ANALYSIS OF FEMALE-TRANSMITTED MITOCHONDRIAL DNA OPEN-READING FRAMES IN THE FRESHWATER MUSSEL GENUS *PYGANODON*

A thesis submitted to the
Kent State University Honors College
in partial fulfillment of the requirements
for General Honors

by

Andrew Ruminas

May, 2011
Thesis written by

Andrew Ruminas

Approved by

_______________________________________________________, Advisor

_____________________________________, Chair, Department of Biological Sciences

Accepted by

_______________________________________________, Dean, Honors College
TABLE OF CONTENTS

LIST OF FIGURES .................................................................................................................. v

LIST OF TABLES ..................................................................................................................... vi

ACKNOWLEDGMENTS ............................................................................................................. vii

CHAPTERS

I. INTRODUCTION .................................................................................................................. 1

II. MATERIALS AND METHODS .......................................................................................... 7

   Populations Sampled ............................................................................................................ 7

   Amplification of the F-orf ................................................................................................... 9

   Sequence Analysis ............................................................................................................. 10

   Evolutionary Analysis ....................................................................................................... 11

III. RESULTS ........................................................................................................................... 13

   Nucleotide Analysis .......................................................................................................... 13

   Protein Analysis ................................................................................................................. 16
Evolutionary Analysis ......................................................................................... 26

IV. Discussion ........................................................................................................... 29

Phylogenetic Implications for Pyganodon Lineage ................................................. 29

Diversifying Selection Acting on the F-orf............................................................... 33

F-ORF Protein Predictions ...................................................................................... 35

REFERENCES CITED ............................................................................................... 39

ABSTRACT ................................................................................................................. 44
LIST OF FIGURES

Figure 1. Maximum-likelihood phylogenetic tree..................................................14

Figure 2. Predicted F-ORF protein motifs .............................................................20

Figure 3. Snake-like diagrams and hydropathy plots.............................................21-23

Figure 4. Examples of F-ORF in Pyganodon species...........................................25

Figure 5. MEC evolutionary model analysis of codon selection............................28
LIST OF TABLES

Table 1. *Pyganodon* sampling sites.................................................................8

Table 2. Average nucleotide sequence composition.................................15

Table 3. Pairwise nucleotide sequence p-distances within and between species....15

Table 4. Pairwise nucleotide and amino acid p-distances for two clades of *P. grandis* ........................................................................................................16

Table 5. Amino acid compositions for Fcox2, Fcox1 and F-ORF sequences........18

Table 6. Pairwise amino acid sequence p-distances within and between species...19
ACKNOWLEDGMENTS

I would like to thank Dr. Walter R. Hoeh for being my instructor, advisor and mentor throughout the past two years. He has been extremely helpful, insightful and patient as he has guided me through the processes of performing biological research.

I would like to thank Dr. Sophie S. Breton for her invaluable assistance and patience in the laboratory.

I would also like to thank Dr. Jennifer Marcinkiewicz, Dr. Arne Gericke and Dr. Brett Ellman for serving as my defense committee.
**Introduction**

The mitochondrial genomics of the Bivalvia (Mollusca) has become a significant subject of interest in the study of animal evolutionary genetics and cellular biology (e.g, Doucet-Beaupre et al. 2010). The mitochondria of eukaryotic cells are the sites where cellular respiration occurs through oxidative phosphorylation (Scheffler 2008). Due to the mitochondria’s possession of a separate, distinct sequence of DNA (referred to as mitochondrial DNA or mtDNA), these organelles have long been seen as an important source of evidence concerning the evolutionary history of eukaryotes (Moritz et al. 1987; Gissi et al. 2008). The mtDNA in animals has been observed to be inherited strictly from the mother of an organism; this process is called Strict Maternal Inheritance or SMI (Birky 2001). Three orders of the Bivalvia, i.e. freshwater mussels (*Unionoida*), marine clams (*Veneroida*) and marine mussels (*Mytiloida*), present an exception to this rule.

In bivalve orders, both maternal and paternal inheritance of the mt genome has been observed, a phenomenon described as Doubly Uniparental Inheritance or DUI (Hoeh et al. 1991, 1996; Zouros et al. 1994, 2000; Skibinski et al. 1994). These organisms have been observed to possess distinct maternal and paternal mt genomes present in different tissues dependent upon the gender of the animal. The maternal mt genome (F genome) is transmitted by the mother to her offspring and is present in the somatic tissues of males and females, as well as in the female gonadal tissue. The paternal mt genome (M genome) is only transmitted by the father to his sons and is
principally found in the male gonads (Obata et al. 2005, 2007; Chakrabarti et al. 2009; Kenchington et al. 2009). It is thought that these distinct mt genomes could play a role in sex-determination, a phenomenon that has not been previously demonstrated in animals (Cao et al. 2004; Obata et al. 2005, 2007; Cogswell et al. 2006; Chakrabarti et al. 2007).

Published research on unionoid bivalve DUI has revealed a high degree of nucleotide sequence divergence (>40%) between the M and F genomes of each species and slight differences in gene order, while both mt genomes possess an almost identical gene content (Breton et al. 2009; Doucet-Beaupre et al. 2010). The M genome appears to display a more accelerated rate of molecular evolution compared to the F genome (Breton et al. 2009). Males, who possess the M genome in their gonads and the F genome in their somatic tissues, display the greatest amount of intra-individual mitochondrial divergence ever observed in an animal between their mt genomes. This difference between the F and M genomes within males is even greater than the amount of divergence required when a mtDNA genome is replaced by a different mtDNA genome to impair mt respiration in model systems for the study of mitonuclear co-evolution (Breton et al. 2007). This appears counter-intuitive to the nature of the mt genome, which must be evolutionarily “aligned” with nucleus-encoded proteins to maintain proper functionality and ensure the organism’s survival.

Identifying gender-specific regions of unionoid M and F genomes is crucial to understanding how this DUI system has been maintained for >100 million years and the mechanisms underlying this system of inheritance. Studies on mytilloid F and M genomes have led to the hypothesis that gender-specific control sequences of the mt genomes
interact in the regulation of maternal and paternal inheritance in early development (Cao et al. 2004; Obata et al. 2005; Cogswell et al. 2006; Kenchington et al. 2009). These results have shown a differential segregation pattern of sperm mitochondria (bearing M genomes) immediately following fertilization. The patterns are strongly correlated with the sex bias of the fertilized mother (i.e. mothers biased to produce more sons or more daughters), suggesting that the distinct mt genomes play a role in the determination of the progeny’s sex (Cao et al. 2004; Obata et al. 2005; Cogswell et al. 2006; Kenchington et al. 2009).

Two gender-specific open-reading frames (ORFs) have been identified in unionoids that potentially interact in the regulation of maternal and paternal inheritance; the F genome-specific sequence has been termed the F-orf and the M genome-specific sequence is termed the M-orf (Breton et al. 2009). The identification of these ORFs is significant, as it is unprecedented in animals to detect new-to-science, ancient protein-coding genes in the mt genome other than those in the typical set of 13 (e.g., Gissi et al. 2008).

The F-ORF protein product is of considerable interest in understanding how the DUI system works. All unionoids surveyed to date possess this protein due to the universality of F genome inheritance (Breton et al. 2009). This protein is predicted to possess a single, conserved trans-membrane helix (TMH) immediately followed by a conserved casein kinase II phosphorylation motif on the C-terminus’-side of the TMH (Breton et al. 2009). Significantly, the F-orf protein has also been identified in V.
*ellipsiformis* as being expressed not only in mitochondria, but in the nuclear membrane and nucleoplasm as well (Breton et al. 2010).

Breton et al. (2010) surveyed a representative sample of unionoid species with different breeding systems, i.e., dioecy and hermaphroditism, and found that macromutations in the F-*orf* gene, absence of M genomes and hermaphroditism were absolutely correlated. Given these results, Breton et al. (2010) predicted that the F-*ORF* protein to be involved in a genetic regulatory network specifying dioecy. This is analogous to the sex-determining, mtDNA-encoded proteins found in angiosperm plants that exhibit cytoplasmic male sterility or CMS (Schnable and Wise 1998; Chase 2007; Delph et al. 2007). In these plants, mtDNA-encoded proteins can negatively affect pollen development, thus transforming hermaphroditic individuals into functional females (Schnable and Wise 1998; Chase 2007; Delph et al. 2007).

To date, no study has focused on the evolution of the F-*ORF* protein and how selection acts upon it. To investigate this phenomenon and gain a greater understanding of the F-*ORF*’s functional significance, three species from the unionoid genus *Pyganodon* (*P. fragilis, P. grandis, P. lacustris*) were chosen for analysis. This genus is of particular interest due to the difficulty in identifying these organisms based on morphological and reproductive characteristics (Clarke and Rick 1963; Kat 1983, 1986; Hanlon and Smith 1999; Cyr et al. 2007).

The goals of this study were three-fold. First, it aimed to generate an updated phylogeny of the genus *Pyganodon* using concatenated *Fcox2-Fcox1* sequences to study the evolution of the F-*orf* in this genus. *Fcox2-Fcox1* refers to the F mt genome’s
cytochrome c oxidase subunits I and II. Cytochrome c oxidase is an inner mt membrane enzyme that is involved in the last step of the electron transport chain, facilitating the transfer of electrons to oxygen (Scheffler 2008). Subunit II has been shown to have distinct structures and functionalities in the F and M genomes, with both the F and M subunits displaying extra-mt functionality (Chakrabarti et al. 2007, 2009). The null hypothesis for this phylogeny was a topology that described the Pyganodon species as distinct, monophyletic groups stemming from a common ancestor that possessed the F-orf.

Secondly, the study aims to evaluate the molecular evolutionary processes affecting the F-orf. This is important to investigate the possibility of reproductive functionality for its protein. Proteins that have a reproductive function are known to be affected by rapid positive (=diversifying) selection (Swanson and Vacquier 2002a, 2002b; Swanson et al. 2003). If codons in the F-orf are undergoing diversifying selection, this represents additional evidence in favor of the Breton et al. (2010) hypothesis that the F-ORF protein is involved in the regulatory network that determines an offspring’s sex.

The last goal of this study was to investigate the hypothesized presences of a single, conserved TMH in the N-terminus’-half of the F-ORF proteins from Pyganodon and to predict other F-ORF protein motifs. The extra-mitochondrial presence of the F-ORF protein in the nuclear membrane and nucleoplasm of V. ellipsiformis, in Breton et al. (2010) suggests a TMH structure as well as a genetic regulatory function. Predictions of conserved TMHs and other motifs based on translated F-ORF amino acid sequences would support the hypothesis of conserved functionality. While this current study could
not confirm the existence of any protein structures, predictions of protein motifs will assist in the investigation of the functionality of the F-\textit{ORF}.
Materials and Methods

Populations Sampled

Males and females, representing three *Pyganodon* species, have been manually collected in northeastern North America from different populations in Michigan, Minnesota, Newfoundland (Canada) and Wisconsin (Table 1): *P. lacustris* (n=32), *P. grandis* (n=35) and *P. fragilis* (n=4). Their tissues were either frozen using liquid nitrogen in the field or whole mussels were preserved in 95–100% non-denatured ethanol in the field. Mussel sex was determined through microscopic examination of gonadal tissues by the presence of eggs or sperm/sperm morulae. Dissected tissues samples were stored either at –80°C or in 95–100% ethanol.
<table>
<thead>
<tr>
<th>No.</th>
<th>Exact Site</th>
<th>Region</th>
<th>State/Province (Country)</th>
<th>Lat.</th>
<th>Long.</th>
<th>Species (N)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Elephant L.</td>
<td>St. Louis Co.</td>
<td>Minnesota (USA)</td>
<td>48.1932</td>
<td>-92.7446</td>
<td>La (2)</td>
</tr>
<tr>
<td>2</td>
<td>Black Duck L.</td>
<td>St. Louis Co.</td>
<td>Minnesota (USA)</td>
<td>48.2034</td>
<td>-92.8193</td>
<td>La (6)</td>
</tr>
<tr>
<td>3</td>
<td>Pelican L.</td>
<td>St. Louis Co.</td>
<td>Minnesota (USA)</td>
<td>48.0595</td>
<td>-92.9098</td>
<td>La (4) Gr (3)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gr (1)</td>
</tr>
<tr>
<td>4</td>
<td>Deer L.</td>
<td>Itasca Co.</td>
<td>Minnesota (USA)</td>
<td>47.3807</td>
<td>-93.6635</td>
<td>La (3) Gr (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La (1)</td>
</tr>
<tr>
<td>5</td>
<td>Pfeiffer L.</td>
<td>St. Louis Co.</td>
<td>Minnesota (USA)</td>
<td>47.7485</td>
<td>-92.4789</td>
<td>La (2)</td>
</tr>
<tr>
<td>6</td>
<td>Prairie R.</td>
<td>Itasca Co.</td>
<td>Minnesota (USA)</td>
<td>47.2391</td>
<td>-93.4821</td>
<td>La (2) Gr (7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La (1) Gr (1)</td>
</tr>
<tr>
<td>7</td>
<td>Rice R.</td>
<td>Aitkin Co.</td>
<td>Minnesota (USA)</td>
<td>45.7219</td>
<td>-94.2194</td>
<td>La (2) Gr (2)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gr (1)</td>
</tr>
<tr>
<td>8</td>
<td>Twin lakes</td>
<td>Itasca Co.</td>
<td>Minnesota (USA)</td>
<td>45.2483</td>
<td>-94.2126</td>
<td>La (1)</td>
</tr>
<tr>
<td>9</td>
<td>Frontenac</td>
<td>Goodhue Co.</td>
<td>Minnesota (USA)</td>
<td>44.5269</td>
<td>-92.3297</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gr (1)</td>
</tr>
<tr>
<td>10</td>
<td>Bowstring R.</td>
<td>Itasca Co.</td>
<td>Minnesota (USA)</td>
<td>47.4972</td>
<td>-93.7435</td>
<td>La (2)</td>
</tr>
<tr>
<td>11</td>
<td>Mendota L.</td>
<td>Dane Co.</td>
<td>Wisconsin (USA)</td>
<td>43.95</td>
<td>-89.43</td>
<td>-</td>
</tr>
<tr>
<td>12</td>
<td>Chetak L.</td>
<td>Sawyer Co.</td>
<td>Wisconsin (USA)</td>
<td>45.7034</td>
<td>-91.4948</td>
<td>Gr (3)</td>
</tr>
<tr>
<td>13</td>
<td>Kentuck L.</td>
<td>Vilas Co.</td>
<td>Wisconsin (USA)</td>
<td>45.9845</td>
<td>-89.0003</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La (3)</td>
</tr>
<tr>
<td>14</td>
<td>Annabelle L.</td>
<td>Vilas Co.</td>
<td>Wisconsin (USA)</td>
<td>46.2154</td>
<td>-89.6843</td>
<td>La (1) Gr (2)</td>
</tr>
<tr>
<td>15</td>
<td>St Germain R.</td>
<td>Vilas Co.</td>
<td>Wisconsin (USA)</td>
<td>45.9147</td>
<td>-89.5275</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gr (4)</td>
</tr>
<tr>
<td>16</td>
<td>N Twin L.</td>
<td>Vilas Co.</td>
<td>Wisconsin (USA)</td>
<td>44.2338</td>
<td>-89.1408</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>La (1) Gr (2)</td>
</tr>
<tr>
<td>17</td>
<td>Prairie L.</td>
<td>Calhoun Co.</td>
<td>Michigan (USA)</td>
<td>41.8586</td>
<td>-85.4037</td>
<td>Gr (1)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Gr (1)</td>
</tr>
<tr>
<td>18</td>
<td>Lyons L.</td>
<td>Co</td>
<td>Michigan (USA)</td>
<td>42.1884</td>
<td>-84.9668</td>
<td>Gr (3)</td>
</tr>
<tr>
<td>19</td>
<td>Pogey L.</td>
<td>Mecosta Co.</td>
<td>Michigan (USA)</td>
<td>43.7962</td>
<td>-93.4821</td>
<td>Gr (4)</td>
</tr>
<tr>
<td>20</td>
<td>Fish L.</td>
<td>Calhoun Co.</td>
<td>Michigan (USA)</td>
<td>42.0494</td>
<td>-85.8600</td>
<td>Gr (2)</td>
</tr>
<tr>
<td>21</td>
<td>Goose Bay</td>
<td>Labrador</td>
<td>Nfld. (Canada)</td>
<td>53.2945</td>
<td>-60.3618</td>
<td>Fr (2)</td>
</tr>
</tbody>
</table>

Table 1. *Pyganodon* sampling sites.
**Amplification of the F-orf**

Total DNA was isolated from either mantle or testis tissue using a QIAGEN DNeasy animal kit (QIAGEN Inc., USA) following the animal tissue protocol.

F-orf DNA was amplified using 2 previously identified primers:
FORFuniversalFor (5’-TTGAAAHYARAARARRAYGTTC-3’) and
FORFuniversalRev (5’-TAYTTWRCGYDGYTCTGA-3’) (Breton et al. 2010).

Additionally, 2 F-orf primers were specifically designed for the Pyganodon genus:
PygaFORFFor (5’-TCGGAGAATATTTGCTTGAA-3’) and PygaFORFRev (5’-
TCCTGGCAAGAAGAAATACAAA-3’). PCR conditions were as follow:

A) *P. lacustris* and *P. fragilis* F-orf amplifications: The PygaFORFFor primer was paired with the FORFuniversalRev primer for these species. Amplifications were performed in 50 μL volumes of solutions containing 1X QIAGEN PCR buffer, 0.2 mM of each dNTP, 0.5 μM of each primer, 1U QIAGEN Taq and ~10-100 ng of template DNA. Reactions had an initial denaturing step of 95 °C for 1 min, followed by 40 cycles of denaturation at 94 °C for 60 s, annealing at 45 °C for 60 s, extension at 72 °C for 90 s, and a final extension at 72 °C for 10 min.

It should be noted that amplification of the F-orf in *P. fragilis* proved to be difficult and due to a smaller number of available tissues, *P. fragilis* only produced four viable PCR products.
B) *P. grandis* F-orf amplification: The PygaFORFFFor primer was paired with the PygaFORFRev primer. Amplifications were performed in 50 μL volumes of solutions containing 1X QIAGEN PCR buffer, 0.2 mM of each dNTP, 0.5 μM of each primer, 1U QIAGEN Taq and ~10-100 ng of template DNA. Reactions had an initial denaturing step of 95 °C for 1 min, followed by 40 cycles of denaturation at 94 °C for 60 s, annealing at 50 °C for 60 s, extension at 72 °C for 90 s, and a final extension at 72 °C for 10 min.

All PCR products were then purified using a QIAquick PCR purification kit following the manufacturer's protocol (QIAGEN Inc., USA). Purified PCR products were sent to Geneway Research (Hayward, CA) for sequencing.

**Sequence Analysis**

F-orf nucleotide sequences were aligned using the ClustalW algorithm in the program BioEdit version 7.0.5.3 (Hall 1999). The nucleotide sequences were translated to amino acid sequences with the program MEGA version 4.0 Alignment Explorer and using the provided invertebrate mt DNA genetic code (Tamura et al. 2007). Pairwise p-distances were then calculated for the nucleotide and amino acid sequences of each species to detect intraspecific and interspecific differences.

Sequence similarity searches of the F-ORF amino acid sequences were performed using pBLAST and PSI-BLAST algorithms against the non-redundant GenBank translated database (Altschul et al. 1997).

Predictions of protein structures were determined using PredictProtein (Rost et al. 2003). This program predicts the secondary structures, motifs and binding sites of the
proteins based on the properties of amino acids. Additional predictions of trans-membrane helices were obtained using the ConPredII platform (Arai et al. 2004). Hydropathy plots were created using the Kyte and Doolittle method (Kyte and Doolittle 1982).

**Evolutionary Analysis**

A phylogenetic tree of *Pyganodon* was constructed by RAxML analysis (Random Axelerated Maximum Likelihood) (Stamatakis 2006) using concatenated Fcox2-Fcox1 nucleotide data that had been previously determined in the laboratory (unpublished data). To optimize the branch lengths of this tree, PAUP* version 4.0 (Swofford 2003) was used to generate a maximum-likelihood phylogenetic tree of the included *Pyganodon* samples.

The program SELECTON (Doron-Faigenboim et al. 2005; Stern et al. 2007) was used to identify whether the F-orf is undergoing positive selection or purifying selection. Specifically, site-specific selection, i.e. the estimation of the ratio of the number of nonsynonymous substitutions per nonsynonymous site (Ka) over the number of synonymous substitutions per synonymous site (Ks) or Ka/Ks at each codon site, was studied using a Bayesian inference approach. Neutrality is indicated by Ka/Ks = 1, purifying selection by Ka/Ks < 1, and positive selection is usually invoked as a possible explanation for rare cases where the pattern Ka/Ks > 1 is observed. This analysis was performed using the MEC evolutionary model. This model of evolution takes into account the differential probabilities of amino acid substitution in molecular evolution, as
well as the transition-to-transversion and synonymous-to-nonsynonymous substitution ratios (Doron-Faigenboim and Pupko 2006). This was compared to the null hypothesis of the M8a evolutionary model, which excludes positive selection (Swanson et al. 2003). These two models were compared using AIC values, with smaller AIC values indicating evidence in favor of the MEC model. For each position, a confidence interval defined by the 5\textsuperscript{th} and 95\textsuperscript{th} percentiles of the posterior distributions inferred for the position was estimated. For positions with an inferred Ka/Ks > 1, the inference of positive selection is considered reliable when the lower bound of the confidence interval is larger than 1.
Results

Nucleotide Analysis

The F-orf nucleotide sequences of three *Pyganodon* species were analyzed. The length of the *P. lacustris* F-orf was 258 bp. *P. fragilis* F-orf sequences were 261 bp long. The F-orf of most *P. grandis* samples (n=32) was 258 bp long, whereas some *P. grandis* F-orf sequences were 234 bp long (n=3). These short *P. grandis* F-orf sequences coincided with a small *P. grandis* clade (Clade B) that belonged to the *P. lacustris* clade previously found in the Fcox2-Fcox1 phylogenetic tree (Figure 1).
Figure 1. Maximum-likelihood phylogenetic tree. Determined using F-orf and concatenated Fcox2-Fcox1 nucleotide sequence data. This tree was constructed by RAxML analysis and branch lengths were optimized by PAUP*.

Note that *P. grandis* is displayed as a paraphyletic group as a smaller clade is found diverged from the *P. lacustris/P. fragilis* lineage.

The total average nucleotide composition of the F-orf for all individuals was 26.725% T, 22.975% C, 24.1% A and 26.25% G (A-T content: 50.825 %) (Table 2).

Pairwise p-distances were calculated from the proportions of different nucleotide sites (Table 3). Sequences from *P. fragilis* displayed the smallest amount of intraspecific divergence (p=0) and *P. lacustris* had the greatest intraspecific divergence (p=.044). The least amount of interspecific divergence was found between *P. fragilis* and *P. lacustris*.
(p=.063). Sequences from *P. grandis* differed the most from other species, as p-distances were greatest with *P. lacustris* (p=.129), followed by *P. fragilis* (p=.1).

After identifying two *P. grandis* clades (A and B) (Figure 1), nucleotide and amino acid pairwise p-distances were calculated within and between clade A and B (Table 4). The results indicated a higher amount of divergence within clade A (p=.051 for nucleotides, p=.036 for amino acids) than within clade B (p=.022 for nucleotides, p=.018 for amino acids). The great amount of divergence between clade A and B (p=.760 for nucleotides, p=.350 for amino acids), which has implications for *P. grandis* phylogeny, will be discussed in further details.

<table>
<thead>
<tr>
<th></th>
<th>T(U)</th>
<th>C</th>
<th>A</th>
<th>G</th>
<th>A-T Content</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grandis</td>
<td>27.5</td>
<td>23.5</td>
<td>25.0</td>
<td>24.0</td>
<td>52.5</td>
</tr>
<tr>
<td>Lacustris</td>
<td>26.6</td>
<td>22.5</td>
<td>24.2</td>
<td>26.7</td>
<td>50.8</td>
</tr>
<tr>
<td>Fragilis</td>
<td>26.4</td>
<td>22.9</td>
<td>23.6</td>
<td>27.1</td>
<td>50.0</td>
</tr>
<tr>
<td>Average</td>
<td>26.725</td>
<td>22.975</td>
<td>24.1</td>
<td>26.25</td>
<td>50.825</td>
</tr>
</tbody>
</table>

**Table 2. Average nucleotide sequence composition.** Values are given in percents. Generated in MEGA 4.0.

<table>
<thead>
<tr>
<th></th>
<th>Grandis</th>
<th>Lacustris</th>
<th>Fragilis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grandis</td>
<td>.041</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Lacustris</td>
<td>.129</td>
<td>.044</td>
<td></td>
</tr>
<tr>
<td>Fragilis</td>
<td>.1</td>
<td>.063</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 3. Pairwise nucleotide sequence p-distances within and between species.** Values given represent the proportions of different nucleotide sites. Generated in MEGA 4.0.
<table>
<thead>
<tr>
<th></th>
<th>Clade A (nucleotide)</th>
<th>Clade B (nucleotide)</th>
<th>Clade A (amino acid)</th>
<th>Clade B (amino acid)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clade A</td>
<td>.051</td>
<td></td>
<td>.036</td>
<td></td>
</tr>
<tr>
<td>Clade B</td>
<td>.760</td>
<td>.022</td>
<td>.350</td>
<td>.018</td>
</tr>
</tbody>
</table>

**Table 4. Pairwise nucleotide and amino acid p-distances for two clades of *P. grandis*.**

Values given represent the proportions of different nucleotide and amino acid sites. Generated in MEGA 4.0.

**Protein Analysis**

The F-orf nucleotide data was translated to amino acids in MEGA 4.0 using the invertebrate mt genetic code. The length of the polypeptide sequence was 86 amino acids in *P. fragilis*; 85 amino acids in most of *P. lacustris* and *P. grandis* clade A and 77 amino acids in *P. grandis* clade B.

Both shared and unique (i.e. specific to 1 individual) amino acid sequences were identified for the three species. *P. grandis* had 5 unique sequences and 7 shared amino acid sequences between 35 individuals. *P. lacustris* had 22 unique sequences and 4 shared sequences between 10 individuals. *P. fragilis* had 1 unique sequence and 1 shared sequence between 3 individuals.

To demonstrate the amino acid differences between Fcox2, Fcox1 and the F-ORF, the total amino acid compositions of these polypeptide sequences were created in MEGA 4.0, using data from all individuals (**Table 5**). Considerably less glycine (0.67% vs. 9.63% and 10.3%) was detected in the F-ORF sequence compared to the Fcox1 and Fcox2 sequences. The F-ORF also had a greater amount of aspartic acid (7.57% vs. 4.48% and 3.34%), isoleucine (11.76% vs. 6.65% and 9.9%), lysine (7.04% vs. 1.79%)
and 0.18%), serine (12.71% vs. 8.22% and 5.48%) and threonine (7.14% vs. 1.23% and 4.74%) than the Fcox2 and Fcox1 sequences.
The table below shows the compositions for FQx2, FQx1, and FQx2 variants. All values are given in percent, determined by MEGA 4.0.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Composition (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FQx2</td>
<td>1.16 2.33 6.98 11.60 16.0 31.3 4.9 13.95 6.98 3.1 1.16 11.6 2.33</td>
</tr>
<tr>
<td>FQx1</td>
<td>4.9 0.94 1.18 4.18 2.96 1.28 1.28 2.96 1.18 4.18</td>
</tr>
<tr>
<td>FQx2-v5</td>
<td>4.6 1.19 3.04 2.72 0.14 2.34 1.87 7.12 1.46 1.09 1.39 5.22 3.52 4.82 11.53 80.4 3.9 1.12 0.20</td>
</tr>
<tr>
<td>FQx2-v6</td>
<td>46.1 0.99 3.27 1.96 6.6 0.67 2.35 1.76 7.14 1.29 1.18 0.76 5.4 4.73 1.27 7.14 6.2</td>
</tr>
<tr>
<td>FQx2-v7</td>
<td>46.1 0.99 3.27 1.96 6.6 0.67 2.35 1.76 7.14 1.29 1.18 0.76 5.4 4.73 1.27 7.14 6.2</td>
</tr>
<tr>
<td>FQx2-v8</td>
<td>52.1 2.24 4.48 5.8 4.93 9.63 3.25 6.65 1.19 1.28 2.43 1.88 3.88 2.65 5.55 3.22 12.3 0.12 0.16 4.8</td>
</tr>
</tbody>
</table>

Note: The compositions are given without specific units as percentages.
Pairwise p-distances were calculated for the F-ORF amino acid sequences, which display the proportions of variable amino acid sites (Table 6). These results echoed the nucleotide p-distance data, as *P. fragilis* and *P. lacustris* displayed the least amount of interspecific difference (p=.089) whereas the greatest divergence was found between *P. grandis* and *P. lacustris* (p=.214). The least amount of intraspecific difference was found in *P. fragilis* (p=0). Unlike the nucleotide p-distance results, *P. grandis* displayed the greatest amount of intraspecific divergence (p=.075).

![Table 6](image.png)

**Table 6. Pairwise amino acid sequence p-distances within and between species.**
Values given represent the proportions of different amino acid sites. Generated in MEGA 4.0.

Both shared and unique amino acid sequences were identified. *P. fragilis* had 1 unique amino acid sequence and 1 shared amino acid sequence, *P. grandis* had 7 shared and 5 unique amino acid sequences and *P. lacustris* had 4 shared and 22 unique amino acid sequences. No significant matches (E-value < .01) of the amino acid sequences were found in GenBank, using the pBLAST and PSI-BLAST algorithms.

Predictions of protein structure were determined to identify potential transmembrane helices and other protein motifs. Previous studies have shown the F-ORF protein products of other freshwater mussels to be located in the inner mt and nuclear membranes (Breton et al. 2010). Proteins with a TMH motif are known to exist in
membranes (Xiong 2006; Breton et al. 2010). All *Pyganodon* amino acid sequences were found to contain a predicted TMH of variable length around the 15\textsuperscript{th} and 38\textsuperscript{th} amino acid position, on the N-terminus’-half of the polypeptide (Figure 2). Hydropathy and snake-diagrams further illustrate this prediction (Figure 3). However, it must be noted that this is only a prediction of a TMH; x-ray crystallography of the F-ORF protein would or reject the hypothesized existence of a TMH.

**Figure 2. Predicted F-ORF protein motifs.** These motifs are predicted in all *Pyganodon* F-ORF sequences.
Figure 3. Snake-like diagrams and hydropathy plots for each *Pyganodon* species.

Produced by ConPredII TMS analysis. Hydropathy plots were made using Kyte & Doolittle analysis. Note that two peaks on some hydropathy plots indicate the majority and minority predictions of the F-ORF protein
A – *P. fragilis.*
B – Clade A of *P. grandis.*
C – Clade B of *P. grandis.*
D – *P. lacustris.*

The F-ORF proteins were predicted to have other protein motifs. Almost all of the F-ORF amino acid sequences were predicted to have a protein kinase C phosphorylation site [SSK] just prior to the predicted TMH, in position 13 to 15 of each polypeptide. However, the short F-ORF sequences of *P. grandis* clade B appear to be missing this predicted motif, as the sequences appear to have a deletion of 11 amino acids from position 2 to 12 of the normal F-ORF polypeptide. Additionally, every F-ORF amino acid sequence in this study was predicted to contain 2 casein kinase II phosphorylation sites [SLSD, SVSD, SADD, SAGD]; in each sequence, one site was predicted just after the TMH, around position 39, while the second site was detected towards the C-terminus of the polypeptide, around position 62.

Additional protein motifs were found in different amino acid sequences. An N-glycosylation site [NTSL, NTSR] was predicted at the C-terminus of two unique *P. lacustris* amino acid sequences in addition to the C-terminus of five shared and two unique *P. grandis* amino acid sequences. An additional protein kinase C phosphorylation site [TRR, TSR] was identified at the C-terminus of two unique *P. grandis* sequences in *P. grandis* clade A. An N-myristoylation site [GLDDTS] was predicted around position
50 of one shared *P. grandis* sequence and one shared and one unique *P. lacustris* sequences.

These results indicate the strong likelihood that the F-ORF protein product contains a conserved TMH domain. Snake-diagrams (Figure 3) generated in the ConPredII TMH analysis predicted the C-terminus’-half of the polypeptide after the TMH on the outside of the inner mt or nuclear membrane. This side of the polypeptide chain also had a greater number of predicted protein motifs, although there is some non-species-specific variation between the type of protein motif. Additionally, it should be noted that Predict Protein analysis predicted the existence of protein-protein binding sites in every polypeptide sequence, but no protein-DNA binding sites (Figure 4).
Figure 4. Examples of F-ORF in *Pyganodon* species. Amino acid sequences were generated in MEGA version 4.0. Predictions were made using PredictProtein and ConPredII platforms.

Note that the second *P. grandis* F-ORF is an example of the shorter F-ORF displayed in the divergent *P. grandis* clade. The F-ORFs found in this clade appears to have lost 11 amino acids (position 2 to position 12 on other F-ORF sequences).
Evolutionary Analysis

The phylogenetic tree generated by RAxML analysis indicates *P. grandis* is not a monophyletic group (Figure 1). *P. grandis* was found to be a paraphyletic group, with clade A containing the majority of the *P. grandis* samples from various locations in Minnesota, Wisconsin and Michigan. However, three *P. grandis* individuals were found in a separate clade, B, that appears to share a common ancestor with *P. lacustris* and *P. fragilis*. These individuals were from 2 locations in Minnesota: one male individual from the Rice River in Aitkin County and one male and one female from Pelican Lake in St. Louis County. These two locations are separated by over 150 miles and are not directly connected. Additionally, other individuals from these locations were found in clade A.

The *P. lacustris* individuals were found to exist in a monophyletic group, composed of one larger clade containing the majority of individuals and one smaller clade sharing a common ancestor. The larger clade included individuals from Minnesota, Wisconsin and Michigan. The smaller clade was composed entirely of individuals from Vilas County, Wisconsin: three males from Kentuck Lake and one male from North Twin Lake. These two locations are not directly connected and all *P. lacustris* samples from Kentucky and North Twin Lakes, fell into this clade.

The MEC evolutionary model, which analyzed the entire data set of Fcox2-Fcox1 and F-orf nucleotide sequences, found that there were several codons undergoing positive selection (Figure 5). Computed AIC values were smaller for the MEC model (10252.7771057514) than the null hypothesis M8a model (30812.6247218789), indicating a better fit for the MEC model. The MEC model determined that 4 codons in...
Fcox2 sequences, 0 codons in Fcox1 sequences and 16 codons in the F-orf are undergoing site-specific, diversifying (i.e. positive) selection. The 16 positively-selected codons in the F-orf are scattered throughout the sequence, however, 4 of these were identified within the predicted TMH region, with mostly hydrophobic amino acid variants. Additionally, there appears to be a concentration of positively-selected codons near the termini.
Figure 5. MEC evolutionary model analysis of codon selection. Codons are represented as amino acids. Position 1 to 226 represents Fcox2, position 227 to 455 represents Fcox1 and position 456 to 541 represents the F-orf.

Fcox2 was found to have 4 positively selected codons. No positively selected codons were detected for Fcox1. The F-orf was found to have 16 positively selected codons. This indicates that while Fcox2 and Fcox1 are under strong purifying selection and evolutionary constraint, the F-orf is undergoing site-specific, diversifying selection.

Note that a trans-membrane helix was predicted in the F-ORF, around position 471 to position 492. This region contains 4 positively selected codons with mostly hydrophobic variants. It also contains a conserved [SFFI] sequence from position 482 to 485.
**Discussion**

The key results of this study were as follows: first, an unusual phylogenetic tree was generated from the data that calls into question the current understanding of the reproductive isolation of species in the *Pyganodon* genus. Second, the results from the SELECTON analysis indicate that the F-orf in the *Pyganodon* lineage is under diversifying selection in 16 codons of the gene. Lastly, predictions made on the F-ORF protein product support the previous prediction of a conserved TMH, in addition to suggesting this protein could carry other functional motifs with extra-mitochondrial functionality.

**Phylogenetic Implications for Pyganodon Lineages**

The phylogenetic tree that was generated for this analysis indicates that *P. grandis* is in fact a paraphyletic group, separated into one clade (clade A) that contains most individuals, and one smaller clade (clade B) that shares a common ancestor with *P. lacustris* and *P. fragilis* (Figure 1). Clade B contains a smaller F-ORF protein product that is 77 amino acids long instead of the average 85 amino acid long F-ORF that is found in the rest of the *P. grandis* lineage (Figure 4). This smaller F-ORF is characterized by the loss of several amino acid sites near the N-terminus. Due to the variability in the length (~66 to ~92 amino acids) of the F-ORF in unionoids and the lack of an outgroup for the *Pyganodon* F-ORF data, it is unclear whether the smaller F-ORF is
the ancestral or derived character state (Breton et al. 2009). Given the optimized branch lengths of the phylogenetic tree, this clade appears to have diverged considerably from the rest of the *Pyganodon* lineage. There is a very high amount of difference in the nucleotide (p=.760) and amino acid (p=.350) sequences between these two clades (Table 4). These results suggest the possibility that hybridization or cryptic speciation could have occurred in the *P. grandis* lineage.

This information contradicts a previous study based on F and M mtDNAs and nuclear DNA for *Pyganodon* species in Quebec, which concluded that *P. grandis* represents a single lineage (Cyr et al. 2007). While other work on *P. grandis* also disputes the presence of two separate lineages (Williams et al. 1992), there are some who believe there is an unidentified, unique *Pyganodon* lineage (Clarke 1981). However, all individuals used in this present study were identified as *P. grandis* based on shell characteristics.

That the individuals in clade B were found in locations that contained *P. grandis* individuals from clade A as well as *P. lacustris* individuals, presents an interesting challenge in identifying the reasons behind the split in lineages. It is possible that sympatric speciation occurred in the past between *P. grandis* sub-populations. However, the amount of divergence between the two *P. grandis* clades would suggest that these two lineages split before the emergence of the *P. lacustris*-*P. fragilis* clade. If sympatric speciation did occur, one would then expect the *P. grandis* lineages to be morphologically distinct and reproductively isolated. However, currently there is no
evidence to suggest morphological distinction or reproductive isolation. Thus, if sympatric speciation did occur in the past, it was a cryptic speciation event.

Alternatively, it is possible *P. grandis* and another species of *Pyganodon* hybridized at some point in the past. The branch length of clade B, along with the rapid rate of evolution of the F-ORF suggested in the SELECTON analysis, would seem to suggest this possibility. Surveys conducted through the laboratory show that *P. grandis* and *P. lacustris* populations live in proximity to each other, making *P. lacustris* the most likely candidate for hybridization, but so far no evidence has been found that indicates these two species have previously hybridized or are currently hybridizing (Hoeh and Burch 1989).

If a hybridization or speciation event occurred in the past, an insertion or deletion of 24 F-orf nucleotides had to have occurred, as the F-orf sequences of clade B were 24 base pairs smaller than the other *Pyganodon* species. However, there is no explanation why there would be such a large deletion due to hybridization. If this deletion is correlated to a speciation event, it would be expected that this mutation would take a long time to fix in a lineage, meaning that a cryptic speciation event happened a very long time ago. However, as stated earlier, there is no existing evidence that suggests any reproductive isolation between different lineages of *P. grandis*.

The greatest amount of interspecific divergence of nucleotide and amino acid sequence divergence was found between *P. grandis* and *P. lacustris* (p=.129 for nucleotides and p=.214 for amino acids), making it less likely that these species could successfully hybridize. However, that does not necessarily rule out a past hybridization
event. To gain a stronger understanding of whether hybridization between these two species occurred in the past, future research should focus on investigating additional genetic data and observing whether these species are reproductively isolated.

The results also indicate that *P. lacustris* is a monophyletic group divided into 2 clades. The evolution and variability of the F-orf in this species is intriguing because this species displayed the greatest number of unique amino acid sequences (n=22) and the greatest intraspecific nucleotide divergence (p=.044) of any species (Table 2). While *P. lacustris* displayed the greatest number of unique sequences, each unique sequence was defined by a relatively small number of substitutions. Small changes in polypeptide composition and a high amount of intraspecific nucleotide divergence indicate the *P. lacustris* F-orf has the greatest amount of synonymous substitutions. This could mean that the F-orf in *P. lacustris* is more evolutionarily constrained than in other species.

In contrast to *P. lacustris*, the F-orf is less evolutionarily constrained in *P. grandis*, considering the results demonstrate the lowest amount of intraspecific nucleotide divergence (p=.041) and greatest intraspecific amino acid divergence (p=.075) (Table 2 and Table 6). *P. grandis* had more shared amino acid sequences than unique sequences but between these sequences, there are many differences in composition and polypeptide length. Thus, the *P. grandis* F-orf has had the least amount of evolutionary constraint than the F-orf in any other species of *Pyganodon*.

There is still uncertainty in understanding the evolutionary history of *Pyganodon*. The results indicate there could be a conflict in the taxonomical classification of this genus between genetic and morphological characters. It is possible this issue in taxonomy
and phylogeny could apply to other unionoid bivalves and other groups of invertebrates where differences in morphology are difficult to determine. In the future, it will be important to identify species-specific genetic markers in the nuclear genome of *Pyganodon*, construct a phylogeny based on this information and compare it to the *Fcox2-Fcox1* phylogeny in order to resolve these taxonomical issues.

**Diversifying Selection Acting on the F-orf**

The SELECTON results from the MEC evolutionary model demonstrate that within the F-orf there are 16 codons undergoing diversifying selection (Figure 5). This shows the F-orf to be rapidly evolving, as opposed to the *Fcox2* and *Fcox1* genes, which only display 4 and 0 codons respectively undergoing diversifying selection. Due to the importance of *cox* proteins during oxidative phosphorylation as the last step in the electron transport chain, it is not surprising these genes are heavily conserved. The 16 rapidly evolving codons of the F-orf appear to be scattered throughout the nucleotide sequence, though there are several that are located within the predicted TMH zone, as well towards the N’ and C’ termini. This positive selection of the F-orf in *Pyganodon* provides a key insight into the nature of the F-ORF protein.

There is a growing amount of evidence suggesting proteins unique to the unionoid DUI system, such as the F-ORF and M-ORF, are involved in reproduction (Breton et al. 2007, 2009, 2010). Additionally, there is evidence suggesting that the *Fcox2* protein has a reproductive function in unionoids, due to the strong expression of *Fcox2* in the vitelline matrix of egg cells immediately prior to fertilization (Chakrabarti 2009).
Breton et al. 2010 showed that the F-ORF protein is found not only in the mitochondria, but also in the nuclear membrane and nucleoplasm in mature ovarian eggs of *V. ellipsiformis*. These results support the hypothesis that the F-ORF protein has a role to play in the reproduction and sex determination of unionoids.

Proteins that have a reproductive function are known to undergo rapid positive selection (Swanson and Vacquier 2002; Swanson et al. 2003). It is known that angiosperm plants exhibiting cytoplasmic male sterility (CMS) possess mt genes that play a role in sex determination (Schnable and Wise 1998; Chase 2007; Delph et al. 2007). Additionally, other mt proteins in unionoids have been shown to be correlated with sex determination, such as Fcox2 (Chakrabarti et al. 2009) and Mcox2e (Chakrabarti et al. 2006, 2007; Chapman et al. 2008). Given this information, along with the correlation between the F-ORF and sex determination presented in Breton et al. 2010, the identification of 16 F-orf codons undergoing diversifying selection is further evidence in favor of a reproductive function in the F-ORF protein.

Corresponding to the predicted TMH region in the F-ORF protein, four positively-selected codons and a small conserved sequence were identified (Figure 5). Positions 19, 20, 24 and 33 of the polypeptide are under strong positive selection and most amino acid variants have hydrophobic residues (though position 24 does vary between valine, a hydrophobic amino acid, and threonine, a hydrophilic amino acid) (Figure 3). In position 26 to 29, an [SFI] sequence was identified that is under strong purifying selection. Regardless of whether a site was conserved or under positive selection, the majority of amino acid variants in the predicted TMH region are
hydrophobic, increasing the likelihood of the presence of a TMH. This also suggests that if a TMH actually does exist in the F-ORF, it is evolutionarily conserved, despite positive-selection. It could be the case while there is positive selection in the TMH, only hydrophobic amino acids are retained.

**F-ORF Protein Predictions**

The prediction of a single, conserved TMH in the N’ half of the F-ORF protein is consistent with the results found in Breton et al. (2009) and (2010)’s studies of the F-ORF. Interestingly, this current study found that several codons undergoing diversifying selection were present in this predicted TMH region (**Figure 5**). Although x-ray crystallography must be performed in the future to confirm the hypothesized existence of a single TMH, if a TMH does in fact exist in a conserved region on the F-ORF protein, this TMH could have an adaptive function due to the positively selected codons. The M-ORF protein of unionoids has also been predicted to have a single, conserved TMH (Breton et al. 2009). It would be interesting to determine if the codons in the M-ORF TMH predicted region are also undergoing diversifying selection, as this would suggest that TMHs in both ORFs carry some type of adaptive function that is possibly related to reproduction.

This study was not able to incorporate any techniques to identify where the F-ORF protein is expressed in *Pyganodon* cells. Breton et al. (2010) utilized immunoelectron microscopy to show that the F-ORF in *V. ellipsiformis* is expressed not only in the mitochondria, but in the nuclear membrane and nucleoplasm as well. I-
TASSER analysis, a state-of-the-art hierarchical protein structure modeling approach that is based on the secondary-structure enhanced profile-profile threading alignment, also predicted the involvement of the F-ORF in DNA binding and/or DNA replication (Breton et al. 2010). Those results support the hypothesis that the F-ORF protein is involved in a genetic regulatory network that specifies sex. However, the *Pyganodon* F-ORF protein structures were predicted to contain no protein-DNA binding sites, only protein-protein binding sites, though it is possible the F-ORF has an unknown protein-DNA binding site. Furthermore, with the exception of lysine, positively charged amino acids that could bind DNA were not found in greater abundance in the F-ORF compared to Fcox1 and Fcox2 (Table 5). In the future, immunoelectron microscopy of *Pyganodon* F-ORFs should be used to determine where the F-ORF is located in the cell. Additionally, chromatin immunoprecipitation should be used to determine if the F-ORF actually binds DNA. If the F-ORF is found to be present on the nuclear membrane and binds DNA, this will support the hypotheses of a conserved TMH and a role in genetic regulation.

If the F-ORF is not found to bind or directly regulate DNA, this would not negate the possibility the F-ORF is involved in sex determination as an upstream regulatory protein. It is interesting to note that Chakrabarti et al. 2009’s study of the Fcox2 protein of *V. ellipsiformis* identified the expression of Fcox2 not only in the mitochondria and cytoplasm of mature ovarian eggs, but most heavily in the vitelline envelope, vitelline matrix and plasma membrane microvilli prior to fertilization. While these results do not directly address the function of the F-ORF, they support the hypothesis that
mitochondrially-encoded proteins involved in the DUI system have extramitochondrial and reproductive functions.

In agreement with Breton et al. 2009’s findings, this study predicted a conserved casein kinase II phosphorylation motif at the C’ end of the predicted TMH zone (Figures 2 and 4). Additionally, a second casein kinase II phosphorylation (CKII) motif was predicted towards the C’ terminus of the polypeptide as well as a conserved protein kinase C (PKC) phosphorylation motif was predicted immediately prior (N’ side) to the predicted TMH region. N-glycosylation and N-myristoylation motifs were predicted in only a few amino acid sequences in *P. lacustris* and *P. grandis* clade A, suggesting that if these predicted motifs exist, they are not essential to the F-ORF’s functionality in *Pyganodon* since they were not found in the majority of F-ORF sequences.

According to the snake-like diagrams provided by the ConPred TMS analysis, the N-terminus’ side of the predicted TMH is likely to exist on the inside of a membrane, while the C-terminus’ side is likely to exist on the outside of the membrane (Figure 3). This means that the CKII phosphorylation motifs would exist on the outside of a membrane, while the PKC phosphorylation motifs would exist on the inside of a membrane. CKII has recently been shown to be involved in regulating the importation of proteins into the mitochondria by promoting the activation of the TOM importation complex in the outer mt membrane (Harbauer et al. 2011). No conclusions can currently be made as to whether the F-ORF participates in this process because the conserved CKII and PKC phosphorylation motifs are only predictions and if they do exist in a functional capacity, the cellular interactions must be explored in depth through enzyme assays. This
coincides with the conclusion that while the F-ORF is predicted to have functional characteristics, especially in relation to sex determination, there is still much work to be done to further elucidate the evolution and functionality of this novel mt protein.
References Cited


Abstract

Previous studies of the unique system of mitochondrial inheritance of freshwater mussels (Bivalvia: Unionoida) known as Doubly Uniparental Inheritance (DUI) have indicated the existence of two novel gender-specific protein-coding genes, the F-orf and the M-orf. Although there is evidence to support the hypothesis that these genes are involved in sex determination, very little is known about the evolution and structure of their protein products. In this thesis, the F-orf of the genus Pyganodon was sequenced and analyzed. Phylogenetic analysis of concatenated Fcox2-Fcox1 and F-orf sequences reveal P. grandis as a paraphyletic group and P. lacustris and P. fragilis are monophyletic groups. Furthermore, results indicate a considerable contrast in the degree of positive selection acting on amino acid positions in the F-orf versus the Fcox2-Fcox1 sequences, as multiple amino acid positions in the F-orf are experiencing diversifying selection. Additional findings are the predictions that all F-ORF proteins in Pyganodon possess a single, conserved trans-membrane helix (TMH) near the N-terminus, as well as other conserved, functional motifs. These results represent information critical to investigating the functional nature of the F-orf and its evolutionary history.