesized to be the result of a greater number of meiotic divisions in a short generation organisms during any given time period. Differences due to generation time have been observed in comparisons of gophers and their parasitic lice (Hafner *et al.* 1994), between rodents and primates (Li *et al.* 1996), primates and sharks (Martin, Naylor, and Palumbi 1992), and turtles and primates (Avise *et al.* 1992). Differences in rate of substitution have also been observed between Y-linked loci and X or autosomal homologues in humans (Shimmin, Chang, and Li 1993; but see Bohossian, Skaletsky, and Page 2000; Makova and Li 2002). Martin and Palumbi (1993) observed that generation time accounted for a proportion of the variance in differences in rate and showed that remaining variation is accounted for by differences in metabolic rate. The metabolic rate hypothesis predicts that organisms with a slower metabolic rate will exhibit slower rates of nucleotide substitution due to lower levels of oxidative damage of mtDNA. More recently there has been great interest in

selection as a cause of rate variation; in particular, use of tests such as the McDonald-Kreitman test has found favor (Rand 2001).

Bivalves of the families Unionidae and Mytilidae exhibit a unique mode of mitochondrial DNA (mtDNA) inheritance. Sons and daughters both receive a mitochondrial genome from their mother, while sons receive an additional mitochondrial genome from their father. These two genomes evolve independently and presumably diverge from conspecific genomes at the time of organismal speciation. A direct comparison can be made between rates of evolution of the two genomes through phylogenetically independent comparisons, without the need for fossil divergence estimates or the use of the relative rate test. Thus, the dual mitochondrial genomes of the doubly-uniparental inheritance system (DUI) are an ideal system to examine the causal mechanisms of differences in rates of evolution.

Differences in rates of evolution of the male and female lineages have been observed in *Mytilus* and *Pyganodon* (Stewart *et al.* 1995; Hoeh *et al.* 1996b; Stewart *et al.* 1996; Liu, Mitton and Wu 1996; Quesada *et al.* 1999). In *Mytilus* the causal mechanisms of rate differences have been examined. A relative increase in the rate of evolution of the male lineage relative to the female lineage was observed in comparisons of *M. edulis* and *M. trossolus* for the cytochrome *c* oxidase III (COIII) locus and for a small region of nucleotide sequence of unassigned function (Stewart *et al.* 1996). The difference in rate at the COIII locus is attributed to relaxed selection of the male genome based on greater non-synonymous to synonymous substitution ratios in the male sequence, a greater polymorphic to fixed ratio for synonymous substitutions and a higher frequency of non-synonymous substitutions of sites deemed "variable" by the authors

(Stewart *et al.* 1996). This increase in rate was extended to the female genome relative to other metazoans in a comparison of a combined cytochrome oxidase I (COI) and COIII data set that indicated that the *Mytilus* mitochondrial genomes show an approximately two-fold increase in rate relative to other metazoan taxa (Hoeh *et al.* 1996b).

In *Pyganodon* the causes of rate differences are not as well understood. It has been shown that the rate of evolution of the male lineage exceeds that of both the female lineage and nuclear allozymes (Liu, Mitton and Wu 1996), but the cause of this difference is unclear. The large number of species of Unionoida (Williams *et al.* 1992) allows for a rich and varied data set of phylogenetically independent pairwise comparisons to estimate the rates of synonymous and non-synonymous substitutions for the M and F lineages. There are currently 281 recognized species of freshwater Unionidae in 32 genera (Williams *et al.* 1992), and many genera consist of ten or more species. In addition, unionid systematics are poorly understood and the possibility that there are cryptic species suggests that additional comparisons may be available. We examined pairwise rate differences in four unionid taxa to test the hypothesis that the same forces causing rate differences in *Mytilus* are responsible for the observed rate differences in the Unionidae.

Methods

Data Collection

Animals were collected from the Amite River basin near Port Vincent, LA and genomic DNA was extracted using a modified CTAB protocol. PCR was performed on extracted DNA using the CO2.2 primer or the male specific COI.13 primer and the Folmer HCO primer. Reactions consisted of 1 X PCR buffer, 2.5mM $MgCl_2$, 0.18mM each dNTP and 0.4μm each primer. Reactions were cycled at 94° for 15 seconds, 45° for 30 seconds and 72° for 60 seconds for a total of 30 cycles. In the case of paternal sequences DNA extracted from male gonads undergoing active spermatogenesis were used. Maternal sequence was derived from female gonads or from mantle tissue. Paternal genome PCR using the CO2.2 primer produced 2 bands (see Chapter 2). The paternal PCR product was gel excised, cleaned of contaminating agarose and cloned using the Promega TA cloning kit. Three clones were sequenced for each species. Paternal genome PCR using the COI.13 primer produced only a single paternal PCR product due to the male specific nature of the primer. These fragments and the maternal products were gel excised, cleaned of contaminating agarose and sequenced directly. Clones and PCR products were sequenced using the Amersham cycle sequencing kit at 1/2 dilution and visualized on an ABI 377 automated fluorescent sequencer.

Data Analysis

Sequences were translated into putative amino acid sequences and then aligned using Clustal X. Gaps were then inserted into the appropriate areas of the nucleotide sequences. The effective number of codons (ENC) (Wright 1990) was calculated using

codonW v1.3 on the Mac OS X Unix BSD variant. ENC is analogous to the effective number of alleles in a population and measures codon usage independent of adaptation indices. ENC ranges from a maximum of 61 (random codon usage) to a minimum of 20 (one codon for each peptide).

Sequences were imported into the program MEGA (Kumar *et al.* 2001) and pairwise synonymous and non-synonymous Jukes-Cantor corrected distances were estimated. The Nei-Gojobori method was used to estimate the number of synonymous and non-synonymous sites. Maximum likelihood pairwise analysis was done using the Phylogenetic Analysis using Maximum Likelihood (PAML) software package (Yang 1997) on the Mac OS X Unix BSD variant. PAML estimates pairwise rates of synonymous and non-synonymous substitutions based on the methods of Yang and Goldman (1994). Pairwise distance comparisons used the estimates and standard errors generated by method 1 of the Yang and Goldman (1994) method, although there was often no difference between method 1 and method 2 estimates. Distances between phylogentically independent taxa were subjected to a z-test to determine the level at which distances were significantly different (Berry and Lindgren 1996).

PAML also estimates within locus variation of non-synonymous/synonymous (ω) ratios based on several models (Goldman and Yang 1994; Nielsen and Yang 1998; Yang *et al.* 2000). Only a portion of the models available were used for analysis as more complex models did not result in better likelihood scores. The one-ratio model assumes a single ω ratio for all codons (Goldman and Yang 1994). The neutral model assumes a portion of sites with an $\omega = 0$ and the remaining sites with an $\omega = 1$. The selection model includes an additional class of sites with ω estimated from the data. Lastly, the discrete

model estimates ω ratios for K specified classes based on an unconstrained discrete model. Initial analysis indicated that only two classes were necessary to adequately explain the variation in ω in all of the data sets analyzed and only these results are presented.

Results

Sequence Alignment

The 3' end of COII is moderately conserved at the amino acid level such that the alignment is unambiguous. A three nucleotide indel (a single amino acid) is present near the 5' end of the alignment; the location of this indel is unambiguous as the amino acids flanking the indel are conserved. The extension at the 3' end of the COII gene causes some ambiguity of homology in this region. The male sequences were truncated at the last amino acid that aligned with the final female peptide. This ambiguity in alignment should not affect the results as all comparisons are within sex and not between. The high level of amino acid conservation at the 5' end of COI makes this alignment unambiguous as well.

Codon Usage

Codon usage was not random for any of the eight sequences examined (Table 3.1). The ENC ranged from a high of 43 to a low of 31.6 at the COI locus. The two *Quadrula* F type sequences have nearly identical measures of ENC. The *Q. refulgens* M type also has a similar measure of ENC, but the *Q. quadrula* M type sequence has slightly more codon bias. The two *Lampsilis* F type sequences have very different levels of codon bias. The *L. claibornensis* F and M type sequence have the most random codon usage and are higher than most of the other sequences. The *L. teres* M and F type sequences are similar to the *Quadrula* sequences in their level of codon bias.

Codon bias was nearly identical at the COII locus for the *Quadrula* F type sequences (Table 3.1). Overall the range of ENC values was smaller than at COI (31.26-36.62). In contrast to the large difference in codon usage at the COI locus for the

Lampsilis F type sequences, ENC was very similar for the two sequences at the COII locus. This is also true of the *Lampsilis* and *Quadrula* M type sequences, which had nearly identical ENC for COII.

Pairwise Comparisons

The Nei-Gojobori (NG) Jukes-Cantor corrected pairwise comparisons of male and female COI and COII sequences produced contradictory results. Synonymous distances for the 360 bp of the 5' end of COI indicated that rates of synonymous substitutions are similar in the male and female genomes (Table 3.2; Figure 3.1). Synonymous distances for the two sets of taxa compared were around 60% and were not significantly different from one another (*L.c./L.t.* 0.012±0.015, Z=0.07, p=0.95; *Q.r./Q.q.* 0.044±0.166, Z=0.26, p=0.79). The estimate of synonymous substitutions was slightly greater for the *Lampsilis* F type sequences as compared with the M type sequences, but this difference is very small and not significant. The *Quadrula* M type sequences showed higher levels of synonymous substitution, although again this difference was relatively small and not significant. In contrast, the male sequences exhibited significantly greater rates of non-synonymous substitution (Table 3.3; Figure 3.1). Although the female sequences showed less than 1% differentiation for both pairwise comparisons, the male sequences are greater than 1% different for both pairwise comparisons (Table 3.3). The *Lampsilis* male sequence distance is 5%, much greater than the 0.7% differentiation of the female sequence (0.045± 0.015, Z = 3.02, p = 0.0026); the *Quadrula* female sequences are identical at the amino acid level, whereas the male sequences are 1.4% divergent (significantly different from 0, Z = 2, p = 0.0456).

Maximum likelihood (ML) analysis produced slightly different results but the pattern of significance was similar. Estimates of synonymous divergence are greater overall for the four pairwise distances (Table 3.2) but the difference between male and female sequences is not significant for either comparison (*L.t./L.c.* 0.475±0.795, p = 0.551; *Q.r./Q.q.* 0.043±0.858, p = 0.97). Although the difference between the *Lampsilis* taxa is substantial, this difference is not significant because of the large standard errors, particularly the large standard error for the male estimate. It is interesting that for this analysis the difference between the *Lampsilis* F and M type sequences represents a greater rate for the F type sequence. Maximum likelihood estimates of pairwise non-synonymous divergence are very similar to the NG estimates (Table 3.3). As such, the pattern of significance is similar to the pattern of the NG estimates. Non-synonymous distances are significantly different for the *Lampsilis* comparison, but are only suggestive for the *Quadrula* comparison (*L.t./L.c.* 0.044±0.015, p=0.0027; *Q.r./Q.q.* 0.0124±0.007, p=0.084).

This pattern of similar rates of synonymous substitution is not reflected in the 218 bp of the 3' end of the COII gene. The difference in synonymous distances for the male and female genomes is 18.3% and 32.1% respectively for the *Lampsilis* and *Quadrula* comparisons (Table 3.4; Figure 3.2). These differences are not significant for the *Lampsilis* comparison (18.3 ± 13.5%, Z = 1.35, p = 0.177) but are significant for the *Quadrula* comparison (32.1 ± 13%, Z = 2.46, p = 0.0138). The large difference in synonymous rates is the result of a two-fold decrease in the rate of substitution for the F type sequences at COII. Non-synonymous pairwise distances are similar to the COI locus for the M type sequences, but are much greater for the F type sequences (Table 3.5;

Figure 3.2). In agreement with this non-synonymous differences are not significant for either the *Quadrula* comparison (0.01±0.014, Z = 0.735, p = 0.4654) or for the *Lampsilis* comparison (0.025±0.02, Z = 1.25, p = 0.211). In both cases the difference in rate between the M and F type sequences is smaller than at the COI locus.

As with COI, ML analysis produced similar results to the NG analysis for the COII locus. In contrast, comparisons that are significant with the NG analysis are not significant using ML estimation although the differences are larger. The difference in synonymous rate for the *Lampsilis* taxa is 21.4% but the estimated ML standard error is much larger (±0.392) and as such the associated p value is even greater than for the NG comparison (p=0.585). The *Quadrula* comparison produced similar results; the difference in estimated rates of synonymous substitution is large (3.89) but the estimated standard error is equally large (±3.78) and the comparison is not significant (p=0.303). Again, ML estimates of non-synonymous differences are similar to the NG estimates and are not significant for either of the comparisons. The ML estimates were slightly smaller for both the M and F type *Lampsilis* sequences, resulting in a smaller but insignificant difference between the two lineages (0.021±0.019, Z=1.14, p=0.254). Maximum likelihood estimates of non-synonymous divergence were also smaller for the M and F type *Quadrula* sequences (Table 2); however, in this case the ML estimate for the M type sequences was much lower than the NG estimate resulting in a larger, but still insignificant difference (0.016±0.01, Z=1.55, p=0.121).

Likelihood Analysis of Among Codon Variation

Likelihood values for the F type COI sequences were similar for the neutral, selection and discrete models (Table 3.6), although none of these was significantly better

than the likelihood for the one-ratio model (likelihood ratio test [LRT] = 3.52, d.f. = 2, p = 0.17); thus, the one-ratio model with only a single parameter (ω = 0.0008) is adequate to describe the ω ratio variation at this locus. The nearly identical likelihood values for the neutral, selection and discrete models is due to identical estimated classes and proportion of sites in each class (Table 3.6). All estimated a class with an ω ratio of zero (fixed sites) or nearly zero and classified an overwhelming majority of sites into this class (Table 3.6). The remaining handful of sites are estimated to have an ω ratio of nearly one (in the case of the neutral model this class has a fixed ω value of 1). That none of these models are significantly better than the one-ratio model is likely the result of the limited data set (analysis not shown; homologous sequence for the M type COI locus is not available).

Similar to the F type sequences, the one-ratio model did not have the best likelihood, but none of the other models had significantly better likelihoods (LRT = 5.28, d.f. = 2, p = 0.07 for selection; LRT = 5.26, d.f. = 2, p = 0.07 for two-category discrete). The selection and discrete models again had nearly identical estimates of ω for each class (ω is fixed at 0 for the lower rate class of the selection model) and distribution of sites into each of the two classes (Table 3.6). Similar to the F type sequences the lower rate class was nearly zero; however, only 65% of sites are classified into this class in contrast to 98% of sites for the F type sequences. Also different is the estimated ω value for the higher rate class; the M type sequences had a much larger proportion of sites in this class, but its estimated ω value is much lower, only being 0.08 as compared with nearly 1 (Table 3.6). In contrast to the F type sequences, the neutral model was significantly worse than all the other models for the M type sequences (LRT = 47.76, d.f. = 1, p =

$4.4X10^{-11}$ *vs.* one-ratio; LRT = 53, d.f. = 2, $p = 3.1X10^{-12}$ *vs.* selection and discrete). Thus, although the LRT for the selection and two-category discrete models *vs.* the one-ratio model are nearly significant, the one-ratio model with a single ω value of 0.0275 is adequate to describe variation in ω for this data set.

As with the COI sequences, the selection and discrete models had the best likelihoods of the models examined for the COII sequences. Also as with the COI sequences, the selection and discrete models had nearly identical estimates of ω and the proportion of sites in each class. At the F type COII locus, the selection and discrete models produced identical results (Table 3.6; thus, the identical likelihoods), grouping 75% of sites into a category with an estimated ω value of zero or nearly so. The remaining one-quarter of sites are assigned to a category with an estimated ω value of 0.163. In contrast to the COI sequences, both of these models are significantly better than the one-ratio model, which estimated a single ω ratio of 0.0297 (LRT = 8.8, d.f. = 2, $p = 0.01$ for both comparisons). The neutral model was also significantly worse (LRT = 19, d.f. = 2, $p = 0.000075$ for both comparisons), predicting a very large proportion of sites (nearly 19%) in the class with an ω value fixed at one. Thus, in contrast to the M and F type COI sequences, a single ω ratio for all sites is not adequate to explain among codon variation in this data set and the partitioning of sites into two classes provides a better explanation of the data.

This is also true for the M type COII sequences; in particular, this was the only locus where the discrete model was significantly better than all other models. The discrete model grouped approximately 86% of sites into the low ω rate ratio category; however, for this data set this class has an estimated ω ratio (0.0042) that is two orders of

magnitude greater than that of the other sequences (Table 3.6). The remaining 14% of sites were grouped into a class that had a higher ω rate ratio than any of the other sequences with the exception of the higher rate class of the F type COI discrete model. In contrast to the other loci, the selection model estimated a small proportion (1%) of neutral sites (ω fixed at 1) with a majority of sites in the fixed class and the rest in a class with ω = 0.1, but this model was significantly worse (LRT = 4.2, d.f. =1, p = 0.04). The one-ratio model and the neutral model both have significantly larger likelihoods as compared with the discrete model and as such are strongly rejected (LRT = 28, d.f. = 2, p = $8.3X10^{-7}$ for one-ratio; LRT = 43.2, d.f. = 2, p = $4.2X10^{-10}$ for neutral). The one-ratio model estimated an ω rate ratio nearly identical to the one-ratio model for the F type COII sequences and the M type COI sequences. Also similar to the F type COII and M type COI sequences, the neutral model classified a large number of sites (23%) into the neutral (ω fixed at 1) category.

Discussion

If the level of interest of a hypothesis can be gauged by the amount of work published, then variation in rates of evolution and the ultimate causes of this variation are of great interest (Li 1997). Since it's inception the molecular clock rate hypothesis has been subject to an extensive level of testing (Li 1993; Li 1997). Rejection of the clock hypothesis has been ascribed to metabolic rate, generation time and effective population size as causes of rate variation; these hypotheses all posit a neutral force as the ultimate causative factor. In contrast, there has been a recent interest in selection as a driving force in the rate of evolution (Fay and Wu 2001) and in the action of selection on the mitochondrial genome (Rand 2001).

Selection has been implicated as the cause of rate differences in the *Mytilus* M and F genomes. Evaluation of 813bp of COIII sequence indicated significantly different rates of non-synonymous substitution between the M and F types, whereas the difference in synonymous rates between the M and F types was not significant (Stewart *et al.* 1996). In addition, there were more fixed non-synonymous substitutions than expected and in the M lineage there is a greater number of amino acid substitutions at sites deemed "variable" (Stewart *et al.* 1996). In contrast, comparisons between American *M. trossulus* and European *Mytilus* populations indicated a significant excess of non-synonymous polymorphism and a significantly greater rate of substitution for the M genome at both synonymous and non-synonymous sites (Quesada, Warren and Skibinski 1998).

Although an increased rate of evolution of the M lineage has been shown for the unionid *Pyganodon*, the causes of this rate difference have not been identified. The large number of species in the Unionidae allows for a wide range of pairwise comparisons allowing thorough testing of the causes of rate differences. Thus, we examined patterns of substitution in four unionid taxa, *Quadrula quadrula, Q. refulgens, Lampsilis teres* and *L. claibornensis*, using approximate and likelihood methods.

Alignments of the COII and COI loci were relatively straightforward. The putative COII amino acid sequence contained a single indel whose location is considered unambiguous as it is flanked by conserved sites. Over half of the amino acids of COII are conserved across gender mitotype indicating the conservative evolutionary nature of this region of the COII locus. The COI locus showed even greater levels of conservation across gender mitotypes also indicative of a strong level of purging selection at the replacement substitution level; thus, the COI alignment was also straightforward and unambiguous.

Pairwise Distance Comparisons

We estimated the proportion of synonymous and non-synonymous substitutions between two pairs of taxa, *Quadrula refulgens/Q. quadrula* and *Lampsilis teres/L. claibornensis* using both approximate and likelihood methods. These pairwise comparisons are phylogenetically independent, allowing for orthologous contrasts of the M and F type distances between the pairs of taxa. These comparisons are also the first for the M and F lineages of the freshwater Unionidae.

Pairwise comparisons of rates of synonymous and non-synonymous substitutions are often used to identify the underlying mechanisms of rate differences between

organisms (Li 1997). Stewart *et al.* (1996) were the first to apply this comparative method to the M and F type sequences of bivalves with DUI; however, they estimated d_S and d_N using only the NG approximate method with a single pair of taxa and the estimates obtained were extremely high (for a critique of this analysis see Chapter 5). Here we have estimated d_S and d_N between two pairs of unionids using both the NG approximate method and a ML method.

Comparisons of NG estimates of d_S for the F type sequences provide different pictures of the underlying rates of synonymous evolution at the COII and COI loci. Differences between d_S for the M and F type sequences at COII indicated an overall greater rate of synonymous substitution for the M type sequences, although this difference is only significant for the *Quadrula* comparison. In contrast, comparisons of pairwise estimates of d_S at the COI locus suggest nearly identical rates of synonymous substitution for the M and F types. This inconsistency is a result of nearly two-fold greater estimates of d_S for the F type COI sequences as compared with COII; in contrast, estimates of d_S are within the same range for the COI and COII M type sequences. Hypotheses have been proposed to explain differences in synonymous rate; however, many of these hypotheses fail to explain differences in rate within genomes. Differences in mutation rate due to a greater number of germ cell divisions or metabolic rate would be expected to affect all loci equally (Li 1997). Pairwise divergence has been shown to increase linearly in the mitochondrial genome in proportion to distance from the replication origin (Tanaka and Ozawa 1994; Bielawski and Gold 1996; Reyes *et al.* 1998; Bielawski and Gold 2002). This is hypothesized to be due to the time spent as a single stranded molecule during replication and transcription of mtDNA (Tanaka and Ozawa

1994). The origin of replication is unknown for these genomes; however, in both the F and M genome the COII and COI loci are adjacent and located on the same strand with only a few base pairs between them. Thus, it is unlikely that the origin of replication model is responsible for the differences in synonymous rate.

Non-random codon usage has been shown to lead to biased estimates of d_S and this bias is greater for the NG approximate method (Dunn, Bielawski, and Yang 2001). Maximum likelihood methods account for non-random codon usage; thus, this method is less prone to bias, although when codon usage is highly non-random the bias is still severe. If non-random codon usage is biasing estimates of d_S and leading to different estimates of d_S for the COI and COII F type sequences then ML estimates of d_S should be larger than NG estimates and the F type COII sequences should consistently have lower estimates of the ENC. Although the ML estimates of d_S are consistently greater than the NG estimates suggesting codon bias, the F type COII sequences do not have consistently lower estimates of ENC. The *L. claibornensis* F type COI sequence has a much larger ENC than the respective COII sequence but ENC's for the remainder of the sequences are nearly identical. In particular, the *Quadrula* F type ENC's are nearly indistinguishable for the COI and COII loci; however, these sequences showed the largest difference in d_S between the COII and COI loci (33% for NG and 65% for ML estimates). Thus, these data do not support the hypothesis that differences in synonymous rates of substitution between the F type COI and COII loci are caused by non-random codon usage.

Comparisons of d_S between the M and F types indicated conflicting patterns resulting from different estimates of synonymous substitution between the F type COI and COII loci. Paired tests of COII M and F sequences suggest that the rate of

synonymous substitution is two-fold greater for the M genome, although this difference is only significant for the *Quadrula* NG comparison. In contrast, paired tests of COI M and F sequences indicate that the rate of synonymous substitution is not significantly different between the two genomes. As noted above, this is due to a substantial difference between the COI and COII estimates of d_S for the F type sequences. Classical hypotheses of rate differences fail to account for the difference in *intra-genomic* rates and this difference does not appear to be the result of intrinsic factors such as non-random codon usage. This presents a conundrum on which rests interpretation of the mechanisms of rate differences between the M and F genomes. One interpretation of the data is that the differences in synonymous rate at the COII locus reflect the true difference in substitution mechanisms between the M and F genomes. A corollary of this hypothesis is that levels of substitution between the M type sequences at the COI locus are saturated and do not reflect the true divergence. This hypothesis is supported by ML estimates of d_S where the transition/transversion rate ratio (κ) is fixed for all pairwise comparisons. In this case, estimates of d_S at both the COI and COII loci are greater for M type sequences as compared with the F type (analysis not shown); however, these differences are not significant due to the large standard errors. Alternatively, differences in estimates of d_S may be due to intrinsic factors such as non-random codon usage or unequal saturation. We have excluded non-random codon usage as a cause of differences in rate *within* the *Quadrula* F genome. It is possible that the F type *Quadrula* COII sequences have saturated at a level of divergence half that of the M type COII sequences. This hypothesis is not supported by the pairwise ML or codon usage analyses. Estimation using ML methods account for some amount of codon bias, and despite this correction

estimates of d_S are still greater for the M type sequences although these differences are not significant. Additionally, estimates of the ENC for the COII locus are greater for the F type sequences than for the M type sequences; therefore, the F type sequences should saturate at higher levels of divergence but the opposite is observed (Dunn, Bielawski and Yang 2001). Thus, we cautiously conclude that synonymous rates of substitution are greater for the M genome, but add that this hypothesis is in need of further testing. In particular, we suggest that comparisons between sequences with less than 10% divergence will be most useful, as ML methods are relatively robust at this level of divergence even with extreme codon bias (Dunn, Bielawski and Yang 2001).

Several hypotheses, both neutral and non-neutral, have been advanced to explain synonymous differences in rates of evolution. In *Drosophila* estimates of synonymous rates of substitution have been shown to correlate with codon bias and this result has been interpreted as evidence of translational selection (reviewed in Skashi and Eyre-Walker 1998). This hypothesis is not consistent with estimates of the ENC; the F type sequences consistently have a greater estimate of the ENC but lower rates of synonymous substitution. Additionally, it has been shown that the correlation of d_S and ENC is a result of the failure of approximate methods to account for codon bias (Dunn, Bielawski and Yang 2001). Neutral explanations of rate differences include metabolic rate, generation-time and male-driven evolution (reviewed in Li 1993 and Li 1997). Given that the two genomes examined are derived from individuals of the same species we argue that metabolic rate and generation time can be excluded.

Male-driven evolution, hypothesized to be the result of a greater number of germ cell divisions in males, has been proposed as the causal mechanism of rate differences

between Y-linked loci and autosomal and X-lined homologues in primates (Shimmin, Chang, and Li 1993; Bohossian, Skaletsky, and Page 2000; Makova and Li 2002) and Z/W homologues in birds (Ellegren and Fridolfsson 1997). In primates, the ratio of male to female mutation (α_m) has been estimated to be between 3 and 10 (Hurst and Ellegren 1998). Estimates of unionid α_m using the NG pairwise distances for the COII locus are below this range. The male to female ratio estimated from the *Quadrula* data set is approximately 2 and α_m estimated from the *Lampsilis* data set is about 1.5. The male to female bias is hypothesized to be the result of a greater number of germ cell divisions in males. The relative reproductive output of these two taxa is unknown; based on analysis of the COII locus we hypothesize that if male driven evolution is the cause of rate differences then the relative reproductive output should favor males by a factor of at least two. We consider this to be a conservative estimate because these sequences are highly diverged and likely saturated; thus, we propose that α_m is likely greater than estimated using this data.

If male-driven evolution is responsible for the proposed difference in rates of synonymous substitution then population level comparisons of mitotype diversity should reveal greater levels of diversity for the M type as compared with the F type. Unfortunately, these data are not available for unionids; however, several studies have examined mitotype diversity in populations of *Mytilus* (Skibinski, Gallagher and Beynon 1994; Quesada, Warren and Skibinski 1998; Ladoukakis *et al.* 2002). Populations of *M. edulis* from coastal England and *M. galloprovincialis* from the Atlantic and Mediterranean all showed higher mitotype diversity for the M genome (Skibinski, Gallagher, and Beynon 1994b; Quesada, Warren, and Skibinski 1998); in contrast, Black

Sea *M. galloprovincialis* had lower haplotype diversity for the M genome(Ladoukakis *et al.* 2002). Thus, as with most other aspects of the evolution of DUI these data are inconclusive.

Interpretation of differences in rates of non-synonymous substitution are confounded by the ambiguity of patterns of evolution at the synonymous level. If underlying base substitution rates are different between the M and F genomes because of mutation rate then a model of neutral evolution at the non-synonymous level is adequate to explain the differences observed. Alternatively, if rates of base substitution are equivalent between the two genomes then non-neutral forces may be responsible for the differences in non-synonymous rates observed at the COI locus. Although we will withhold interpretation of these differences until additional evidence weighs in, below we provide a guarded interpretation of a more dynamic analysis of the ratio of synonymous to non-synonymous substitutions.

Maximum Likelihood Analysis of intra-genic variation in ω

Methods that average estimates of the ω ratio over all codons in a locus have recently been criticized because this analysis may obscure identification of positive selection acting on a few codons within the locus (Nielsen and Yang 1998). Although we expect that the proportion of codons under positive selection in mitochondrial loci is near zero (Yang *et al.* 2000), this analysis does provide the opportunity to estimate the number of classes of codons and the proportion of codons in each class, similar to the COVARION analysis of Fitch and Margoliash (1967). We interpret these results with the caveat that it is unclear what affect the observed saturation may have on the results.

All loci and gender types were similar in that a large proportion of sites are estimated to have an ω ratio of nearly zero (less than 0.005). The proportion of sites ranged from a minimum of 65% for the M type COI sequences to 98% for the F type COI sequences. This is consistent with previous analyses of mammalian mitochondrial loci (Yang *et al.* 2000).

The proportion of sites in each class as well as their ω values vary between sequences for the COII locus; 85% of codons in the M type sequences are assigned to a class with an average ω value of 0.0042, but 75% of codons in the F type sequences are assigned to a class with an average ω value of 0.00001. This may reflect subtle differences in levels of selection on the two gender type sequences; however, the remaining 25% of codons in the F type sequences are assigned to a class with an average ω value of 0.16, but only 14% of M type codons are assigned to a class with the similar ω value of 0.21. Surprisingly, this leads to an average ω ratio of 0.041 for the F type COII sequences and 0.034 for the M type COII sequences. On average these sequences appear to be under similar levels of selection given that 4% non-synonymous divergence per 100% synonymous divergence is not substantially different than 3.4% per 100%; however, there may be subtle differences in how selection is acting on these sequences. In particular, these results indicate that certain non-synonymous substitutions may be acceptable at most codons in the M type sequences but not acceptable in the F type sequences. Examination of the posterior probabilities for each site also reveals interesting trends. Probabilities for sites in the higher rate class for the F type sequences are nearly one (average probability = 0.996, S.D.[ω]=0.00045), indicating that these sites form a cohesive class with very similar ω ratios. In contrast, probabilities for sites in the

lower rate class for the F type sequences are more variable (average probability = 0.901, S.D.[ω]=0.00790), suggesting that this rate class is more heterogeneous than the higher rate class. Thus, in the F type COII sequence there is a class of sites whose ω ratios are consistently high and a class of sites whose ω ratios are low but variable within this range. In contrast, the two rate classes for the M type COII sequences appear to be equally heterogeneous. The average probabilities for sites in each class are nearly identical (94.5 and 93.6%) and the variances in estimated ω ratios are identical (S.D.=0.021). Thus, although there are subtle differences in the patterns of substitution for the two gender types, overall they exhibit very similar patterns of evolution.

In this region of the COII locus a large majority of sites are under stringent selection with less than 0.5% non-synonymous divergence per 100% synonymous divergence; the remaining sites are more free to vary and over time will make a significant contribution to non-synonymous evolution at this locus, respectively 16% and 20% non-synonymous divergence per 100% synonymous divergence for the F and M types. Amazingly this suggests that the contribution of the greater ω sites to non-synonymous evolution is approximately 8.3X (calculated as $P[\omega_1]*\omega_1/P[\omega_2]*\omega_2$) that of sites in the smaller rate class for the M type sequences and 5,000X that of sites in the smaller rate class for the F type sequences. It should be pointed out that if the rate of synonymous substitution has been underestimated for the M type sequences, the difference in ω ratios for the M and F types could be an artifact. In particular the contribution of the greater rate class to non-synonymous evolution may be similar to that of the F type sequences.

Analysis of inter-codon variation in ω at the M type COI locus provides a picture similar to that of COII. This locus had the smallest proportion of sites in the low ω ratio class and *de facto* the highest proportion of sites in the higher ratio class. The estimated ω ratio for the low class is identical to that for the F type COII sequences; however, the ω ratio for the high ratio class is the lowest among all loci examined (ω=0.086). Thus, patterns of evolution are similar at this locus as compared with COII and the average ω rate ratio is not substantially different from the two COII loci, again in the range of 3-4% non-synonymous divergence per 100% synonymous divergence. Also, similar to the F type COII sequences the contribution of the higher ratio class to rates of non-synonymous evolution is approximately 4,700X that of the lower ratio class.

In contrast to COII, the COI F type sequences exhibit a much more conserved pattern of evolution. The two category discrete model was not significantly better than the one-ratio model; however, the results of the two category discrete model were similar to the one-ratio model. Thus, for sake of comparison with the COII locus we will include the two category discrete model in our discussion. The average ω ratio for the one-ratio model is indicative of the conservative evolution at this locus and although the average ω ratio for the two category discrete model is 20X greater, this greater ratio is still one-half that of the average ω ratio for the other three loci. The one-ratio model estimates only 0.08% non-synonymous divergence per 100% synonymous divergence and the two category discrete model estimates 1.6% per 100% synonymous divergence. The difference in average ω ratios for the two models is the result of a small (1.7%) proportion of sites with an ω ratio near unity. The contribution of this small class of sites to overall rates of non-synonymous substitution is 1,630X that of the lower rate class.

These analyses indicate that patterns of molecular evolution in these genomes are not easily understood through comparisons of average non-synonymous/synonymous ratios. The M and F type COII loci and the M type COI locus have similar ω ratios when averaged across all codons; however, there are subtle differences in the rate underlying rate classes. Despite the slightly smaller average ω ratio for the M type COII sequences, the smallest ω ratio for codons at this locus is two orders of magnitude greater than the smallest ω ratio at the COII F type locus. Thus, although the COII F type locus has a larger proportion of sites that are under reduced selective constraint, almost all codons in the M type COII locus are under less selective constraint than the same codon in the F type COII locus. Additionally, codons in the higher ω ratio class for the F type COII locus are incredibly uniform in their ω ratios, possibly indicating a very consistent level of selection at these sites. In contrast, codons of the M type COII locus and, to a greater extent, codons of the smaller ω ratio class for the F type COII locus are more heterogeneous possibly indicating either varying levels of selection between these codons or fluctuating selection levels at these sites. Lastly, based on these analyses it is clear that a small proportion of sites are contributing to the estimated levels of non-synonymous divergence at these loci. Although this may not be surprising, the estimated relative contribution of this class of sites to non-synonymous evolution is extraordinary and has implications for saturation at non-synonymous sites. In particular for the loci examined, by the time synonymous sites have reached complete saturation (*i.e.* 100% true divergence) non-synonymous sites in the higher ω ratio category are approaching moderate levels of saturation and non-synonymous sites in the lower ω ratio category make effectively no contribution.

Elucidating Causes of Molecular Evolutionary Rate Differences of Sex-Associated mtDNA in Bivalves-Conclusions

It is clear that the elucidation of the causes of rate differences in the F and M type genomes of bivalves requires additional work. In particular, comparisons between M and F genomes with lower levels of divergence are essential. It is unclear what, if any, bias substitutional saturation may have introduced into this data set. Not only do differences in rates of synonymous substitution between the M and F genome require more attention, differences in rate within the F genome remain perplexing. Additional comparisons will also test hypotheses of the subtle differences between these genomes in their patterns of evolution. Modeling of the effects of substitutional saturation, in particular at synonymous sites, may also help test if the observed patterns are indicative of the true underlying patterns or bias introduced by differential saturation. These tests will not only have implications for the underlying causes of differences in rates of synonymous evolution and the patterns of substitution in these genomes but, possibly, also for the poorly understood phenomenon of non-synonymous saturation.

Table 3.1. Estimates of the effective number of codons.

	L.t. F	L.c. F	Q.r. F	Q.q. F	L.t. M	L.c. M	Q.r. M	Q.q. M
COI	33.89	43.23	35.24	34.04	32.9	38.46	34.32	31.62
COII	32.44	32.01	35.05	34.23	35.8	36.62	31.26	30.07

Table 3.2. Pairwise synonymous distances of male and female COI region. Likelihood estimates above the diagonal, Nei-Gojobori Jukes-Cantor corrected values below the diagonal. Standard errors in parentheses. Asterisks indicate that the standard error is undefined.

	Q.r.F	Q.q.F	L.t.F	L.c.F	Q.r.M	Q.q.M	L.t.M	L.c.M
Q.r.F		1.77	1.77	1.68	7.20	4.36	4.89	5.66
		(0.139)	(0.489)	(0.615)	(4.53)	(2.61)	(1.94)	(4.9)
Q.q.F	0.599		2.23	2.79	13.21	9.11	6.21	61.86
	(0.119)		(0.696)	(1.11)	(38.33)	(16.42)	(4.81)	(86.33)
L.t.F	1.028	1.024		1.40	4.50	77.76	31.87	70.71
	(0.213)	(0.212)		(0.758)	(1.95)	(95.16)	(52.17)	(108.7)
L.c.F	0.854	0.929	0.585		6.46	6.76	6.23	71.36
	(0.170)	(0.187)	(0.117)		(3.86)	(3.49)	(3.23)	(****)
Q.r.M						1.81	2.12	1.73
						(0.85)	(0.692)	(0.485)
Q.q.M					0.643		2.08	2.21
					(0.126)		(0.60)	(0.823)
L.t.M					0.824	0.854		0.925
					(0.163)	(0.169)		(0.243)
L.c.M					0.914	0.927	0.573	
					(0.183)	(0.186)	(0.114)	

Table 3.3. Pairwise non-synonymous distances of male and female COI region. Likelihood estimates above the diagonal, Nei-Gojobori Jukes-Cantor corrected values below the diagonal. Standard errors in parentheses. Asterisks indicate that the standard error is undefined.

	Q.r.F	Q.q.F	L.t.F	L.c.F	Q.r.M	Q.q.M	L.t.M	L.c.M
Q.r.F		0.018	0.008	0.004	0.229	0.232	0.193	0.216
		(0.000)	(0.005)	(0.004)	(0.034)	(0.034)	(0.030)	(0.033)
Q.q.F	0.000		0.008	0.004	0.229	0.232	0.192	0.213
	(0.0000)		(0.005)	(0.004)	(0.034)	(0.034)	(0.030)	(0.032)
L.t.F	0.022	0.026		0.004	0.231	0.234	0.191	0.218
	(0.009)	(0.010)		(0.004)	(0.034)	(0.034)	(0.030)	(0.033)
L.c.F	0.015	0.018	0.007		0.231	0.234	0.198	0.218
	(0.007)	(0.008)	(0.005)		(0.034)	(0.035)	(0.031)	(0.033)
Q.r.M						0.014	0.072	0.054
						(0.007)	(0.017)	(0.014)
Q.q.M					0.015		0.072	0.061
					(0.007)		(0.017)	(0.015)
L.t.M					0.086	0.086		0.048
					(0.018)	(0.018)		(0.014)
L.c.M					0.064	0.070	0.052	
					(0.016)	(0.016)	(0.014)	

Table 3.4. Pairwise synonymous distances of male and female COII region. Likelihood estimates above the diagonal, Nei-Gojobori Jukes-Cantor corrected values below the diagonal.

	Q.r.F	Q.q.F	L.t.F	L.c.F	Q.r.M	Q.q.M	L.t.M	L.c.M
Q.r.F		0.644	1.978	2.069	93.66	92.51	87.71	83.93
		(0.225)	(0.693)	(0.884)	(101.8)	(112.0)	(150.0)	(147.3)
Q.q.F	0.255		1.900	1.209	94.41	93.58	87.879	83.47
	(0.069)		(0.628)	(0.354)	(****)	(107.1)	(110.7)	(88.91)
L.t.F	0.603	0.635		0.746	96.83	98.00	91.05	86.02
	(0.132)	(0.138)		(0.258)	(201.9)	(105.0)	(****)	(****)
L.c.F	0.487	0.460	0.301		94.61	94.61	40.84	84.26
	(0.110)	(0.105)	(0.077)		(149.2)	(****)	(****)	(151.9)
Q.r.M	0.939	0.922	0.897	0.877		4.534	5.365	5.355
	(0.212)	(0.207)	(0.199)	(0.195)		(3.772)	(4.004)	(4.232)
Q.q.M	0.918	1.162	0.877	0.858	0.553		5.532	2.707
	(0.206)	(0.281)	(0.194)	(0.189)	(0.125)		(4.334)	(1.064)
L.t.M	0.951	0.925	0.671	0.831	0.753	0.756		0.960
	(0.216)	(0.207)	(0.146)	(0.182)	(0.195)	(0.167)		(0.295)
L.c.M	0.953	0.817	0.947	0.877	0.871	0.798	0.490	
	(0.206)	(0.181)	(0.214)	(0.195)	(0.196)	(0.178)	(0.113)	

Table 3.5. Pairwise non-synonymous distances of male and female COII region. Likelihood estimates above the diagonal, Nei-Gojobori Jukes-Cantor corrected values below the diagonal. Standard errors in parentheses. Asterisks indicate that the standard error is undefined.

	Q.r.F	Q.q.F	L.t.F	L.c.F	Q.r.M	Q.q.M	L.t.M	L.c.M
Q.r.F		0.021	0.053	0.054	0.247	0.245	0.271	0.270
		(0.009)	(0.016)	(0.016)	(0.039)	(0.039)	(0.042)	(0.042)
Q.q.F	0.024		0.063	0.064	0.241	0.239	0.274	0.288
	(0.011)		(0.017)	(0.017)	(0.038)	(0.038)	(0.043)	(0.044)
L.t.F	0.068	0.078		0.026	0.226	0.223	0.256	0.263
	(0.018)	(0.020)		(0.011)	(0.037)	(0.037)	(0.040)	(0.041)
L.c.F	0.063	0.073	0.031		0.221	0.219	0.251	0.257
	(0.018)	(0.019)	(0.012)		(0.036)	(0.036)	(0.040)	0.040
Q.r.M						0.005	0.093	0.110
						(0.005)	(0.021)	(0.024)
Q.q.M					0.014		0.101	0.120
					(0.008)		(0.023)	(0.026)
L.t.M					0.118	0.118		0.047
					(0.025)	(0.025)		(0.015)
L.c.M					0.144	0.142	0.056	
					(0.027)	(0.027)	(0.016)	

Table 3.6. Results of likelihood intra-genic ω ratio analysis. Table is divided by locus and gender type. Models are described in the methods section, κ is the transition/transversion rate ratio, ω is as defined in the text, and *f* is the frequency of sites.

		COI				COII			
	Model	LnL	κ	ω	*f*	LnL	κ	ω	*f*
F	One-ratio	-730.046	34.05	0.0008	1.00	-600.539	13.56	0.0297	1.00
T	Neutral	-728.277	99	0	0.983	-605.553	17.98	0	0.815
Y				1.00	0.017			1.00	0.185
P									
E	Selection	-728.276	99	0	0.983	-596.080	17.57	0	0.749
				1.00	0			1.00	0
				0.94282	0.017			0.1625	0.250
	Discrete	-728.319	99	0.00001	0.983	-596.080	17.58	0.00001	0.749
				0.9479	0.017			0.16258	0.251
M	One-ratio	-856.625	5.05	0.0275	1.00	-718.969	8.06	0.0269	1.00
T	Neutral	-880.501	8.48	0.0000	0.810	-726.576	14.37	0.0000	0.770
Y				1.0000	0.190			1.0000	0.230
P									
E	Selection	-853.990	5.48	0.00000	0.645	-707.133	13.17	0	0.723
				1.00	0.000			1.00	0.010
				0.08563	0.355			0.10042	0.267
	Discrete	-853.990	5.48	0.00001	0.645	-704.997	13.23	0.0042	0.857
				0.08565	0.355			0.20905	0.143

Figure 3.1. Bar graph of synonymous and non-synonymous distances for the COI locus.

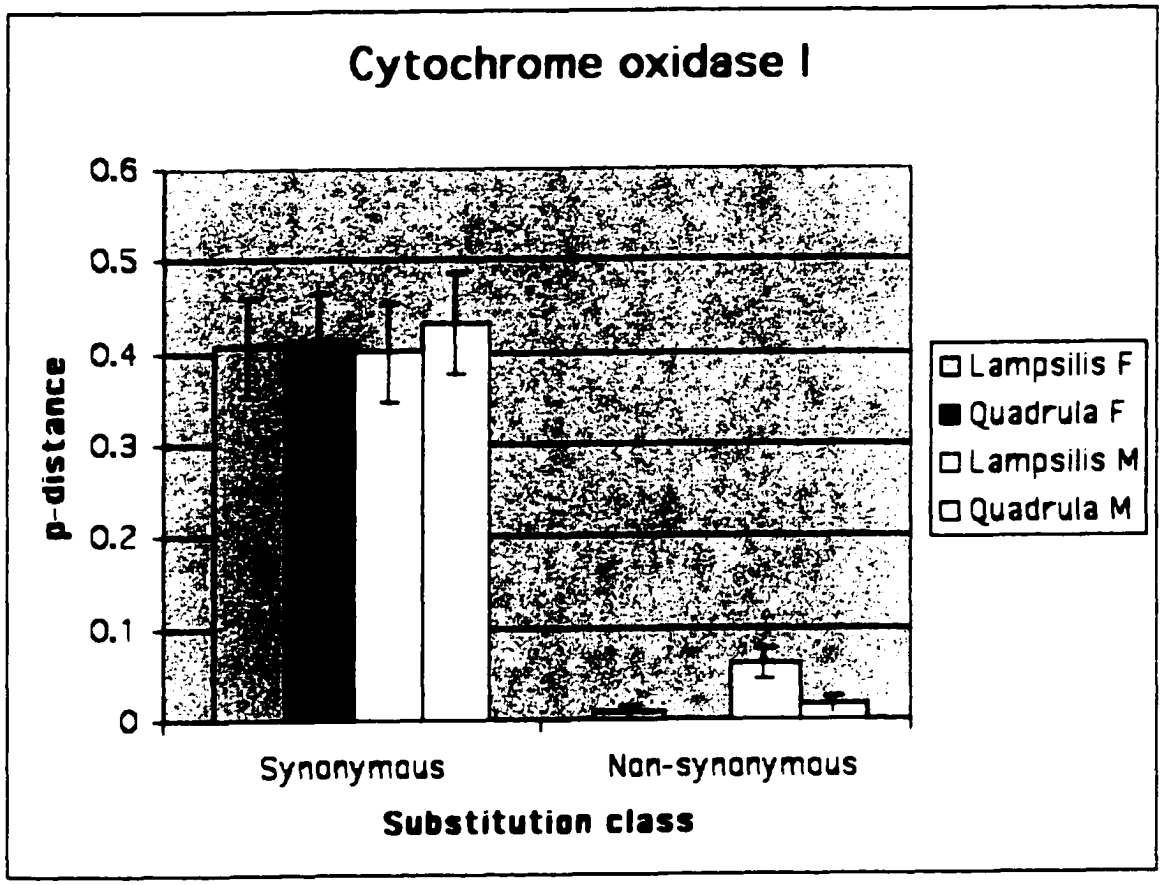
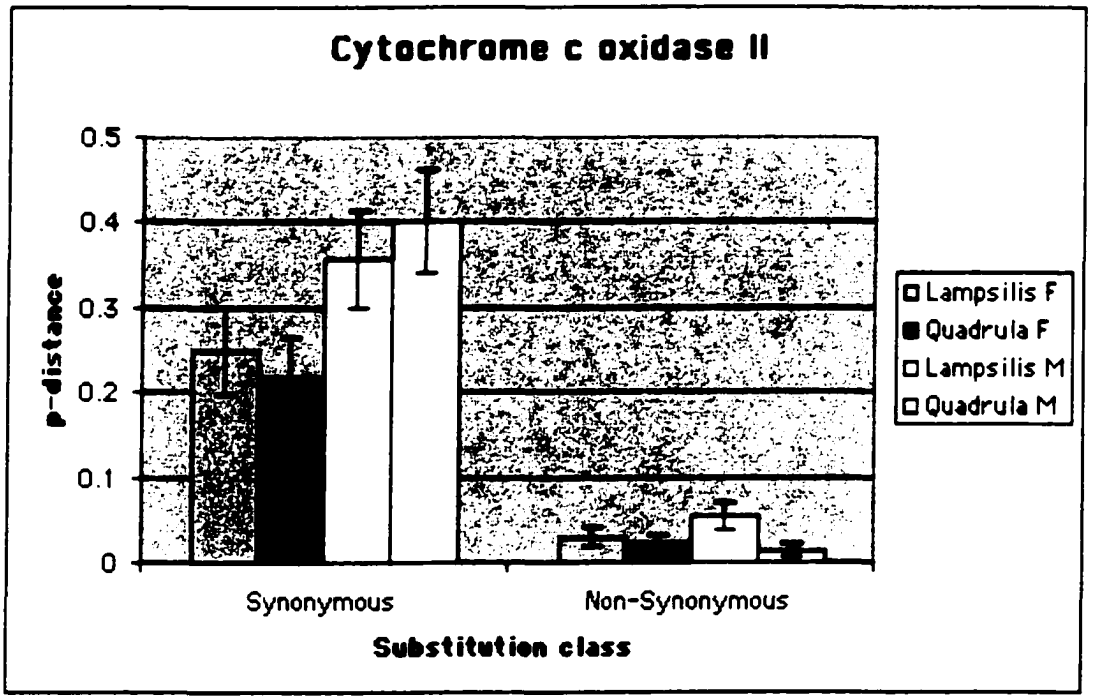


Figure 3.2. Bar graph of synonymous and non-synonymous distances for the COII locus.



CHAPTER IV

RELAXED SELECTION DOES NOT EXPLAIN THE RAPID EVOLUTON OF THE *MYTILUS* F MITOCHONDRIAL GENOME

Abstract

Mytilus maternally inherited (F type) mitochondrial DNA (mtDNA) has been shown to have an unusually high rate of evolution. This has been hypothesized to be the result of relaxed selection on the M type mitochondrial genome due to the presence of a paternally inherited mitochondrial genome (M type). The presence of two mtDNA lineages in a separate family of bivalves, the Unionidae, allows for a test of this hypothesis. We sequenced 270 base pairs from the cytochrome *c* oxidase II gene and 708 base pairs from the cytochrome *c* oxidase I gene from two species of Unionidae, *Quadrula refulgens* and *Q. quadrula*. Distance analysis with a data set comparable to the previous analysis, but including the unionids, indicates that the high rate of evolution of the *Mytilus* F type mtDNA is not observed in the *Quadrula* F type lineages. In particular, the *Quadrula* F type mtDNA has a rate of evolution similar to that of two gastropods. The *Quadrula* M type mtDNA does show an increased rate of evolution relative to the F type, but is still slower than the *Mytilus* M type. Alternative hypotheses, including that the newly discovered phenomenon of recombination in *Mytilus* may be responsible, are presented.

Introduction

Mytiloid (Bivalvia: Pteriomorpha: Mytiloida) and unionoid (Bivalvia: Palaeoheterodonta: Unionoida) bivalves possess a unique form of mitochondrial DNA (mtDNA) inheritance, named doubly-uniparental inheritance (DUI) (Fisher and Skibinski 1990; Skibinski, Gallagher and Beynon 1994a, b; Zouros *et al.* 1994a, b; Liu, Mitton and Wu 1996; Hoeh *et al.* 1996a; Curole and Kocher 2002). With DUI, as with standard-maternal inheritance (SMI), mothers transmit their mtDNA type (the F mitotype) to all offspring. In contrast to SMI, males exhibiting DUI carry a separate mtDNA type (the M mitotype) that is transmitted patrilineally.

It has been shown that in both the mytiloid *Mytilus* and the unionoid *Anodonta* nucleotide substitutions accumulate at a faster rate in the M genome than the F genome (Skibinski *et al.* 1994; Rawson and Hilbish 1995; Stewart *et al.* 1996; Liu, Mitton and Wu 1996); in addition, an increased rate of substitution is not a property of the M genome alone (Hoeh *et al.* 1996b). Comparisons of 660bp of cytochrome *c* oxidase I (COI) and 762bp of cytochrome *c* oxidase III (COIII) sequence with several other metazoan taxa indicate that both the *Mytilus* M and F genomes evolve at a increased rate relative to other metazoan taxa; relative rate tests between the gastropod *Albinaria turrita* and the chiton *Katharina tunicata* indicated a statistically significant increase in the rate of nucleotide substitutions (Hoeh *et al.* 1996b). Only the nematode *Ascaris suum* had a rate approaching that of the *Mytilus* M and F genomes. Hoeh *et al.* (1996b) proposed that the presence of two mtDNA lineages in *Mytilus* may result in a "division of labor"; thus, selection on both genomes may be relaxed relative to metazoans that have only a single mtDNA lineage. Hoeh *et al.* (1996b) pointed out that the presence of DUI in the

Unionoida allows for an independent test of this hypothesis; however, there are some distinct differences in the evolution of DUI in this freshwater group (Curole and Kocher 2002).

Occasionally a male will fail to pass his M mitotype on to his sons; in this case the male will transmit his F mitotype. This has been referred to as a gender-switching event (Hoeh *et al.* 1996a). Whereas there are frequent gender-switching events in the mytiloids (Hoeh *et al.* 1996a; Hoeh *et al.* 1997) such that the divergence of the most common M and F genomes is only 5.3MYA (Rawson and Hilbish 1995), these lineage capture events appear to be absent in the Unionoida (Hoeh *et al.* 1996a; Curole and Kocher 2002). The absence of lineage capture events has lead to two highly differentiated (average 33% p-distance for COI region in several taxa; Hoeh *et al.* 1996a) mitochondrial genomes that are over 400MY old (Curole and Kocher 2002). The apparent absence of gender-switching may be due to the presence of a 600bp sex-specific extension of the cytochrome *c* oxidase II (COII) gene in the M genome (Curole and Kocher 2002). Regardless of this difference in the evolution of DUI, if relaxed selection is operating on the female genome as hypothesized (Hoeh *et al.* 1996b) then both the M and F unionoid lineages should show an increased rate of nucleotide substitution similar to the *Mytilus* M and F lineages.

To test this hypothesis, we analyzed 978 bp of nucleotide sequence from the M and F genomes of two species of Unionoida. Included in this analysis are the taxa analyzed by Hoeh *et al.* (1996b) to allow for comparison between the two data sets. Here we show that the higher rate of evolution of both *Mytilus* genomes is limited to the M genome in the two species on Unionoida examined; therefore, relaxed selection is not

responsible for the difference in rate between the *Mytilus* F and *Albinaria* mitochondrial lineages and we offer alternative hypotheses to explain this phenomenon.

Methods

We determined 270 bp and 708 bp of nucleotide sequence from the 3' and 5' ends of the COII and cytochrome *c* oxidase I (COI) genes, respectively, for the M and F genomes of two species of Unionoida, *Quadrula refulgens* and *Quadrula quadrula*. The COII sequences were reported previously (Curole and Kocher 2002). Sequence of 708 bp of COI was determined by PCR amplification of the M and F genomes using the CO2.2 (Curole and Kocher 2002) and HCO (Folmer *et al.* 1994) primers to prime synthesis. In these species the COII gene is upstream of the COI gene and this primer pair produces a 1.2 kbp fragment for the F genome and a 1.8 kbp fragment for the M genome; the size difference is due to the 600 bp sex-specific extension. PCR, gel isolation, cloning and sequencing protocols followed those in Curole and Kocher (2002). Sequence of the F genome was determined by direct sequencing of PCR products from mantle tissue. PCR amplification of male gonadal tissue resulted in the presence of two bands; one of the size of the female band and a larger band corresponding to the M genome. The larger band was gel excised, cloned using the Promega TA cloning kit and three clones were sequenced for each individual.

Homologous COII and 5' COI sequences are not available for the *Mytilus* M mitotype, so the same taxa that were previously analyzed by Hoeh *et al.* (1996b) were included in order to make comparisons across data sets. Homologous sequence for *Albinaria coerule* (Genbank Accession # NC_001761), *Pupa strigosa* (Genbank Accession # NC_002176), *Katharina tunicata* (Genbank Accession # U09810), *Anopholes gambiae* (Genbank Accession # NC_002084), *Drosophila yakuba* (Genbank

Accession # NC_001322), *Artemia franciscana* (Genbank Accession # NC_001620), *Gallus gallus* (Genbank Accession # NC001323), *Homo sapiens* (Genbank Accession # NC_001807), *Xenopus laevis* (Genbank Accession # M10217) and *Paracentrotus lividus* (Genbank Accession # NC_001572) were obtained from Genbank. Amino acid sequences were aligned using Clustal X and any necessary gaps were inserted into the nucleotide sequences at the appropriate sites. The alignment was examined manually to determine regions that were adequately unambiguous for analysis. The M type Unionoid mitochondrial genome contains a 600 bp extension of the COII gene (Curole and Kocher 2002); thus, this region was excluded from the alignment and only the aligned region 5' of the Unionoid F type stop codon is included. Only 1st and 2nd positions of codons were used for analysis as 3rd codon positions are undoubtedly saturated at this level (Kocher and Carleton 1997). Phylogenetic estimation and bootstrap support (1000 replicates) were done using p-distance estimation and the Neighbor-Joining algorithm with the program PAUP* (Swofford 2000). Distances and standard errors were also calculated using the program MEGA (Kumar *et al.* 2001). The relative rate test (Li 1997) was used to assess whether the male and female Unionid lineages were evolving more rapidly than either the gastropod *Albinaria coerule* or the chiton *Katharina tunicata*. *Drosophila yakuba* was used as the outgroup for all relative rate tests.

Results

The alignment of the fourteen taxa included is given in Figure 1. Three indels were inserted into the amino acid alignment, one each in *Gallus*, *Albinaria* and *Ascaris*; these codons were excluded from the analysis as were 3rd codon positions. The total length of the dataset (including 3rd positions) is 978 bp, 270 bp from the COII gene and 708 bp from the COI gene.

The Neighbor-Joining (NJ) phylogenetic tree is presented in Figure 2. Similar to the previous analysis, *Ascaris* grouped with mollusks in contradiction to presumed evolutionary relationships (Hoeh *et al.* 1996b). In the NJ tree, *Ascaris* groups with the two *Quadrula* lineages with weak bootstrap support (53%). This aberrant clustering is hypothesized to be due to long branch attraction as has been argued for elsewhere (Hoeh *et al.* 1996b). In pairwise comparisons with the same taxon, the *Mytilus* M and F genomes generally showed greater rates of nucleotide substitution than the *Ascaris* mitochondrial genome (Hoeh *et al.* 1996b). In contrast to *Mytilus*, neither the M nor the F genome of the *Quadrula* species have branch lengths longer than *Ascaris*; in addition, pairwise distances between *Ascaris* and several of the taxa included in the analysis are greater than distances between these same taxa and the *Quadrula* M and F types (Table 1). Thus, in contrast to *Mytilus*, the two *Quadrula* species do not appear to have higher rates of evolution than *Ascaris*. In particular, the *Quadrula* F genomes are clearly slower than *Ascaris* in Figure 2 and are smaller in all of the pairwise distance comparisons. Relative rate tests between the *Quadrula* M and F genomes and *Ascaris* are significant

for the F genome (*Q.q*.M vs. *Ascaris*, 0.02±0.018, p=0.25; *Q.q*.F vs. *Ascaris*, 0.061±0.0163, p<0.0002).

To determine if the *Mytilus* M lineage was evolved more rapidly than the *Quadrula* M lineage, we tested differences in rate using a 600bp subset of the data set for which homologous *Mytilus* M sequence is available. Based on this limited data set, the *Mytilus* M genome is 8% faster than the *Quadrula* M genome (±0.02; $p = 6.3 \times 10^{-5}$).

Differences in rate between the M and F genomes also are apparent in relative rate tests with *Albinaria coerule*. The difference in rate between *Albinaria* and the *Q. quadrula* M genome is 0.05±0.0169, which is significant at the $p = 0.003$ level. Not surprisingly the small difference in rate between the *Q. quadrula* F genome and *Albinaria* (0.009±0.0151) is not significant. Thus, the *Q. quadrula* M genome follows the pattern observed in *Mytilus* where the relative rate test indicates a higher overall rate of nucleotide substitution as compared with *Albinaria*; in contrast, the *Q. quadrula* F genome does not show an overall greater rate of substitution.

Relative rate tests between the *Quadrula quadrula* M and F genomes and *Katharina tunicata* depart from the pattern observed with *Albinaria*. As compared with *Drosophila*, the *Q. quadrula* M genome has a 9.5% (±1.62%) greater rate of evolution than *K. tunicata* ($p = 4.5 \times 10^{-9}$). The difference in rate between the *Q. quadrula* F genome and *K. tunicata* was slightly over half that of the M genome, at 5.4% (±1.51%), but was still significant ($p = 0.00035$). In these comparisons both relative rate tests are significant, as was the case with the *Mytilus* M and F genomes.

To better understand why the difference in rate between the *Q. quadrula* F genome and *Katharina* is significant but the difference in rate between the *Q. quadrula* F

genome and *Albinaria* is not, a relative rate test was calculated between *Albinaria* and *Katharina* using this data set and the published distances and standard errors from the *Mytilus* data set (Hoeh *et al.* 1996b). Relative rate tests for both data sets are significant (Hoeh *et al.* [1996b], $p = 0.0002$; this study $p = 0.0028$) with an equivalent difference in rate for the two data sets (*Albinaria-Katharina*; Hoeh *et al.* [1996b] 3.2±0.87%; this study 4.5±1.51%).

Discussion

Mytiloid and unionoid bivalves possess an extraordinary mtDNA inheritance system. This system consists of two genomes, one of which is transmitted matrilineally (the F genome) with the other being transmitted patrilineally (the M genome). In comparisons with other metazoans both the M and F genomes exhibited significantly higher rates of evolution for a combined data set of 660 bp of COI and 762 bp of COIII sequence (Hoeh *et al.* 1996b). The significantly greater rate of evolution is hypothesized to be a consequence of DUI: that is, the possession of multiple mitochondrial DNA lineages leads to a "division of labor" and thus relaxed selection on each of the genomes (Hoeh *et al.* 1996b).

The observation that DUI occurs in a separate bivalve lineage, the Unionoida, which has been separated from the Mytiloida for over 450MY (Newell 1969) offers the opportunity for an independent test of the hypothesis of relaxed selection (Hoeh *et al.* 1996b). There are significant differences between the DUI systems of these two bivalve lineages. In mytiloids, DUI frequently breaks down such that capture of an F genome by the paternal lineage is common (Hoeh *et al.* 1996a; Hoeh *et al.* 1997; Quesada *et al.* 1999; Ladoukakis *et al.* 2002). These events are referred to as masculinization events and are evident in both pair crosses of individuals and phylogenetic analyses of mytiloid species (Zouros *et al.* 1994; Hoeh *et al.* 1997; Quesada *et al.* 1999; Ladoukakis *et al.* 2002). These masculinization events have not been observed in the unionoids, the result being that the unionoid lineages are much older (almost two orders of magnitude) than the mytiloid lineages (Hoeh *et al.* 1996a; Curole and Kocher 2002). Regardless of these

differences in functional mechanisms, both bivalve lineages possess two separate mtDNA genomes; thus, if a "division of labor" does lead to relaxed selection this should be evident in both lineages. In order to test this hypothesis, we analyzed 270 bp of COII and 708 bp of COI sequence totaling 970 bp of protein coding sequence.

Neighbor-joining analysis of the *Quadrula quadrula* M and F genomes with several other metazoans produces a phylogeny that is mostly consistent with accepted relationships (Fig. 2). The *Q. quadrula* M and F types cluster by gender and indicate an ancient divergence of the M and F types in concordance with previous analyses (Hoeh *et al.* 1996a; Curole and Kocher 2002). As expected these lineages cluster with other mollusks and in particular cluster with the two gastropods to the exclusion of the chiton. Perhaps not so unexpectedly, the worm *Ascaris suum* also clusters within the mollusks. This unorthodox relationship has been observed previously and has been proposed to be due to long branch attraction (Hoeh *et al.* 1996b). All other taxa cluster according to accepted phylogenetic relationships.

Pairwise distance comparisons indicate that both the *Mytilus* M and F genomes are evolving marginally faster than the *Ascaris* mitochondrial genome, although these comparisons are not significant (Hoeh *et al.* 1996b). The *Q. quadrula* M lineage appears slower than the *Ascaris* lineage, although this difference is marginal when the pairwise distances are averaged (0.268 for *Ascaris* vs. 0.257 for *Quadrula quadrula* M) and a relative rate test is not significant. In contrast, the *Q. quadrula* F genome is slower than the *Ascaris* genome when tested (0.061 ± 0.0163, $p<0.0002$).

This difference in rate is also reflected in relative rate tests between the two *Quadrula* lineages and the gastropod *Albinaria*. The *Quadrula* M lineage is 5% faster

than the *Albinaria* lineage and is significant ($p = 0.003$). The difference in rates between the *Q. quadrula* F lineage and *Albinaria* is less than 1% and not significant. Based on these comparisons the hypothesis that the *Quadrula* M and F lineages are evolving at an equivalent rate can be rejected. This conclusion is consistent with observations from the mytiloids (Skibinski *et al.* 1994b; Rawson and Hilbish 1995; Stewart *et al.* 1996) and unionoids (Liu, Mitton and Wu 1996). In contrast to *Mytilus*, the hypothesis that the M and F lineages are evolving at the same rate as the gastropod *Albinaria* can only be rejected for the M lineage.

Similar to Mytilus, a significant difference in rate for both genomes is observed in relative rate comparisons with the chiton, *Katharina tunicata*. There is a 9.5% difference in rate between *Katharina* and the *Q. quadrula* M genome and a 5.4% difference between *Katharina* and the *Q. quadrula* F genome. This is intriguing given the nearly identical rates of evolution of the F genome and *Albinaria*. Thus, in order to determine if the difference in rate between the *Quadrula/Mytilus* lineages and *Katharina* is specific to DUI clades or this dataset we calculated a relative rate test between *Albinaria* and *Katharina* for both data sets. The COI/COIII data set of Hoeh *et al.* (1996b) and the COI/COII data set of this study indicate a significant difference in rate between *Albinaria* and *Katharina*.

Thus, there appears to be four classes of nucleotide substitution rates for the various molluscan lineages. The fastest class of lineages includes the *Mytilus* M and F lineages. This class is the fastest evolving group of sequences among the metazoans compared and is marginally greater than the rate of nucleotide substitution of *Ascaris*; thus, this class is likely the fastest known among metazoans. The second class consists

solely of the *Quadrula* M lineages and is marginally slower than the *Ascaris* lineage. The third class includes the *Quadrula* F lineage and the gastropods *Albinaria coerule* and *Pupa strigosa*, which are nearly clocklike in their rates of evolution. Indeed there is a less than 1% difference in rate between the *Quadrula* F lineage and *Albinaria*; however, this class is slower than the *Quadrula* M lineage by 5% for the COII-COI data set. The final and slowest class includes *Katharina* whose rate is similar to that of other metazoans and is about 5% slower than the *Quadrula* F/*Albinaria*/*Pupa* class and 10% slower than the *Quadrula* M class for the COII-COI data set.

These data reject the hypothesis that the difference in rate between the *Mytilus* F genome and *Albinaria* is due to relaxed selection on the F genome. If the presence of two mtDNA lineages resulted in a "division of labor" and therefore relaxed selection on each of the two lineages, the two *Quadrula* gender associated mitotypes should both exhibit rates greater than the *Albinaria* lineage. Given that both the *Mytilus* M and F mitotypes have significantly greater rates of nucleotide substitution as compared with *Albinaria* (Hoeh *et al.* 1996b) but only the *Quadrula* M lineage shows the same increased rate, we reject the hypothesis of relaxed selection as a general consequence of DUI.

Rejecting the hypothesis of relaxed selection, the rate differences between the *Mytilus* M and F class, the *Quadrula* M class, the *Quadrula* F/gastropod class and the chiton class remain unexplained. An increased rate of nucleotide substitution for the M genome as compared with the F genome has already been observed in the Unionoida (Liu, Mitton and Wu 1996); in *Mytilus* this difference has been attributed to differences in both non-neutral and neutral mechanisms (Stewart *et al.* 1996; Quesada, Warren and Skibinski 1998) and we consider it likely that these mechanisms are responsible for the

difference reported here (causes of rate differences between the Unionoid M and F lineages will be explored in a separate manuscript). It is possible that classical explanations of differences in rate (*e.g.* effective population size, baseline mutation rate, metabolic rate, generation time) could be responsible for differences between the other classes; however, it would be extraordinary that each of these factors would affect each lineage to the necessary levels to produce similar rates within classes while maintaining differences between classes. Below we explore these hypotheses and propose additional testable hypotheses as causes of the differences in rates between these lineages.

That the *Quadrula* F lineage and the two gastropod lineages exhibit nearly identical rates of nucleotide substitution is intriguing. To our knowledge no gastropod taxon has been rigorously tested for the presence of gender associated mtDNA lineages. Foot tissue of the gastropod *Cepaea nemoralis*, has been PCR sampled for heteroplasmy (Davidson 2000); however, given the relatively low concentration of the M molecule in somatic tissues (Garrido-Ramos *et al.* 1998) and the observed failure of "universal" primers to amplify some M genomes, we do not consider this a rigorous test. A more rigorous test would be DNA-DNA hybridization of ripe gonadal tissue. In light of the difference in rate between the *Quadrula* F lineages and *Katharina* and the nearly identical rates of the *Quadrula* F lineages and the gastropods, a test of DUI in these gastropod lineages is imperative.

The *Mytilus* M and F genomes may be subject to similar life history parameters that are very different from Unionoid life histories. *Mytilus* males and females free spawn gametes (Gosling 1992), whereas unionid females brood larvae (Bauer 2001) and males release sperm in spermatazeugmata (Lynn 1994). *Mytilus* life spans are also short

compared to many unionoids whose life spans can reach 100 years (Gosling 1992 ;Bauer 2000). Estimates of effective population sizes of the two *Mytilus* genomes are inconsistent. Haplotype diversity measured with a PCR/RFLP approach in European *Mytilus galloprovincialis* indicate that the M genome has a smaller effective population size than the F genome (Ladoukakis *et al.* 2002); however, mitotype sequence diversity estimated in European *M. edulis* and *M. galloprovincialis* is greater for the M genome, inconsistent with the hypothesis of a smaller effective population size for the M genome (Quesada, Warren and Skibinski 1998). In particular, haplotype diversity for non-synonymous sites are not significantly different for the M and F genomes, in agreement with the hypothesis of similar effective population sizes for the M and F genomes in these populations; in contrast, haplotype diversity for synonymous sites are significantly greater for the M genome. Assuming equal population sizes for the M and F genomes, life history parameters may explain the differences in rates of nucleotide substitution between these lineages; however, we contend that the differences in rates in these lineages could also be due to the inherent differences in their DUI systems (Hoeh *et al.* 1996a; Curole and Kocher 2002).

In the Unionoida, the M and F lineages have remained separate for hundreds of millions of years (Curole and Kocher 2002); this ancient divergence is reflected in the tree topology in Figure 2. In contrast, in the marine lineage (*Mytilus* and *Geukensia*) gender-switching events appear to be common (Hoeh *et al.* 1996a; Hoeh *et al.* 1997; Quesada *et al.* 1999; Ladoukakis *et al.* 2002). Three of these events are unambiguous masculinization events (events 4, 5 and 6 in Figure 4 of Hoeh *et al.* 1997); although the remaining two events are assumed to have been masculinization events (events 2 and 3 in

Figure 4 of Hoeh *et al.* 1997), it is important to note that these events are ambiguous with respect to polarity. Thus, there is the possibility that feminization events (the capture of the M genome by the female lineage) may have occurred in the history of the mytiloid mussels and have since been erased by the frequent masculinizations. The capture of the M genome by the female lineage would accelerate the relative rate of the F genome; this would be due to the greater rate of evolution of the *Mytilus* M lineage (Skibinski *et al.* 1994; Rawson and Hilbish 1995; Stewart *et al.* 1996). In contrast, the lack of lineage capture in the Unionoida (Curole and Kocher 2002) would prevent acceleration and lead to a relatively slower rate of evolution.

This hypothesis suffers from the lack of observed feminization events. Masculinization appears to be far more common than feminization and it is most parsimonious to assume that if feminization events occur, they occur with a very low frequency. The M type is present in the gonadal tissues of some females, but at a very low frequency and it has not been observed in eggs (Garrido-Ramos *et al.* 1998). Thus, the probability of feminization still remains low.

In contrast, recombination between M and F molecules has been observed in the mussel *Mytilus galloprovincialis* (Ladoukakis and Zouros 2001). Although the effects of recombination would not be genome wide, it is likely that recombination occurs with regular frequency, given that in the ten individuals examined four possessed recombinant sequences and two of these individuals were heteroplasmic for two recombinant haplotypes (Ladoukakis and Zouros 2001). Recombination would not be expected to affect the unionid mitotypes given the large levels of divergence between these molecules

(Ladoukakis and Zouros 2001a, b); the ancient divergence of the unionid mitotypes supports this hypothesis (Curole and Kocher 2002).

Recombination events in *Mytilus* produced haplotypes with a range of 4-14 nucleotide differences as compared with the ancestral haplotypes for 681 bp of COIII sequence. Thus, recombination could effectively speed the evolution of the F genome by introducing substitutions from the M genome. A complication of this hypothesis is that *Mytilus* females are generally homoplasmic for the F genome; males are heteroplasmic but recombinant mitotypes would not be introduced into the F lineage as males only transmit their mitotypes to their sons. Heteroplasmic females have been observed in pair crosses (Zouros *et al.* 1994b) and in natural populations (Skibinski *et al.* 1994b). That a female individual with an M/F recombinant haplotype has been observed supports the hypothesis that recombinant M/F mitotypes, which are significantly divergent from the ancestral F mitotype, can enter into the F lineage (Ladoukakis *et al.* 2002). The apparent increase in substitution rate in the *Mytilus* F genome relative to the *Quadrula* F genome may be the first insight into how recombination between the highly mutated sperm mtDNA contribution (**refs.**) and the maternal mtDNA contribution may affect the rate of evolution of the latter, particularly between organisms with different levels of paternal leakage. As additional bivalve lineages with gender-associated mitochondrial DNA are discovered (Passamonti and Scali 2001), this hypothesis will be subject to more rigorous testing.

Table 4.1. Uncorrected distances (below diagonal) and S.E. (above diagonal) for 1st and 2nd positions of the COII-COI data set.

Albinaria		0.015	0.016	0.017	0.015	0.016	0.016	0.015	0.016	0.016	0.013	0.016	0.017	0.016	0.017	0.016
Anopholes	0.189		0.013	0.017	0.008	0.015	0.015	0.013	0.015	0.014	0.015	0.015	0.017	0.015	0.017	0.014
Artemia	0.21	0.129		0.017	0.014	0.015	0.015	0.015	0.016	0.014	0.015	0.016	0.017	0.016	0.017	0.015
Ascaris	0.243	0.261	0.273		0.017	0.018	0.018	0.017	0.018	0.017	0.017	0.017	0.018	0.017	0.018	0.017
Drosophila	0.183	0.048	0.14	0.253		0.015	0.015	0.014	0.015	0.014	0.016	0.015	0.017	0.015	0.017	0.014
Gallus	0.214	0.167	0.169	0.28	0.167		0.011	0.015	0.016	0.014	0.016	0.016	0.018	0.016	0.018	0.01
Homo	0.22	0.174	0.181	0.278	0.175	0.08		0.015	0.016	0.014	0.016	0.016	0.018	0.016	0.018	0.011
Katharina	0.186	0.127	0.166	0.253	0.138	0.175	0.189		0.015	0.015	0.015	0.016	0.016	0.015	0.016	0.014
Loligo	0.223	0.184	0.209	0.312	0.19	0.212	0.224	0.167		0.016	0.017	0.017	0.018	0.017	0.018	0.016
Paracentrotus	0.206	0.152	0.163	0.267	0.151	0.141	0.14	0.169	0.198		0.015	0.016	0.017	0.016	0.017	0.013
Pupa	0.135	0.187	0.192	0.266	0.195	0.198	0.21	0.192	0.235	0.192		0.016	0.017	0.016	0.017	0.016
Q. quadrula F	0.218	0.194	0.217	0.267	0.192	0.227	0.226	0.195	0.253	0.214	0.212		0.016	0.004	0.016	0.016
Q. quadrula M	0.253	0.238	0.247	0.292	0.233	0.286	0.28	0.217	0.292	0.261	0.255	0.209		0.016	0.005	0.017
Q. refulgens F	0.21	0.192	0.215	0.266	0.194	0.227	0.226	0.19	0.253	0.212	0.209	0.012	0.212		0.016	0.016
Q. refulgens M	0.249	0.238	0.247	0.292	0.237	0.287	0.28	0.212	0.286	0.258	0.253	0.206	0.018	0.203		0.017
Xenopus	0.21	0.149	0.164	0.267	0.151	0.077	0.092	0.163	0.198	0.137	0.204	0.221	0.266	0.22	0.267	

Figure 4.1 Alignment of putative amino acid sequence for the portions of the COII and COI loci examined.

```
Albinaria          YRLLEVDNRPMVPYGLDINVITTSADVIHAWALPSMGVKMDAVPGRLNSMGFHANLPGIYY
Pupa               YRLLEADNRAVVPFGLDTTIISTSADVLHAFAMPSAGVKMDAVPGRLNSMSTLFNRPGVFY
Anopheles          FRLLDVDNRVVLPMNNQIRILVTATDVLHSWTVPSLGVKVDATPGRLNQLNFLINRPGLFF
Drosophila         FRLLDVDNRVILPMNSQIRILVTAADVIHSWTVPALGVKVDGTPGRLNQTNFFINRPGLFY
Artemia            YRLLDVDNRSQCPMIKAIRLMITSDAVLHSWAVPSLGIKMDADPGRLNQSSLLVNMPGVFY
Gallus             FRLLEVDHRIVIPMESPIRVIITADDVLHSWAVPALGVKTDAIPGRLNQTSFITTRPGVFY
Homo               LRLLDVDNRVVLPIEAPIRMMITSQDVLHSWAVPTLGLKTDAIPGRLNQTTFTATRPGVYY
Paracentrotus      PRLLEVDNRLMLPMQNPIRVLVSSADVLHSWAVPSLGVKMDAVPGRLNQTTFFAARAGLFY
Katharina          YRLLEVDHRSVVPMKTKVRVLVTAADVLHSWTVPSLGVKADAVPGRLNQLSFFANYPGVFY
Ascaris            PRLLEVDNRCVVPCDVNIRFCITSGDVIHSWALPSMSIKLDAMSGILSTLSYSFPVVGVFY
Q. refulgens F     YRLLEVDNRCVVPYGVDSRVLISSADVIHAWALPSIGVKVDAIPGRINQLGVHLMGSGVMF
Q. quadrula F      YRLLEVDNRCVIPYGVDSRILVSSADVIHAWALPSIGVKVDAIPGRINQLGVHLIGSGVMF
Q. quadrula M      YRLLEVDNRCIVPSLLQMRGLVTSDDVIHSWAIPSSSVKVDGVPGRINQVGLCFIYSGVFY
Q. refulgens M     YRLLEVDNRCIVPSLLQMRGLVTSDDVIHSWAIPSSSIKVDGVPGRINQVGLCFIYSGVFY
                    ***:.*:*   *         ::  *:*::::*: .:* *. .* :.        *: :

                                        End COII-Begin COI
Albinaria          GQCSEICGANHSFMPITVEAIDVKDFIKM-RWFYSTNHKDIGTLYMLFGIWCGMVGTGLSL
Pupa               GQCSEICGANHSFMPIVIESINLHDFVST-RWLCSTNHKDIGTLYMVFGMWCGLVGTGLSL
Anopheles          GQCSEICGANHSFMPIVIESIPMNYFIKW-QWLFSTNHKDIGTLYFIFGAWAGMVGTSLSI
Drosophila         GQCSEICGANHSFMPIVIESVPVNNFIKW-QWLFSTNHKDIGTLYFIFGAWAGMVGTSLSI
Artemia            GQCSEICGSGHSFMPIVIEAVGESDFLKW-RWFYSTNHKDIGTLYFIFGAWAGMVGTSLSM
Gallus             GQCSEICGANHSYMPIVVESTPLKHFEAW-RWLFSTNHKDIGTLYLIFGTWAGMAGTALSL
Homo               GQCSEICGANHSFMPIVLELIPLKIFEMG-RWLFSTNHKDIGTLYLLFGAWAGVLGTALSL
Paracentrotus      GQCSEICGANHSFMPILMESVPFSNFENW-RWLFSTNHKDIGTLYLIFGAWAGMVGTAMSV
Katharina          GQCSEICGANHSFMPIVLEVVDSSSFIKW-RWIFSTNHKDIGTLYILFGIWAGLVGTALSL
Ascaris            GQCSEICGANHSFMPVALEVTLLDNFKSW-VWLESSNHKDIGTLYFLFGLWSGMVGTSLSL
Q. refulgens F     GQCSEICGVNHSFMPIGLESVSPEVFYCW-RWLCSTNHKDIGTLYLLLALWSGLIGLALSL
Q. quadrula F      GQCSEICGVNHSFMPVGLESVSPEVFYCW-RWLCSTNHKDIGTLYLLLALWSGLIGLALSL
Q. quadrula M      GQCSELCGVNHSFMPVCVEAVSTKVFLNW-RWMFSTNHKDIGTLYMVSSVWWGLVGLSLSI
Q. refulgens M     GQCSELCGVNHSFMPVCVEAVSTKVFLNW-RWMFSTNHKDIGTLYMVSSIWWGLVGLSLSV
                   *****:** .**:**: :*      *     *: *:**********:: . * *: * .:*:
```

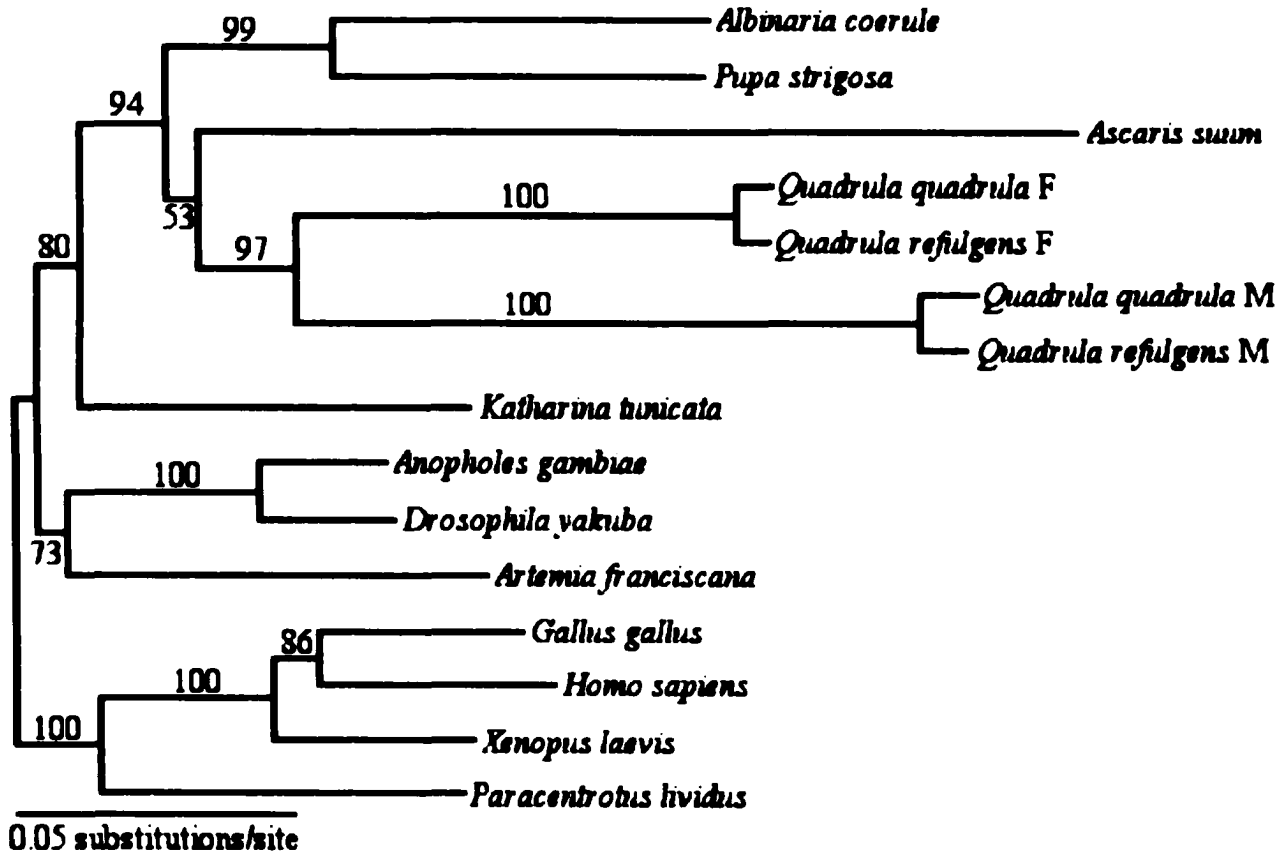
Figure 4.1 Alignment of putative amino acid sequences for the portions of the COII and COI loci examined (cont.).

```
Albinaria          LIRLELGTSGTLT-DDHFYNVIVTAHAFVMIFFMVMPIMIGGFGNWMVPLLIGAPDMSFPR
Pupa               LIRFELGTAGALLGDDHFYNVIVTAHAFVMIFFMVMPLMIGGFGNWMVPLLIGAPDMSFPR
Anopheles          LIRAELGHPGAFIGDDQIYNVIVTAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPR
Drosophila         LIRAELGHPGALIGDDQIYNVIVTAHAFIMIFFMVMPIMIGGFGNWLVPLMLGAPDMAFPR
Artemia            LIRAELGQPGSLIGDEQVYNVIVTAHAFIMIFFMVMPILIGGFGNWLVPIMLGAPDMAFPR
Gallus             LIRAELGQPGTLLGDDQIYNVIVTAHAFVMIFFMVMPIMIGGFGNWLVPLMIGAPDMAFPR
Homo               LIRAELGQPGNLLGNDHIYNVIVTAHAFVMIFFMVMPIMIGGFGNWLVPLMIGAPDMAFPR
Paracentrotus      IIRAELAQPGSLLKDDQIYKVVVTAHALVMIFFMVMPIMMGGFGNWLIPLMIGAPDMAFPR
Katharina          LIRAELGQPGALLGDDQLYNVIVTAHAFVMIFFLVMPMMIGGFGNWLVPLMLGVPDMAFPR
Ascaris            VIRLELAKPGLLLGSGQLYNSVITAHAILMIFFMVMPTMIGGFGNWMLPLMLGAPDMSFPR
Q. refulgens F     LIRAELGQPGSLLGDDQLYNVIVTAHAFMMIFFLVMPMMIGGFGNWLIPLMIGAPDMAFPR
Q. quadrula F      LIRAELGQPGSLLGDDQLYNVIVTAHAFMMIFFLVMPMMIGGFGNWLIPLMIGAPDMAFPR
Q. quadrula M      LIRVELGHPGGVMCSEQLYYSIVTAHAFVMIFFVVMPAMMGGMGNWLVPIMLGSPDMSFPR
Q. refulgens M     LIRVELGHPGGVMRNEQLYYSIVTAHAFVMIFFVVMPAMMGGMGNWLVPIMLGSPDMSFPR
                   :** **. .* .  . :.*  ::****::****:*** ::**:***::*:::* ***:***

Albinaria          MNNMSFWLLPPAFILLICSSMVEGGAGTGWTVYPPLSSSLAHSGASVDLAIFSLHLAGMSS
Pupa               MNNMSFWLLPPSLILLLSSSLMEGGAGTGWTVYPPLSGAMGHTGCSVDLAIFSLHLAGMSS
Anopheles          MNNMSFWMLPPSLTLLISSSMVENGAGTGWTVYPPLSSGIAHAGASVDLAIFSLHLAGISS
Drosophila         MNNMSFWLLPPALSLLLVSSMVENGAGTGWTVYPPLSSGIAHGGASVDLAIFSLHLAGISS
Artemia            LNNLSFWMLPPSLTLLLASSMVESGAGTGWTVYPPLSSAIAHAGPSVDLAIFSLHLAGVSS
Gallus             MNNMSFWLLPPSFLLLLASSTVEAGAGTGWTVYPPLAGNLAHAGASVDLAIFH-YLAGVSS
Homo               MNNMSFWLLPPSLLLLLASAMVEAGAGTGWTVYPPLAGNYSHPGASVDLTIFSLHLAGVSS
Paracentrotus      MNNMSFWLIPPSFILLLASAGVESGAGTGWTMYPPLSSKIAHAGGSVDLAIFSLHLAGASS
Katharina          LNNMSFWLLPPALCLLLASGAVESGAGTGWTVYPPLAGNVGHAGGSVDLAIFSLHLAGVSS
Ascaris            LNNLSFWLLPTAMFLILDACFVDMGCGTSWTVYPPLS-TMGHPGGSVDLAIFSLHCAGVSS
Q. refulgens F     LNNLSFWLLVPALFLLLSSSMVESGVGTGWTVYPPLSGNIAHSGASVDLAIFSLHLAGASS
Q. quadrula F      LNNLSFWLLVPALFLLLSSSMVESGVGTGWTVYPPLSGNIAHSGASVDLAIFSLHLAGASS
Q. quadrula M      LNNVSFWLLMGSGLLLGWSMFVESGCGTGWTIYPPLSSVGYHSGVCMDIVIFSLHLAGASS
Q. refulgens M     LNNVSFWLLMGSGLLLGWSMFVESGCGTGWTIYPPLSSVGYHSGVSMDIVIFSLHLAGASS
                   :**:***::  :  *:  :  :: * **.**:****:    * * .:*:.**  : ** **
```

Figure 4.1 Alignment of putative amino acid sequences for the portions of the COII and COI loci examined (cont.).

```
Albinaria          ILGAINFITTIFNMRSPGMTMERVSLFVWSILVTVFLLLLSLPVLAGAITMLLTDRNFNTS
Pupa               LLGAVNFITTIFNMRAPGMTMERVSLFVWSVLVTAFLLLLSLPVLAGAITMLLTDRNFNTS
Anopheles          ILGAVNFITTVINMRSPGITLDRMPLFVWSVVITAVLLLLSLPVLAGAITMLLTDRNLNTS
Drosophila         ILGAVNFITTVINMRSTGITLDRMPLFVWSVVITALLLLLSLPVLAGAITMLLTDRNLNTS
Artemia            ILGAVNFITTIINMRPQSMSIDRMPLFVWAVGITAVLLLLSLPVLAGAITMLLTDRNLNTS
Gallus             ILGAINFITTIINMKPPALSQYQTPLFVWSVLITAILLLLSLPVLAAGITMLLTDRNLNTT
Homo               ILGAINFITTIINMKPPAMTQYQTPLFVWSVLITAVLLLLSLPVLAAGITMLLTDRNLNTT
Paracentrotus      ILASINFMTTIINMRTPGMSFDRLPLFVWSVFVTAFLLLLSLPVLAGAITMLLTDRKINTT
Katharina          ILGAVNFITTIVNMRSEGMQLERLPLFVWSVKITAILLLLSLPVLAGGITMLLTDRNFNTS
Ascaris            ILGAINFMTTTKNLRSSSISLEHMSLFVWTVFVTVFLLVLSLPVLAGAITMLLTDRNLNTS
Q. refulgens F     ILGAINFISTVGNMRSPGLVAERIPLFVWAVTVTAVLLVAALPVLAGAITMLLTDRNINTS
Q. quadrula F      ILGAINFISTVGNMRSPGLVAERIPLFVWAVTVTAVLLVAALPVLAGAITMLLTDRNINTS
Q. quadrula M      ILGALNFITTILNMRVEALRVERMTLFSWSVLSTAGLVLLSFPVLAGAITMLLTDRNFNTS
Q. refulgens M     ILGALNFITTILNMRVEALRVERMTLFTWSVLSTAGLVLLSFPVLAGAITMLLTDRNFNTS
                   :*.::**::*  *::  .:   : .** *::  *. *:: ::****..********::**:

Albinaria          FFDPAGGGDPILYQHLFWFFGHPEV
Pupa               FFDPAGGGDPILYQHLFWFFGHPEV
Anopheles          FFDPAGGGDPILYQHLFWFFGHPEV
Drosophila         FFDPAGGGDPILYQHLFWFFGHPEV
Artemia            FFDPAGGGDPILYQHLFWFFGHPEV
Gallus             FFDPAGGGDPILYQHLFWFFGHPEV
Homo               FFDPAGGGDPILYQHLFWFFGHPEV
Paracentrotus      FFDPAGGGDPILFQHLFWFFGHPEV
Katharina          FFDPAGGGDPILYQHLFWFFGHPEV
Ascaris            FFDPSTGGNPLIYQHLFWFFGHPEV
Q. refulgens F     FFDPVGGGDPILYMHLFWFFGHPEV
Q. quadrula F      FFDPVGGGDPILYMHLFWFFGHPEV
Q. quadrula M      FFDPSGGGDPILFMHLFWFFGHPWS
Q. refulgens M     FFDPSGGGDPILFMHLFWFFGHPEV
                   ****  **:*::: *********
```

Figure 4.2 Neighbor-Joining phylogeny constructed using p-distances for 1st and 2nd positions at the combined COII and COI data set.



CHAPTER V

SEX-LIMITED MITOCHONDRIAL DNA INHERITANCE IN BIVALVES: A REASSESSMENT AND FUTURE DIRECTIONS

DUI in Bivalves-A Brief Review

The discovery of independently inherited, sex-limited mitochondrial DNA lineages in some bivalves offers rich opportunities to understand the mechanisms of molecular evolution of mtDNA. These opportunities have been seized upon and much data has been collected. As such, we are at a point where prudence dictates that we review and assess this data, determine if conclusions stand and ask, what, if anything does DUI mean.

Doubly-uniparental inheritance was first discovered via population studies of *Mytilus* (Fisher and Skibinski 1990; Hoeh, Blakley and Brown 1991; Zouros *et al.* 1992). Through studies examining the correlation of the two haplotypes with sex and tissue (Skibinski, Gallagher and Beynon 1994a, b) and determination of the inheritance patterns of the two mitotypes in pair crosses (Zouros *et al.* 1994a, b) this unique mode of inheritance was confirmed. Similar to other metazoans, mothers passed their mitotype on to all offspring. Unlike other metazoans, fathers also transmitted their mitotype but only to their sons. The discovery that the freshwater species, *Anodonta*, also exhibited high levels of heteroplasmy (Hoeh, Blakley, and Brown 1991) and the association of this

heteroplasmy with sex strongly suggested that a similar inheritance mechanism was working in the Unionidae (Liu, Mitton and Wu 1996).

During these initial studies it was noted that males were not always heteroplasmic and in rare cases some females were heteroplasmic (Fisher and Skibinski 1990; Zouros *et al.* 1994). Additional studies confirmed that hybridization between individuals from divergent populations led to a breakdown of DUI. These observations laid the groundwork for a model of DUI: after ingression of the sperm into the egg, interactions between the sperm mitochondrial membrane and the egg cytoskeleton determine the fate of paternal mitochondria. The paternal mitochondria are then preferentially replicated in the male germ line. In populations where the egg and sperm mitochondrial factors have diverged adequately, this interaction is diminished, leading to a breakdown of DUI.

Phylogenetic studies indicated that breakdown of DUI is not solely a population level phenomenon. Phylogenetic analysis of M and F COI sequences from three unionid species, two *Mytilus* species and *Geukensia demissa* suggested three independent origins of DUI (Hoeh *et al.* 1996a). Analysis of additional mytilid species indicated that it was likely that these independent origins were actually lineage capture events (Hoeh *et al.* 1997). These lineage capture events, or masculinizations, were the result of the fixation of a female genome that had been captured by the male inheritance system. These data also suggested that although lineage capture events are very common in mytilids they are scarce in the Unionidae; thus, the ancient divergence of the unionid sex-associated mitotypes in contrast to the recent divergence of the mytilid mitotypes (Hoeh *et al.* 1996a).

The separation of the two genomes into independent lineages also allows for independent evolutionary mechanisms. Initial observations suggested that the rate of evolution in the male lineage was two-fold greater than the female lineage (Skibinski, Gallagher and Beynon 1994; Rawson and Hilbish 1995; Liu, Mitton and Wu 1996) and relaxed selection was initially proposed as the causative mechanism (Stewart *et al.* 1995). A more thorough examination suggested that although the difference in rates between the M and F genomes at synonymous sites was not significant, the difference at non-synonymous sites was significant consistent with the hypothesis of relaxed selection (Stewart *et al.* 1996). Relaxed selection was also proposed as the cause of the greater observed rate of evolution of the *Mytilus* F genome relative to other metazoan lineages (Hoeh *et al.* 1996b). Thus, three tiers of selection were proposed: the slowest lineages were those with strict maternal inheritance, followed by the slightly greater rate of the *Mytilus* F genome and the fastest tier being the *Mytilus* M genome.

Mode of inheritance

Through the use of careful pair crosses a genetic model of the mechanism of DUI has been developed. *Mytilus* has proved to be an ideal organism for a mapping approach to understanding the mechanism of DUI. In contrast, the difficulty in rearing unionids and the lack of observed breakdown of DUI in this group suggests that a mapping approach with these organisms would be less fruitful; however, the presence of a uniquely identifiable male specific product (*i.e.* the extension) and the rich developmental work that has been done with this group suggests a qualitative approach.

The *Mytilus* system satisfies the requirements for a quantitative trait locus (QTL) mapping approach. The annual generation time is short for a bivalve, individuals are easily crossed and can be cage reared in the wild, progeny are copious and, most importantly, mutant phenotypes have been observed. In crosses of *M. galloprovincialis* one father sired 44 of the 50 M negative progeny although his sperm were M positive (Saavedra, Reyero and Zouros 1997). If the factor that ensures survival of the M type mitochondrion in the egg is autosomally inherited then crosses of a male with this phenotype should result in F_2 segregating for this factor. Unfortunately, phenotyping of the F_2 requires an additional cross between the F_2 and a female with an alternate mitotype to determine if male offspring retain the paternal mitotype. Thus, undertaking of this project is a formidable task and it is well advised that the penetrance of the original mutant male phenotype be determined from at least a couple of spawnings before a large number of individuals are reared.

The unionid system is refractory to a QTL mapping approach. Generation times are longer, larvae must pass through a parasitic stage on a host fish effectively limiting the number of progeny that can be reared and mutant phenotypes are unknown in this system. The last hurdle may simply be a result of the first two, as pair crosses have not been undertaken in this system with the objective of analyzing of DUI phenotypes. The extensive work in the embryology of the Unionidae may make up for the difficulty in a quantitative approach. A map of cell lineages up to the 45 cell stage is available (Lillie 1895) and recent work on fertilization in *Unio elongatulus* (Rosati *et al.* 2000) may provide a foundation for identification of temporal and spatial location of the sperm mitochondrion. Additionally, the presence of a large protein coding insertion in the unionid M type mitochondrial genome may offer a lead into understanding why a male genome persists in these bivalves.

The available developmental work already suggests interesting differences between the two systems and possible avenues for additional studies. The unionid egg is polarized such that the site of sperm entrance is through a "crater like" region, also known as a micropyle, located at a constant site on the vegetative pole (Lillie 1901; Focarelli, Renieri, and Rosati 1988; Focarelli, Rosa, and Rosati 1990); this is one of the few known exceptions to the rule that the egg surface is generally uniform with no landmarks to identify the vegetative and animal poles (Focarelli *et al.* 2001). This site-specific entrance is ensured by the localized distribution of the primary binding polypeptide to the crater region (Focarelli and Rosati 1995). The sperm path through the micropyle appears constant between eggs, leaving the sperm structure in a consistent position (Lillie 1910). Lillie (1910, p. 230) writes, "The usual path of the spermatozoon

is ... always towards one side of the egg and in the direction of the animal pole. The path is frequently quite direct ..." and later (p. 248), adding, "I am convinced, from its subsequent behavior, that the penetration path of the sperm-head is conditioned by dynamic relations with the egg-substance dependent on the orientation of the latter." These statements strongly suggest that in unionids the egg plays an important role in the localization of the sperm head within the developing embryo. *Mytilus* eggs do not have a micropyle, but instead possess a number of vitelline coat spikes which are evenly distributed on the surface of the egg, resembling the discarded seed pod of a gumball tree. Sperm entrance is not polarized. These differences may require underlying differences in the mechanisms of sperm mitochondrial segregation in the developing zygote. The consistent location of the sperm body in unionids may lend itself to a simple mechanism, whereby the sperm body is localized to the region of the egg destined to become germ plasm. In contrast, active migration of the mytilid sperm mitochondria post fertilization and differential replication of the M genome (Rawson, Secor and Hilbish 1996; Saavedra, Reyero and Zouros 1997) may be required to ensure localization to the gonad. The segregatory path of sperm mitochondria is work that should be undertaken.

Mode of evolution-Mitogenesis

The frequent lineage capture events in the mytilids as compared with the relative absence, or at least persistence, of such events in the unionids also present interesting differences in these two systems. This dichotomy raises several questions: why do masculinization events occur in the mytilids, why are these events unobserved in the unionids and what does it mean for the evolution of the sex limited genomes in these two systems.

Breakdown of DUI at the individual level in *Mytilus* is dependent on two factors, the level of divergence between the two individuals (*e.g.* the level of inter-specific hybridization) and the male genotype (Zouros *et al.* 1994b; Fisher and Skibinski 1990; Rawson, Secor and Hilbish 1996; Saavedra, Reyero and Zouros 1997). The relative propensity of *Mytilus* spp. towards hybridization may be a cause of the frequent invasion of the F genome into the M lineage; however, male specific breakdown of paternal inheritance has also been observed in a homospecific cross of *M. galloprovincialis* and this mode of breakdown may represent an underestimated contribution to invasion. Regardless of the mechanism of invasion into the male lineage, it is expected that once the F type genome is present in the male line of descent, its probability of fixation is dependent on population level forces. If the newly masculinized F genome (hereafter referred to as M') is substantially less fit than the ancestral M genome, then we expect that the M' genome will be selected against and removed from the population. The frequency of fixation of these M' genomes suggests that this is not always the case. At the other end of the spectrum the M' genome may exhibit a greater fitness than the M

genome and would likely become fixed in the population; however, given the high frequency of breakdown of DUI we would expect that all populations examined would be fixed for an M' mitotype if this were the case. The exception might be if the fitness differences between genomes are correlated with the levels of divergence. In this case, the older an M genome the more likely it is to become replaced by an M' mitotype. Equivalent or nearly equivalent fitness levels for the two genomes would place the fixation process in the realm of population drift. Evaluation of fitness differences between sperm with M and M' mitotypes is essential to understanding this process. This is no small task, as our understanding of what phenotypic characteristics are important to sperm fitness is embryonic; however, this work could begin by looking at characteristics that have been shown to be dependent on sperm mitotype, such as sperm motility and oxidative respiration rates (Ruiz-Pesini *et al.* 2000). That the F genome is significantly less fit in the sperm environment, as has been shown in humans (Ruiz-Pesini *et al.* 2000), than the M genome, is unlikely. Despite the fact that, in contrast to standard maternal inheritance, with DUI the F genome is released from any selection might be acting to maintain function in the sperm environment, at least some of the F type mitochondrial genomes maintain adequate fitness for the sperm environment. Whether this indicates that selection pressures for function in the two lineages is largely overlapping remains to be determined. Additional insight into the processes responsible for fixation of M' mitotypes could also come from temporal studies of the relative frequencies of M and M' mitotypes.

In contrast to the mytilids, the objective with the unionids is understanding why M' mitotypes do not become fixed in populations, if they exist at all. The unique mode of

fertilization in the Unionidae may be integral to the fidelity of DUI in this system. If localization of the sperm body to a particular region of the egg results in more effective segregation of the sperm mitochondrion to the germ line, then it is possible that the presence of M mitotype negative males is rare. In *Mytilus*, breakdown of DUI also appears to be strongly correlated with interspecific hybridization. Hybridization may be rare in unionids, again leading to a paucity of males with breakdown phenotypes; evaluation of the fidelity of DUI in known unionid hybridization zones will be important in testing this hypothesis (Hoeh, Stewart and Guttman, submitted). The identification of a relatively large protein coding polymorphism between the M and F genomes of unionoids offers an additional hypothesis (Curole and Kocher 2002). If this sex-specific extension provides a significant fitness advantage to the M genome in the sperm environment, the F mitotype would be selected against should it invade the male transmission route. The extension may function in maintaining segregation of the sperm mitochondria to the germ line or, if the genetic mechanisms of gonadal development are similar in mytilids and unionids, it may be homologous to the mitochondrial factor proposed by Saavedra *et al.* (1997) to be responsible for gonadal development in *Mytilus* males. Elucidating the temporal and spatial expression of the extension will be important to understanding its function and any possible contribution it may make to the absence of masculinization.

Mode of evolution-molecular

Relaxed selection on the M genome

Early in the study of DUI the observation was made that the M mitochondrial genome evolved more rapidly at the molecular level than the F mitochondrial genome. It was suggested that this difference in rate might be due to male driven evolution (*i.e.* differences in the substitution rate of the two genomes) (Rawson and Hilbish 1995) or relaxed selection (Rawson and Hilbish 1995; Stewart *et al.* 1995). In a later paper, Stewart *et al.* (1996) presented three lines of evidence in favor of relaxed selection and suggested that separation of the milieu in which the two mitochondrial genomes function may lead to relaxed selection on the M genome.

Examination of rates of evolution in the freshwater Unionidae does not present as clear a picture (Chapter 3). Inter-specific pairwise comparisons of a portion of the COII and COI loci suggest differences in rates of synonymous substations at COII but not at COI. In contrast, differences in non-synonymous rates are observed at COI but not at COII. These differences are consistent regardless of the model used to evaluate distances (Nei-Gojobori Jukes-Cantor corrected distances or maximum likelihood estimation). It was pointed out that saturation at the COI locus may be masking differences in rates of synonymous substitution; however, this does not explain the F type intra-genomic differences in rates of synonymous substitution between the COII and COI loci. Here we tentatively conclude that rates of synonymous substitution are greater in the M genome by the same factor of two found in *Mytilus* (Rawson and Hilbish 1995; Stewart *et al.*

1996). The examination of synonymous and non-synonymous rates of substitution between more closely related taxa is essential for testing this hypothesis.

In light of our observations and the previous observations that synonymous substitution rates are greater for the *Mytilus* M genome, although not significantly, we have re-examined the data of Stewart *et al.* (1996) as this publication has served as a benchmark in understanding evolution of the M and F genomes. Stewart *et al.* (1996) presented three lines of evidence to suggest that evolution is non-neutral and that relaxed selection may be the mechanism responsible for the increased rate of substitution of the M genome.

The first line of evidence was a comparison of synonymous and non-synonymous rates of substitution between *M. edulis* and *M. trossulus* for the M and F lineages. Although the rate of synonymous substitution for the M genome was nearly two-fold greater than the F genome (2.105 *vs.* 1.237) this difference was not significant. Estimation of synonymous rates was done using the NG method with a Jukes-Cantor correction; in contrast, tests of the difference in synonymous rates are significant using both the NG method with uncorrected p-distances or maximum likelihood (ML) estimation (NG p-distance 0.142 ± 0.0499, $p = 0.0037$; ML 1.1015 ± 0.39, $p = 0.0093$). Thus, it does appear that synonymous rates of substitution are greater in the M genome than the F genome. This is consistent with Quesada, Warren and Skibinski (1998) observations that estimates of population level synonymous haplotype diversity are from 1.5-4 fold greater for the M genome and are overall significantly greater for all populations. This should perhaps serve as a cautionary tale; if conclusions are to be drawn on the failure to reject a null hypothesis, the test of the null hypothesis must have

the appropriate power. It is disconcerting that while a great amount of attention is given to Type I error rates, sometimes to the point where researchers make unnecessary corrections, Type II error rates are practically ignored, although failure to reject a hypothesis is often considered to be a conclusive result.

The second line of evidence of non-neutral evolution comes from McDonald-Kreitman tests of neutrality. Analysis of the M and F COIII sequences indicated a significant excess of fixed replacement substitutions. Unfortunately, this result is an expectation of the McDonald-Kreitman test when comparing highly divergent species (in this case species refers to the M and F lineages) (Maynard Smith 1994; Nachman 1998; Blouin 2000). This results from the rapid saturation of synonymous sites; thus, while fixed non-synonymous differences continue to accumulate between the two lineages synonymous divergence reaches saturation, ultimately leading to an underestimate of the number of fixed synonymous differences. Indeed, McDonald-Kreitman tests of the *Lampsilis* and *Quadrula* COI and COII data sets all give the same result-a highly significant excess of fixed replacement substitutions. It should also be noted that in all of these cases-the *Mytilus* COIII, *Quadrula* and *Lampsilis* COII and COI- that the results of the McDonald-Krietman test indicate a neutrality index of substantially less than 1 (range 0.040-0.237) indicating extremely strong positive selection and not relaxed selection (Rand 2001). Also suggestive of a bias in these comparisons is that they are opposite of tests on more closely related *Mytilus* genomes, where significant comparisons indicate an excess of replacement polymorphism (Quesada, Warren and Skibinski 1998; Quesada, Wenne and Skibinski 1999) consistent with that observed in the mitochondrial genomes

of many other organisms (Rand 2001) and suggestive of a nearly neutral mode of evolution.

The last line of evidence is the estimation of rates of amino acid substitution at sites deemed "variable" by the authors. Variable sites were determined by alignment of several invertebrate COIII sequences (not including the *Mytilus* sequences) and identifying sites at which amino acid identity was not complete. At the 114 codons deemed variable, the F sequences were different at 5 of these sites and the M sequences were different at 19 codons. Although this is evidence of a greater rate of non-synonymous substitution at variable sites, it is by no means solely compatible with the hypothesis of relaxed selection. These observations are consistent with the hypothesis that the replacements occurring at these sites are selectively neutral, or nearly so, and that the observed increase in rate is due to some neutral mechanism.

Given this diminishing support for relaxed selection as the mechanism of rate differences between the M and F genomes, we suggest that additional comparisons are necessary before firm conclusions can be made. The family Unionidae, with its rich species diversity in North America is an ideal system for such comparisons. The key to identifying the appropriate comparisons may come from an initial evaluation of levels of divergence of the F genome. Comparisons between taxa whose F genomes are divergent by less than 10% at synonymous sites (at the COII or COI loci) will be ideal. Thus, it may be prudent to begin by comparing species with similar morphotypes as they appear to exhibit lower levels of divergence (Roe and Lydeard 1998). In addition, the fortuitous discovery of cryptic species may suggest additional taxa for comparison (Curole, unpublished data).

Relaxed selection on the F genome

Relative rate tests between *Mytilus* F/M COI and COIII sequences and other molluscan lineages indicated that both *Mytilus* genomes were evolving at a significantly greater rate (Hoeh *et al.* 1996b). This difference was attributed to the possibility of relaxed selection on the F lineage because of the presence of the M lineage (Hoeh *et al.* 1996b). The presence of the M lineage would relieve the F genome of selection that might occur in the sperm environment (Gemmell and Allendorf 2001).

We examined this hypothesis by comparing relative rates of evolution for two *Quadrula* species using a combined COII/COI dataset (see Chapter 3). Our results indicate that the accelerated rate of evolution observed for the *Mytilus* F sequences is not observed for the *Quadrula* F sequences. Rates of evolution for the *Quadrula* F sequences are nearly identical to those for the gastropods included, but are significantly greater than that of the chiton. In particular, the *Mytilus* sequences exhibit a rate of evolution significantly greater than that of the *Quadrula* M genome. These differences in rate may be the result of a number of factors, including different mutation rates, life history strategies, effective population sizes or metabolic rates. Another hypothesis is that the recently observed phenomenon of recombination between *Mytilus* M and F genomes (Ladoukakis and Zouros 2001a) is a causative factor. Recombination is not expected to occur in unionids because of the levels of divergence between the M and F genomes (Ladoukakis and Zouros 2001a, b). Analysis of rates of evolution and patterns of mitogenesis in the heterodont bivalve *Tapes* may provide additional tests of this hypothesis.

DUI: What does it mean?

Since its discovery, the ultimate reason as to why some bivalves would exhibit separate mtDNA lineages has been of interest. The identification of a large protein coding insertion in the M type unionid mitochondrial genome may provide the first clue to understanding the cause of DUI. Alternatively, this protein coding extension may simply be the result of the ancient separation of these two lineages. The identification of a unionid species with a significantly reduced extension (Curole, unpublished data) does suggest a model for the evolution of DUI, where the presence of an extension is the ancestral condition. Reduction or loss of the extension via transfer to the nuclear genome, a phenomenon known to have occurred throughout the evolution of the mitochondrial genome, leads to a transitory stage where the extension is not present in the mitochondrial genome but DUI remains faithful. Eventually masculinization eliminates any remaining evidence of the presence of the extension, resulting in an M genome that is functionally equivalent to the F genome. Breakdown of DUI is frequent and elimination of the paternal transmission route is not selectively disadvantageous. The implications of this model are extensive, but its foundation is weak; only through careful study of the differences between the mytilid and unionid systems will testing of this model proceed and bring us closer to the answer of the elusive question, "why DUI?"

LITERATURE CITED

Amikura, R., M. Kashikawa, A. Nakamura, and S. Kobayashi. 2001. Presence of mitochondria-type ribosomes outside mitochondria in germ plasm of *Drosophila* embryos. Proceedings of the National Academy of Sciences, USA 98:9133-9138.

Avise, J., B. W. Bowen, T. Lamb, A. B. Meylan, and E. Bermingham. 1992. Mitochondrial DNA evolution at a turtle's pace: evidence for low genetic variability and reduced microevolutionary rate in Testudines. Molecular Biolgy and Evolution 9:457-473.

Bauer, G. 2001. Life-history variation on different taxonomic levels of naiads. Pp. 83-91 *in* G. Bauer, and K. Wachtler, eds. Ecology and Evolution of the Freshwater Unionoida. Springer, Berlin.

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

Berry, D. A., and B. W. Lindgren. 1996. Statistics: Theory and Methods. Wadsworth Publishing Company, Belmont, CA.

Bielawski, J. P., and J. R. Gold. 1996. Unequal synonymous substitution rates within and between two protein coding mitochondrial genes. Molecular Biolgy and Evolution 13:889-892.

Bielawski, J. P., and J. R. Gold. 2002. Mutation patterns of mitochondrial H- and L-strand DNA in closely related cyprinid fishes. Genetics 161:1589-1597.

Blouin, M. S. 2000. Neutrality tests on mtDNA: unusual results from nematodes. The Journal of Heredity 91:156-158.

Bohossian, H. B., H. Skaletsky, and D. C. Page. 2000. Unexpectedly similar rates of nucleotide substitution found in male and female hominids. Nature 406:622-625.

Brown, J. R., K. Beckenbach, A. T. Beckenbach, and M. J. Smith. 1996. Length variation, heteroplasmy and sequence divergence in the mitochondrial DNA of four species of sturgeon (Acipenser). Genetics 142:525-535.

Curole, J. P., and T. D. Kocher. 2002. Ancient sex-specific extension of the cytochrome *c* oxidase II gene in bivalves and the fidelity of doubly-uniparental inheritance. Molecular Biology and Evolution 19:1323-1328.

Davidson, A. 2000. The inheritance of divergent mitochondria in the land snail, *Cepaea nemoralis*. Journal of Molluscan Studies 66:143-147.

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

Dunn, K. A., J. P. Bielawski, and Z. Yang. 2001. Substitution rates in Drosophila nuclear genes: implications for translational selection. Genetics 157:295-305.

Ellegren, H., and A.-K. Fridolfsson. 1997. Male-driven evolution of DNA sequences in birds. Nature Genetics 17:182-184.

Fay, J. C., and C.-I. Wu. 2001. The neutral theory in the genomic era. Current Opinion in Genetics and Development 11:642-646.

Fisher, C., and D. O. F. Skibinski. 1990. Sex-biased mitochondrial DNA heteroplasmy in the marine mussel *Mytilus*. Proceedings of the Royal Society of London. Series B: Biological Sciences 242:149-156.

Fitch, W. M., and E. Margoliash. 1967. A method for estimating the number of invariant amino acid coding positions in gene using cytochrome *c* as a model case. Biochemical Genetics 1:65-71.

Focarelli, R., G. B. La Sala, M. Balasini, and F. Rosati. 2001. Carbohydrate-mediated sperm-egg interaction and species specificity: a clue from the *Unio elongatus* model. Cell Tissues Organs 168:76-81.

Focarelli, R., T. Renieri, and F. Rosati. 1988. Polarized site of sperm entrance in the egg of a freshwater bivalve, *Unio elongatus*. Developmental Biology 127:443-451.

Focarelli, R., D. Rosa, and F. Rosati. 1990. Differentiation of the vitelline coat and the polarized site of sperm entrance in the egg of *Unio elongatus* (Mollusca, Bivalvia). The Journal of Experimental Zoology 254:88-96.

Focarelli, R., and F. Rosati. 1995. The 220-KDa vitelline coat glycoprotein mediates sperm binding in the polarized egg of *Unio elongatus* through O-linked oligosaccharides. Developmental Biology 171:606-614.

Folmer, O., M. Black, W. R. Hoeh, R. Lutz, and R. C. Vrijenhoek. 1994. DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3:294-299.

Frohman, M., M. K. Dush, and G. R. Martin. 1988. Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. Proceedings of the National Academy of Sciences, USA 85:8998-9002.

Gach, M. H., and W. M. Brown. 1997. Characteristics and distribution of large tandem duplications in Brook Stickleback (*Culaea inconstans*) mitochondrial DNA. Genetics 145:383-394.

Garrido-Ramos, M. A., D. T. Stewart, B. W. Sutherland, and E. Zouros. 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. Molecular Marine Biology and Biotechnology 3:334-337.

Gemmell, N. J., and F. W. Allendorf. 2001. Mitochondrial mutations may decrease population viability. Trends in Ecology and Evolution 16:115-117.

Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Molecular Biolgy and Evolution 11:725-736.

Gosling, E. 1992. Mussel Mytilus: ecology, physiology, genetics and culture. Elsevier, New York.

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

Haas, F. 1969. Superfamily UNIONACEA. Pp. N411-N470 *in* R. C. Moore, ed. Treatise on Invertebrate Paleontology: Mollusca 6. Geological Survey of America and University of Kansas, Kansas.

Hafner, M. S., P. D. Sudman, F. X. Villiblanca, T. A. Spradling, J. W. Demastes, and S. A. Nadler. 1994. Disparate rates of molecular evolution in cospeciating hosts and parasites. Science 265:1087-1090.

Harrison, R. G., D. M. Rand, and W. C. Wheeler. 1985. Mitochondrial DNA size variation within individual crickets. Science 228:1446-1448.

Hisatomi, O., T. Satoh, L. K. Barthel, D. L. Stenkamp, P. A. Raymond, and F. Tokunaga. 1996. Molecular cloning and characterization of the putative ultraviolet-sensitive visual pigment of goldfish. Vision Research 36:933-939.

Hoeh, W. R., K. H. Blakley, and W. M. Brown. 1991. Heteroplasmy suggest limited biparental inheritance of *Mytilus* mitochondrial DNA. Science 251:1488-1490.

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

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

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

Hurst, L., and H. Ellegren. 1998. Sex biases in the mutation rate. Trends in Genetics 14:446-452.

Iida, T., and S. Kobayashi. 1998. Essential role of mitochondrially encoded large rRNA for germ-line formation in *Drosophila* embryos. Proceedings of the National Academy of Sciences, USA 95:11274-11278.

Kaneda, H., J. Hayashi, S. Takahama, C. Taya, K. F. Lindahl, and H. Yonekawa. 1995. Elimination of paternal mitochondrial DNA in intraspecific crosses during early mouse embryogenesis. Proceedings of the National Academy of Sciences, USA 92:4542-4546.

Kashikawa, M., R. Amikura, and S. Kobayashi. 2001. Mitochondrial small ribosomal RNA is a component of germinal granules in *Xenopus* embryos. Mechanisms of Development 101:71-77.

Kocher, T. D., and K. C. Carleton. 1997. Base substitution in fish mtDNA: patterns and rates. Pp. 13-24 *in* T. D. Kocher, and C. A. Stepien, eds. Molecular Systematics of Fishes. Academic Press, San Diego.

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

Kumar, S., K. Tamura, I. B. Jakobsen, and M. Nei. 2001. MEGA2: Molecular Evolutionary Genetics Analysis. Arizona State University, Tempe, Arizona, USA.

Ladoukakis, E. D., C. Saavedra, A. Magoulas, and E. Zouros. 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. Molecular Ecology 11:755-769.

Ladoukakis, E. D., and E. Zouros. 2001a. Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. Molecular Biolgy and Evolution 18:1168-1175.

Ladoukakis, E. D., and E. Zouros. 2001b. Recombination in animal mitochondrial DNA: evidence from published sequences. Molecular Biology and Evolution 18:2127-2131.

Li, W.-H. 1993. So, what about the molecular clock hypothesis? Current Opinion in Genetics and Development 3:896-901.

Li, W.-H. 1997. Molecular Evolution. Sinauer Associates, Sunderland, MA.

Li, W.-H., D. L. Ellsworth, J. Krushkal, B. H.-J. Chang, and D. Hewett-Emmett. 1996. Rates of Nucleotide Substitution in Primates and Rodents and the Generation-Time Effect Hypothesis. Molecular Phylogenetics and Evolution 5:182-187.

Lillie, F. 1895. The embryology of the Unionidae. Journal of Morphology 10:1-100.

Lillie, F. 1901. The organization of the egg of Unio, based on a study of its maturation, fertilization, and cleavage. Journal of Morphology 17:227-293.

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

Longo, F. J., and E. Anderson. 1969. Cytological aspects of fertilisation in the lamellibrach, *Mytilus edulis* II. Development of the male pronucleus and the association of the maternally and paternally derived chromosomes. Journal of Experimental Zoology 172:97-120.

Lynn, J. W. 1994. The ultrastucture of the sperm and motile spermatozeugmata released from the freshwater mussel *Anodonta grandis* (Mollusca, Bivalvia, Unionidae). Canadian Journal of Zoology 72:1452-1461.

Magoulas, A., and E. Zouros. 1993. Restriction-site heteroplasmy in Anchovy (*Engraulis encrasicolus*) indicates incidental biparental inheritance of mitochondrial DNA. Molecular Biology and Evolution 10:319-325.

Makova, K. D., and W.-H. Li. 2002. Strong male-driven evolution of DNA sequences in humans and apes. Nature 416:624-626.

Martin, A., G. J. P. Naylor, and S. R. Palumbi. 1992. Rates of mitochondrial DNA evolution in sharks are slow compared with mammals. Nature 357:153-155.

Martin, A., S. R. Palumbi. 1993. Body size, metabolic rate, generation time and the molecular clock. Proceedings of the National Academy of Sciences, USA. 90:4087-4091.

Maynard Smith, J. 1994. Estimating selection by comparing synonymous and substitutional changes. Journal of Molecular Evolution 39:123-128.

Nachman, M. W. 1998. Deleterious mutations in animal mitochondrial DNA. Genetica 102/103:61-69.

Newell, N. D. 1969. Subclass PALAEOHETERODONTA. Pp. N393-N489 *in* R. C. Moore, ed. Treatise on Invertebrate Paleontology: Mollusca 6. Geological Society of America, University of Kansas.

Nielsen, R., and Z. Yang. 1998. Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148:929-936.

Ogawa, M., R. Amikura, K. Akasaka, T. Kinoshita, S. Kobayashi, and T. Shimada. 1999. Asymmetrical distribution of mitochondrial rRNA into small micromeres of sea urchin embryos. Zoological Science 16:445-451.

Ohara, O., R. L. Dorit, and W. Gilbert. 1989. One-sided polymerase chain reaction: the amplification of cDNA. Proceedings of the National Academy of Sciences, USA 86:5673-5677.

Oka, T., R. Amikura, S. Kobayashi, H. Yamamoto, and H. Nishida. 1999. Localization of mitochondrial large ribosomal RNA in myoplasm of the early ascidian embryo. Development, Growth and Differentiation 41:1-8.

Passamonti, M., and V. Scali. 2001. Gender-associated mitochondrial DNA heteroplasmy in the venerid clam *Tapes philippinarum* (Mollusca: Bivaliva). Current Genetics 39:117-124.

Pitnick, S., and T. L. Karr. 1998. Paternal products and by-products in *Drosophila* development. Proceedings of the Royal Society of London. Series B: Biological Sciences 265:821-826.

Quesada, H., D. A. G. Skibinski, and D. O. F. Skibinski. 1996. Sex-biased heteroplasmy and mitochondrial DNA inheritance in the mussel *Mytilus galloprovincialis* Lmk. Current Genetics 29:423-426.

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

Quesada, H., R. Wenne, and D. O. F. Skibinski. 1999. Interspecies transfer of female mitochondrial DNA is coupled with role-reversals and departure from neutrality in the mussel *Mytilus trossulus*. Molecular Biology and Evolution 16:655-665.

Rand, D. M. 2001. The units of selection on mitochondrial DNA. Annual Review of Ecology and Systematics 32:415-448.

Rawson, P. D., and T. J. Hilbish. 1995. Evolutionary relationships among the male and female mitochondrial DNA lineages in the *Mytilus edulis* species complex. Molecular Biology and Evolution 12:893-901.

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

Reyes, A., C. Gissi, G. Pesole, and C. Saccone. 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Molecular Biology and Evolution 15:957-966.

Roe, K. J., and C. Lydeard. 1998. Molecular systematics of the freshwater mussel genus *Potamilus* (Bivalvia: Unionidae). Malacologia 39:195-205.

Rosati, F., A. Capone, C. Della Giovampaola, C. Brettoni, and R. Focarelli. 2000. Sperm-egg interaction at fertilization: glycans as recognition signals. International Journal of Developmental Biology 44:609-618.

Ruiz-Pesini, E., A.-C. Lapena, C. Diez-Sanchez, A. Perez-Martos, J. Montoya, E. Alvarez, M. Diaz, A. Urries, L. Montoro, M. J. Lopez-Perez, and J. A. Enriquez. 2000. Human mtDNA haplogroups associated with high or reduced spermatozoa motility. American Journal of Human Genetics 67:682-696.

Saavedra, C., M.-I. Reyero, and E. Zouros. 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., D. T. Stewart, R. R. Stanwood, and E. Zouros. 1997. Species-specific segregation of gender-associated mitochondrial DNA types in an area where two mussel species (*Mytilus edulis* and *Mytilus trossulus*) hybridize. Manuscript.

Sato, K., T. Sugita, K. Kobayashi, K. Fujita, T. Fujii, T. Mastumoto, T. Mikami, N. Nishizuka, S. Nishizuka, K. Shojima, M. Suda, G. Taka, H. Himeno, A. Muto, and S. Ishida. 2001. Localization of mitochondrial ribosomal RNA on the chromatid bodies of marine planarian polyclad embryo. Development, Growth and Differentiation 43:107-114.

Shimmin, L. C., B. H.-J. Chang, and W.-H. Li. 1993. Male-driven evolution of DNA sequences. Nature 362:745-747.

Shitara, H., J. I. Hayashi, S. Takahama, H. Kaneda, and H. Yonekawa. 1998. Maternal inheritance of mouse mtDNA in interspecific hybrids. Segregation of the leaked paternal mtDNA followed by the prevention of subsequent paternal leakage. Genetics 148:851-858.

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

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

Solignac, M., M. Monnerot, and J.-C. Mounolou. 1983. Mitochondrial DNA heteroplasmy in *Drosophila mauritiana*. Proceedings of the National Academy of Sciences, USA 80:6942-6946.

Stewart, D. T., E. R. Kenchington, R. K. Singh, and E. Zouros. 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., C. Saveedra, R. R. Stanwood, A. O. Ball, and E. Zouros. 1995. Male and female mitochondrial DNA lineages in the blue mussel (*Mytilus edulis*) species group. Molecular Biology and Evolution 12:735-747.

Sutherland, B., D. Stewart, E. R. Kenchington, and E. Zouros. 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.

Swofford, D. L. 2000. PAUP*. Phylogenetic analysis using parsimony (* and other methods). Sinauer Associates, Sunderland, MA.

Tanaka, M., and T. Ozawa. 1994. Strand asymmetry in human mitochondrial mutations. Genomics 22:327-335.

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

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

Williams, J. D., J. Melvin L Warren, K. S. Cummings, J. L. Harris, and R. J. Neves. 1992. Conservation Status of Freshwater Mussels of the United States and Canada. Fisheries 18:6-22.

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

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. CABIOS 13:555-556.

Yang, Z., R. Nielsen, N. Goldman, and A.-M. K. Pederson. 2000. Codon-substitution models for heterogeneous selection pressure at amino acid sites. Genetics 155:431-449.

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

Zouros, E., A. O. Ball, C. Saavedra, and K. R. Freeman. 1994b. An unusual type of mitochondrial DNA inheritance in the blue mussel *Mytilus*. Proceedings of the National Academy of Sciences, USA 91:7463-7467.

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