Genetic variation in freshwater mussels (Bivalvia: Unionidae) and its implications for species delineation and conservation

Jer Pin Chong
Iowa State University

Follow this and additional works at: https://lib.dr.iastate.edu/etd

Part of the Biodiversity Commons, Genetics Commons, Natural Resources and Conservation Commons, and the Natural Resources Management and Policy Commons

Recommended Citation
Chong, Jer Pin, "Genetic variation in freshwater mussels (Bivalvia: Unionidae) and its implications for species delineation and conservation" (2016). Graduate Theses and Dissertations. 15684.
https://lib.dr.iastate.edu/etd/15684

This Dissertation is brought to you for free and open access by the Iowa State University Capstones, Theses and Dissertations at Iowa State University Digital Repository. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Iowa State University Digital Repository. For more information, please contact digirep@iastate.edu.
Genetic variation in freshwater mussels (Bivalvia: Unionidae) and its implications for species delineation and conservation

by

Jer Pin Chong

A dissertation submitted to the graduate faculty
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Major: Ecology and Evolutionary Biology

Program of Study Committee:
Kevin J. Roe, Co-Major Professor
Jeanne M. Serb, Co-Major Professor
John D. Nason
Dean C. Adams
Julie A. Blanchong

Iowa State University

Ames, Iowa

2016

Copyright © Jer Pin Chong, 2016. All rights reserved.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>LIST OF FIGURES</td>
<td>iv</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>viii</td>
</tr>
<tr>
<td>CHAPTER 1  GENERAL INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>CHAPTER 2  INCONGRUENCE BETWEEN MTDNA AND NUCLEAR DATA IN THE FRESHWATER MUSSEL GENUS CYPROGENIA (BIVALVIA: UNIONIDAE) AND ITS IMPACT ON SPECIES DELINEATION</td>
<td>13</td>
</tr>
<tr>
<td>Abstract</td>
<td>13</td>
</tr>
<tr>
<td>Introduction</td>
<td>14</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>17</td>
</tr>
<tr>
<td>Results</td>
<td>22</td>
</tr>
<tr>
<td>Discussion</td>
<td>26</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>33</td>
</tr>
<tr>
<td>References</td>
<td>42</td>
</tr>
<tr>
<td>Supplemental information</td>
<td>51</td>
</tr>
<tr>
<td>CHAPTER 3  COMPARING THE GENE FLOW PATTERNS OF THE ENDANGERED SCALESHELL (LEPTODEA LEPTODON), THE WIDELY DISTRIBUTED FRAGILE PAPERSHELL (LEPTODEA FRAGILIS) AND THEIR HOST-FISH THE FRESHWATER DRUM (APLODINOTUS GRUNNIENS)</td>
<td>52</td>
</tr>
<tr>
<td>Abstract</td>
<td>52</td>
</tr>
<tr>
<td>Introduction</td>
<td>53</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>55</td>
</tr>
<tr>
<td>Results</td>
<td>62</td>
</tr>
<tr>
<td>Discussion</td>
<td>67</td>
</tr>
<tr>
<td>Acknowledgements</td>
<td>73</td>
</tr>
<tr>
<td>References</td>
<td>81</td>
</tr>
<tr>
<td>CHAPTER 4  ASSESSING SPERM AND ZYGOTE-MEDIATED GENE FLOW IN FRESHWATER MUSSELS (BIVALVIA: UNIONIDAE)</td>
<td>87</td>
</tr>
<tr>
<td>Abstract</td>
<td>87</td>
</tr>
<tr>
<td>Introduction</td>
<td>88</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>93</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
</tr>
<tr>
<td>--------</td>
<td>-------------</td>
</tr>
<tr>
<td>Figure 2-1</td>
<td>Red and brown conglutinates produced by <em>Cyprogenia</em></td>
</tr>
<tr>
<td>Figure 2-2</td>
<td>Map of <em>Cyprogenia</em> mtDNA lineage distribution</td>
</tr>
<tr>
<td>Figure 2-3</td>
<td>Bayesian inference gene tree based on the mitochondrial ND1 gene of two <em>Cyprogenia</em> species</td>
</tr>
<tr>
<td>Figure 2-4</td>
<td>Haplotype network for the ND1 sequences in <em>Cyprogenia</em></td>
</tr>
<tr>
<td>Figure 2-5</td>
<td>Geographic distribution of <em>Cyprogenia</em> ESUs based on results of STRUCTURE analysis</td>
</tr>
<tr>
<td>Figure 3-1</td>
<td>Sampling map for <em>Le. leptodon</em>, <em>Le. fragilis</em>, and <em>A. grunniens</em></td>
</tr>
<tr>
<td>Figure 3-2</td>
<td>Mitochondrial DNA haplotype networks for <em>Le. leptodon</em>, <em>Le. fragilis</em>, and <em>A. grunniens</em></td>
</tr>
<tr>
<td>Figure 3-3</td>
<td>STRUCTURE barplots for <em>Le. leptodon</em>, <em>Le. fragilis</em>, and <em>A. grunniens</em></td>
</tr>
<tr>
<td>Figure 3-4</td>
<td>Isolation by distance figures for <em>Le. leptodon</em>, <em>Le. fragilis</em>, and <em>A. grunniens</em></td>
</tr>
<tr>
<td>Figure 3-5</td>
<td>Pairwise genetic distances linearized $F_{ST}$ estimated among populations of <em>Le. leptodon</em>, <em>Le. fragilis</em>, and <em>A. grunniens</em></td>
</tr>
<tr>
<td>Figure 4-1</td>
<td>Sampling map for <em>Le. leptodon</em> and <em>Le. fragilis</em> in Missouri</td>
</tr>
<tr>
<td>Figure 4-2</td>
<td>Map showing sampling locations of all <em>La. abrupta</em></td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
</tr>
<tr>
<td>2-1</td>
<td>Summary of mtDNA ND1 gene and microsatellite diversity from <em>Cyprogenia</em> samples grouped according to current taxonomy or STRUCTURE analysis.</td>
</tr>
<tr>
<td>2-2</td>
<td>Shimodaira-Hasegawa Test (SH-test) on the likelihoods of constraint and unconstraint trees corresponding to three a priori models</td>
</tr>
<tr>
<td>2-3</td>
<td>AMOVA analysis performed on individuals grouped according to three a priori models.</td>
</tr>
<tr>
<td>2-4</td>
<td>Genetic differentiation among <em>Cyprogenia</em> populations from Ouachita, Ozark, and Eastern clusters.</td>
</tr>
<tr>
<td>Supplemental 2-S1</td>
<td>Sample sizes from all <em>Cyprogenia</em></td>
</tr>
<tr>
<td>Supplemental 2-S2</td>
<td>Outgroup taxa included in the Bayesian analysis</td>
</tr>
<tr>
<td>3-1</td>
<td>Sampling locations and sample size for <em>Le. leptodon, Le. fragilis, and A. grunniens</em> populations.</td>
</tr>
<tr>
<td>3-2</td>
<td>Summary of mtDNA ND1 gene and microsatellite diversity from five <em>Le. leptodon, Le. fragilis, and A. grunniens</em> populations.</td>
</tr>
<tr>
<td>3-3</td>
<td>Analysis of molecular variance (AMOVA) among <em>Le. leptodon, Le. fragilis, and A. grunniens</em> populations.</td>
</tr>
<tr>
<td>3-4</td>
<td>The assignment tests and detection of first-generation migrants for <em>Le. leptodon and Le. fragilis</em>.</td>
</tr>
<tr>
<td>3-5</td>
<td>Mantel tests and the correlations among population structures of <em>Le. leptodon, Le. fragilis, and A. grunniens</em>.</td>
</tr>
<tr>
<td>4-1</td>
<td>Sampling locations and sample size for <em>Le. leptodon and Le. fragilis</em>.</td>
</tr>
<tr>
<td>4-2</td>
<td><em>Lampsilis abrupta</em> samples collected from each sampling location.</td>
</tr>
<tr>
<td>4-3</td>
<td>F-statistics for mtDNA $F_{ST(f)}$, nuclear $F_{ST(n)}$, and sperm flow/glochidia flow ratio for <em>Le. leptodon, Le. fragilis, and La. abrupta</em>.</td>
</tr>
</tbody>
</table>
ACKNOWLEDGEMENTS

I came to U.S. for my bachelor’s degree; in no way could I imagine that someday I would graduate with a Ph.D. degree. There are so many people I would like to thank along the way who helped me to become the researcher and the person I am now.

I would like to express my deepest gratitude to my family for their endless caring and support. I left my hometown for my bachelor’s degree promising that I would return in a few years. It was a promise I could not keep over the past 15 years, yet you never stop loving me and always have maintained faith in me as I pursued my dream. My siblings, especially Jer Hong Chong, has always been there for me. Her beautiful daughters, my cute nieces, have cheered me up countless times with their innocent looks and laughs whenever I have experienced one of those long days.

I could not have completed this degree without the guidance and support of my major advisor, Kevin Roe. Kevin, you have always been very generous in funding my program and supporting my travels to conferences. You’ve taught me to become an independent scientist who is constantly exploring ways to conduct better research. Thank you for the extensive time you’ve spent in editing my manuscripts and discussing the direction of my research. I appreciate your advising style; I think it has helped me to become a better researcher and writer. Because of you, I have learned so much since I started my program here. Thank you.

I am grateful to have had committee members who have been very supportive in my research. John, you have been very helpful and shown so much patience in guiding and mentoring me in my research. You have been a great inspiration, and I thank you. Jeanne, the little things you do and say have always provided great comfort to me. Thank you for your guidance and advice in helping me better understand the process of science. Dean, thank you
for helping me in the morphometric data analyses. Your insight and thoughtful advice added a great depth and dimension in my research. Julie, you always made me think about the implications of my research with respect to wildlife management. It keeps me motivated in my research, and I thank you.

I would like to thank Linda Westgate and Chanda Skelton for being great teaching mentors and helped me discovered my passion in teaching. I will be forever grateful to my master’s advisor Karen Mock and lab mentor Carol Rowe for your immense support, for giving me the opportunity to fulfill my potential and taught me all the lab techniques I know.

Over the years I have been very lucky to have many wonderful friends around me to support and guide me through all my challenges. Davide Faggionato, Elke Brandes, Alvin Alejandrino, Glene Mynhardt, Chelsea Berns, you have been dear friends, and thank you for always believing in me even when I did not have much confidence in myself. Lailatul Ali-Husin, Meghan Gillette, Zhi Yong Yang, and John Wong, you are like family members to me. I could not have completed this degree without your unconditional support. I would also like to thank my EEOB and NREM colleagues Fan Yang, Amy Worthington, Vilma Mateos Remigio, Viraj Muthye, Gerardo Antonio Cordero, Anna Belyaeva, Lynne Gardner, Hannah Julich, Karri Folks, Brittany Tawes, Katherine Espinosa, Morgan Sexton, and Jermaine Mahguib for your caring and support. Thank you to all the EEOB and NREM faculties, secretaries, staffs, graduate students, and postdocs for providing a comfortable and supportive work environment. Thank you for everything.
Dispersal ability and reproductive success are important criteria in determining whether a threatened or endangered species is capable of recovering from bottlenecks and catastrophic population declines. The goal of this dissertation is to understand patterns of population genetic variation and population connectivity (i.e. gene flow), and their implications for the conservation of freshwater mussels (Bivalvia: Unionidae). I generated mitochondrial DNA sequence data and nuclear microsatellite data sets for five mussel species (*Leptodea leptodon, Le. fragilis, Lampsilis abrupta, Cyprogenia aberti, C. stegaria*) and one fish species (*Aplodinotus grunniens*) in order to evaluate genetic diversity and population structure within each of these species.

In the first of three chapters, I assessed the genetic variation of two nominal species in the genus *Cyprogenia* (*C. aberti* and *C. stegaria*) in order to delineate the number of evolutionary lineages present within the genus. Based on the molecular data collected from mitochondrial DNA and microsatellite genotyping, I found evidence of mito-nuclear discordance in *Cyprogenia*. Analyses of mtDNA sequences suggest that two deeply divergent clades co-occur sympatrically in most of our sampling sites; the nuclear microsatellite data support three allopatric clades that correspond to major hydrologic basins. My study also showed that the pigmentation of mussel conglutinates in *Cyprogenia* is highly correlated with the recovered mtDNA clades. Furthermore, the mtDNA sequences appear to be under selection and therefore not reliable for delineating species, and the recognition of species and evolutionarily significant units was based solely on the microsatellite data.

The second chapter investigates the impact of host fish dispersal on the gene flow of a common and a rare mussel species. I compared the gene flow pattern and population
structure of an endangered mussel species *Leptodea leptodon* and its common and widespread congener *Le. fragilis* with their shared host fish, the freshwater drum *Aplodinotus grunniens*. The results showed that population structure of *Le. fragilis* were more congruent with that of the host fish, while the endangered *Le. leptodon* populations displayed higher levels of genetic isolation among drainages.

In the third chapter I investigated contributions to mussel gene flow from two sources: gamete-mediated gene flow contributed through sperm dispersal, and zygote-mediated gene flow contributed through larval (glochidia) dispersal. The purpose of this study was to determine whether sperm dispersal or glochidia dispersal (via host fish) contribute more to maintaining connectivity among mussel populations. I developed an approach for estimating sperm gene flow of mussel populations when a paternally inherited genetic marker is not available. This is the first study that attempts to measure the gamete-mediated gene flow (male gene flow) in mollusks using mtDNA and nuclear markers. The results for three mussel species (*Le. leptodon*, *Le. fragilis*, and *Lampsilis abrupta*) showed that sperm gene flow among populations is higher than previously assumed.
Overview of Conservation Genetics

As the human population continues to grow, the negative impact of human activities on the natural world places increasing pressure on plants, animals, and ecosystems (Vitousek et al. 1997; Cardinale et al. 2012). Maintaining biodiversity while providing adequate and sustainable resources for human use remains an incredibly challenging task (Kellert et al. 2000; Dudgeon et al. 2006). A major component in the field of conservation biology is natural resource management, using a top-down approach (conservation efforts initiated by regional, national, or global agencies and applied on local communities) to preserve natural resources and maintain ecosystem resilience (Holling & Meffe 1996). Although demographic effects on population survival are profound, the impact of genetic factors on small populations should not be underestimated (Frankham 2005). Conservation genetics is an interdisciplinary science aiming to preserve genetic diversity, resolve taxonomic uncertainties, and identify management units in an effort to reduce the risk of population extinction caused by genetic factors (Frankham et al. 2002). In his influential paper, Michael Soulé, the father of modern conservation biology, addressed six major factors that can reduce genetic variation in a population: inbreeding, limited gene flow, increased genetic drift, small effective population size, low variation between niches, and directional selection (Soulé 1973). Three of the factors identified by Soulé (1973) are particularly problematic in small populations: bottlenecks, inbreeding depression, and genetic drift. Inbreeding and genetic drift may contribute to the expression of deleterious alleles which can reduce the fitness and survival of individuals in a population (Lynch & Gabriel 1990). Small populations can
become trapped in a feedback loop known as an extinction vortex (Gilpin & Soule 1986) in which inbreeding depression leads to reduced survival of offspring and thus smaller populations which in turn leads to more inbreeding. This process can eventually lead to extinction of local populations and contribute to reducing the long-term survival of a species.

**Species Delineation**

One of the major challenges in conservation research is to identify taxonomic units for wildlife management. Conservation biologists rely on species as taxonomic units for assigning conservation priorities such as identifying biological “hotspots” (Myers et al. 2000). Species delimitation is often closely associated with the particular species concept adopted by each individual researcher (Wayne 1992; De Queiroz 2007). In fact, biodiversity hotspots assigned based on species richness can be completely altered depending upon which species concept was applied (Peterson & Navarro-Siguenza 1999). The biological species concept (BSC) is the most influential species concept (Hausdorf 2011); it defines a species as a group of individuals reproductively isolated from other groups of individuals (Mayr 1942). The BSC is simple and straightforward, but it is not applicable to fossil organisms, asexual organisms, allopatric populations of sexually reproducing organisms, and instances where interspecific hybridization has occurred (Agapow et al. 2004). Alternatively, phylogenetic species concepts (PSC) define species as a group of organisms sharing a pattern of ancestry and descent (Cracraft 1983), or monophyly (Donoghue 1985). The phylogenetic species concept provides a framework for identifying species boundaries and the hierarchical relationships among a lineage of organisms (Eldredge & Cracraft 1980). The PSC is more applicable than the BSC in defining species boundaries for operational management units
because it can be applied to asexual organisms and allopatric populations (Goldstein et al. 2000). However, application of a PSC has the potential to recognize small populations that are diverging due to inbreeding or genetic drift, which may lead to inappropriate management policies that prevent gene flow between populations and eventually cause unnecessarily high population/species extinction rates (Frankham 2010). De Queiroz (2007) proposed a unified species concept defining species as an evolving metapopulation lineage; it was widely accepted and later described as the Generalized Lineage Concept (GLC). The GLC uses criteria from traditional species concepts (e.g. morphology, monophyly, reproductive incompatibility) as secondary criteria and evidence of lineage diversification, instead of using only one criterion as a single definition of a species.

**Population Structure and Connectivity**

After a species has been properly delimited, the next question is, how many populations reside within the species distribution range? A population is referring to a group of individuals with higher probability of mating with one another than with individuals from another group. Understanding the genetic structure among populations allows wildlife managers to identify whether the populations are fragmented and isolated, which ones are the source and sink populations, whether the populations recently underwent a bottleneck event, and whether populations are currently suffering from inbreeding and genetic drift.

To evaluate the viability of a species, an in-depth assessment with respect to the genetic variation among populations is required to predict the possible responses of populations to environmental stresses. Molecular genetic tools can be used to understand evolutionary processes acting at the population scale, such as profiling individuals with DNA
fingerprinting, understanding social structure and migration among populations, estimating effective population size, examining local adaptation, and assessing evolutionary trajectory of populations under selection. Genetic structure among evolutionary entities can be estimated using DNA sequences, microsatellite data, or genomic data. DNA sequencing can be applied to mitochondrial (mtDNA) or nuclear (nDNA) data. Mitochondrial DNA are maternally-inherited haploid lineages that do not recombine. Nuclear DNA data are biparentally-inherited lineages that are diploid and can recombine. Mitochondrial DNA lineages reflect patterns of deeper evolutionary history, while nDNA in general narrates more recent population history (Rodriguez et al. 2010). Mitochondrial DNA sequencing is most often used to reconstruct phylogenetic trees in evolutionary biology studies, but it is also commonly used to infer genetic structure among populations. Microsatellite data is the most commonly used genetic marker in conservation genetics studies (Ouborg et al. 2010). Microsatellite data provide estimates of allele and heterozygosity diversity to infer the overall genetic diversity among populations. Conservation genomics can provide genome-wide estimates of genetic variation both within and between individuals, and can be used to assess gene expression and examine environmental selection acting on phenotypic traits (Ouborg et al. 2010). Conservation genomics is the future trend of molecular research in natural resource management.

**Gene Flow Pattern**

At its core, gene flow is the dispersal of genes among populations. The mechanisms that organisms use to maintain population connectivity directly influence the viability of a species. Species with high dispersal ability and high reproductive success tend to have higher
gene flow and can potentially recover from bottlenecks or other catastrophic population decline. It is thus important to understand how organisms reproduce and how far they can travel to maintain connectivity among populations.

The focus of this dissertation is to understand how population connectivity (i.e. gene flow) influences species viability and adaptability to environmental changes, with my research primary focusing on freshwater mussels (Bivalvia: Unionidae). The reproductive cycle of freshwater mussels includes a parasitic life stage, where the mussel larvae (glochidia) must parasitize a fish host to complete development. Gene flow in freshwater mussels is thought to occur through the following routes: glochidia dispersal (based on the mobility of host fish), gametic dispersal (sperm dispersal), and movement of adults/juveniles (generally sessile, except for movement downstream during flood events). It is crucial for conservation biologists to understand how different types of gene flow contribute to the genetic differentiation and adaptation of local populations.

**Dissertation outline**

Molecular tools can be used to effectively understand the relatedness among populations and can therefore be used to aid in wildlife conservation and management. In this dissertation, I examine the population structures of five mussel species (*Cyprogenia aberti, C. stegaria, Leptodea leptodon, Le. fragilis, and Lampsilis abrupta*) and one fish species (*Aploeginotus grunniens*). In Chapter I, I attempt to resolve taxonomic uncertainties in mussel genus *Cyprogenia* and to delimit species. In Chapter II, I examine the impact of host dispersal on the gene flow structures of a common and endangered mussel species in genus
In Chapter III, I investigate the male-mediated gene flow (through sperm dispersal) in freshwater mussels.

Chapter 1. Incongruence between mtDNA and nuclear data in the freshwater mussel genus *Cyprogenia* (Bivalvia: Unionidae) and its impact on species delineation

Taxonomic uncertainty is a major challenge, because species or other management units have to be accurately delimited before a proper conservation strategy for protecting the organism in question can be developed. The first chapter of my dissertation involves two species in the freshwater mussel genus *Cyprogenia*, *C. aberti* and *C. stegaria*. *Cyprogenia aberti* is found west of the Mississippi River, while *C. stegaria* is a federally-designated endangered species found in rivers to the east of the Mississippi River. Previous molecular studies have described two deeply diverged and sympatrically distributed mitochondrial DNA (mtDNA) lineage in *Cyprogenia* that did not correspond to current taxonomic designations, suggesting a need for major systematic revision in this genus. Generation and analysis of microsatellite data supported three allopatric clusters that correspond to major hydrologic basins (Ozark, Ouachita, and Ohio river basins) (Chong *et al.* 2016). This research further suggests that different mtDNA lineages of *Cyprogenia* are associated with the color of conglutinates (associated with reproduction), and appear to be under negative frequency-dependent selection.

Chapter 2. Comparing the gene flow patterns of the endangered scaleshell (*Leptodea leptodon*), the widely distributed fragile papershell (*Leptodea fragilis*) and their host-fish the freshwater drum (*Aplodinotus grunniens*)

Genetic variation affects long-term adaptability of populations to environmental changes such as diseases and human-caused stressors. Understanding gene flow patterns
among populations allows conservation biologists to determine the impact of habitat fragmentation and identify population sources-sinks for management purposes. The effect of habitat fragmentation on biodiversity maybe difficult to evaluate (Fahrig 2003), however, it often yields negative impacts on aquatic organisms (Dunham et al. 1997; Hovel & Lipcius 2001; Rizkalla & Swihart 2006). By assessing gene flow among populations, population fragmentation stemming from habitat loss can be identified and mitigated to reduce the risk of losing genetic variation in the isolated populations through inbreeding depression and genetic drift. In this chapter, I have compared the gene flow patterns of a federally endangered mussel species *Leptodea leptodon* with its common sister species *Leptodea fragilis*, to those of their sole host fish, the freshwater drum *Aplodinotus grunniens*. Our results indicated that the population structures of all three species are not correlated. However, there is substantial gene flow between the extant populations of *Le. leptodon* comparable to the common *Le. fragilis* (Chong & Roe 2016, in preparation). This study has provided insight into the cause of decline of the federally endangered *Le. leptodon* by ruling out barriers to gene flow or lack of available host fish as potential threats to the species.

**Chapter 3. Assessing sperm and zygote-mediated gene flow in freshwater mussels (Bivalvia: Unionidae)**

Male-mediated gene flow has been commonly studied in plants (pollen dispersal) (Ellstrand *et al.* 1989; Ellstrand 1992; Watrud *et al.* 2004; Messeguer *et al.* 2006) and in some animals with strong sex-biased dispersal (Pusey 1987; Pardini *et al.* 2001; Goudet *et al.* 2002). For mammalian species, i.e. primates, bears, and cervids, male gene flow can be estimated through Y chromosome markers (Hurles *et al.* 1999; Li *et al.* 2005; Bidon *et al.* 2014). However, paternally inherited markers are often not available for invertebrates. Ennos
(1994) described an approach that indirectly estimated male gene flow by comparing differentiation at bi-parentally inherited nuclear loci (from microsatellites or SNPs) with differentiation at maternally-inherited markers (such as mitochondrial DNA). This chapter of my dissertation attempts to investigate male gene flow in freshwater mussels using this method. Bivalves in general, and freshwater mussels in particular, are considered to be sessile organisms and possess limited mobility as adults. Reproduction in freshwater mussels is remarkable in that female mussels retain unfertilized eggs in modified portions of their gills called marsupia. Male mussels release sperm into the water that is taken in through the incumbent siphon of the female mussels and used to fertilize their eggs. The fertilized eggs remain in the marsupia of the females until the larvae (glochidia) are mature. Glochidia must parasitize an appropriate fish host to complete their development and become juveniles. Based on this life history, there are essentially two types of gene flow among freshwater mussel populations: male-mediated gene flow (through sperm dispersal) and zygotic gene flow (glochidia dispersal through host fish movement).

Despite several decades of study, the scientific community still only has a rudimentarily understanding regarding the impact of glochidia dispersal in freshwater mussels (Lefevre & Curtis 1912; Kat 1984; Watters 1994; Barnhart et al. 2008). However, we have an even more limited understanding of male-mediated gene flow via sperm dispersal (Barnhart & Robert 1997; Ishibashi et al. 2000; Christian et al. 2007). In my dissertation, I have adopted population genetic approaches commonly used to estimate pollen dispersal in plants in an attempt to develop an innovative approach for estimating sperm gene flow in freshwater mussels. This study will be the first to examine sperm gene flow in freshwater mussels using mtDNA and nuclear markers. In this study we would like to determine
whether the gametic phase (sperm dispersal) or the parasitic phase (glochidia dispersal) of mussels contributes the majority of gene flow among mussel populations.

REFERENCES


Watters GT (1994) *An annotated bibliography of the reproduction and propagation of the Unionoidea (primarily of North America).* Ohio Biological Survey, Columbus, Ohio.

CHAPTER II

INCONGRUENCE BETWEEN MTDNA AND NUCLEAR DATA IN THE FRESHWATER MUSSEL GENUS CYPROGENIA (BIVALVIA: UNIONIDAE) AND ITS IMPACT ON SPECIES DELINEATION

This is a published manuscript in peer-reviewed scientific journal, *Ecology and Evolution*

Jer Pin Chong, John L. Harris, Kevin J. Roe

Abstract

Accurately identifying species is a crucial step for developing conservation strategies for freshwater mussels, one of the most imperiled faunas in North America. This study uses genetic data to re-examine species delineation in the genus *Cyprogenia*. Historically, *Cyprogenia* found west of the Mississippi River have been ascribed to *Cyprogenia aberti* (Conrad 1850), and those east of the Mississippi River were classified as *Cyprogenia stegaria* (Rafinesque 1820). Previous studies using mitochondrial DNA sequences indicated that *C. aberti* and *C. stegaria* were not reciprocally monophyletic groups, suggesting the need for systematic revision. We generated a novel data set consisting of ten microsatellite loci and combined it with sequence data from the mitochondrial ND1 gene for 223 *Cyprogenia* specimens. Bayesian analysis of the ND1 nucleotide sequences identified two divergent clades that differ by 15.9%. Members of these two clades occur sympatrically across most sampling locations. In contrast, microsatellite genotypes support recognition of three allopatric clusters defined by major hydrologic basins. The divergent mitochondrial lineages are highly correlated with the color of the conglutinate lures used by mussels to attract and infest host fishes, and tests for selection at the ND1 locus were positive. We infer that the incongruence between mtDNA and microsatellite data in *Cyprogenia* may be the result of a combination of incomplete lineage sorting and balancing selection on lure color.
Our results provide further evidence that mitochondrial markers are not always neutral with respect to selection, and highlight the potential problems of relying on a single-locus-marker for delineating species.

**Introduction**

Taxonomic uncertainties are a major challenge to the conservation of endangered species because errors in the delineation of species may lead to flawed management decisions and incorrect estimates of biodiversity (Goldstein et al. 2000; Isaac et al. 2004; Frankham 2010). The delineation of species can be based on morphological, ecological, behavioral, and genetic information (Coyne and Orr 2004). Molecular taxonomy uses DNA sequences to identify molecular operational taxonomic units (MOTUs) and clarify taxonomic uncertainties by grouping morphologically cryptic organisms into distinct genetic entities (Vogler and Monaghan 2007). However, an increasing number of studies have shown that gene trees generated using mitochondrial data are often incongruent with gene trees constructed using nuclear data (e.g. Sota and Vogler 2001; Weins et al. 2010). Gene trees and species trees can be incongruent for a number of reasons including gene duplication (paralogy), introgression or hybridization between lineages (Doyle 1992; Degnan and Rosenberg 2009), and incomplete lineage sorting (Ting et al. 2008; Rodríguez et al. 2010; Hausdorf et al. 2011; Hobolth et al. 2011). Differentiating between the possible causes of incongruence is not always straightforward, and often is not attempted (Toews and Brelsford 2012).

Freshwater mussels are often considered to be keystone species in the freshwater benthic community (Aldridge et al. 2007; Geist 2010). As filter-feeders, they serve an
important functional role in the river ecosystem through enhancing nutrient cycling and increasing habitat richness for the benthic community (Vaughn and Hakenkamp 2001). Degradation of water quality and other human activities have led to the imperiled status of many freshwater species, and unionid mussels are among the most endangered faunas in the world (Williams et al. 1993; Stein and Flack 1997; Master et al. 1998; Haag 2012). Freshwater mussels are unique among bivalves in that they have a parasitic stage in their lifecycle where the larvae, termed glochidia, attach to a vertebrate host for a period of several weeks (Lefevre and Curtis 1912). Within the North American Unionidae, members have evolved many spectacular methods of attracting their fish-hosts, including packaging their larvae to resemble food items and inducing the host to infest themselves by consuming the mock food item (Kat 1984). Prior to the advent of molecular markers, freshwater malacologists relied primarily on conchological characters (e.g. shell shape, size, and color) for mussel species identifications (Simpson 1914; Haas 1969). However, lineages identified using morphological characters alone have been shown to not always be congruent with evolutionary lineages identified using molecular markers (Roe and Lydeard 1998; Inoue et al. 2013).

The geographic range of the freshwater mussel genus *Cyprogenia* occurs within the Mississippi faunal province (Burr and Mayden 1992), and includes the Eastern, Ozark, and Ouachita highland regions that are characterized by high-gradient streams with coarse substrates and cool water temperatures (Mayden 1988). These regions also exhibit a high degree of faunal endemism (reviewed by Hoagstrom et al. 2014). Current taxonomy recognizes two species in the genus *Cyprogenia*: the Fanshell *Cyprogenia stegaria* (Rafinesque 1820), and the Western Fanshell *Cyprogenia aberti* (Conrad 1850). *Cyprogenia*
*stegaria* is listed as a federally endangered species (USFWS 1991) and is found east of the Mississippi River in tributaries of the Ohio River Basin, whereas the range of *C. aberti* is west of the Mississippi River in the Arkansas, White, Black, and Ouachita river basins (Oesch 1995; Harris et al. 2009). The original species descriptions indicated distinct conchological differences between these two species; however, specimens resembling intermediate forms of both species have been encountered in both the White and Ouachita river drainages in Arkansas (Harris et al. 2009). In *Cyprogenia*, mature glochidia are packaged, along with unfertilized eggs, into structures called conglutinates that resemble worms and facilitate host infection (Fig. 1). The mature glochidia are almost completely transparent, and the color of the conglutinate lure results from the pigmentation of unfertilized eggs (Eckert 2003; Barnhart et al. 2008). In *Cyprogenia*, the colors of conglutinates observed to date include brown, red, and white.

Previous molecular studies of *Cyprogenia* using mitochondrial DNA (mtDNA) sequences have indicated that both *C. aberti* and *C. stegaria* are not reciprocally monophyletic groups (Serb 2006; Grobler et al. 2011). Serb (2006) reported two monophyletic groups within *Cyprogenia*, but each clade included individuals of both nominal species. The two evolutionarily distinct clades (14% sequence divergence) reported were sympatric in several drainages. In addition, these two mtDNA lineages seemed to be correlated with the color of the conglutinate lures. *C. aberti* specimens that produced red-colored conglutinates grouped into one mtDNA clade, and those with brown conglutinates grouped into the other clade (Serb and Barnhart 2008). These observations led to the development of the hypothesis that the mitochondrial lineages of *Cyprogenia* might be maintained by negative frequency-dependent selection by host fish (Barnhart et al. 2008,
Serb and Barnhart 2008). Under this hypothesis host fish learn to avoid conglutinates of the abundant color, and instead select the less common form. Grobler et al. (2011) obtained similar results for their mtDNA analyses, but the microsatellite markers they included (only for the *C. stegaria* specimens) showed little differentiation.

Previous studies have raised doubts about the validity of the two species of *Cyprogenia*. For this study, we employed both mtDNA sequences and nuclear microsatellite loci in an explicit test of alternative hypotheses concerning the number of evolutionary entities within the genus. Our study improves on previous efforts in that we have combined both mitochondrial and nuclear data for both *C. aberti* and *C. stegaria* samples, which had not been achieved previously, and increased the numbers of sampling sites and sample sizes over previous efforts. Finally, we discuss the implications of our findings regarding the reliability of mtDNA markers and for the conservation and management of *Cyprogenia*.

**Materials and Methods**

**Sample collection and DNA extraction**

A total of 223 *Cyprogenia* samples were included in this study. 144 *Cyprogenia aberti* individuals were collected in 2010 and 2011 in collaboration with the Arkansas Game and Fish Commission, Arkansas State Highway and Transportation Department, Missouri Department of Conservation, and U.S. Fish and Wildlife Service (Fig. 2 and Table S1, Supporting Information). Samples for DNA extraction were collected non-destructively using cytology brushes (Henley et al. 2006), and genomic DNA was extracted using the Puregene Buccal Cell Kit (Qiagen). The color of the conglutinate lures was recorded if female individuals were collected during the breeding season. Additional genomic DNA was
obtained from 26 individuals (24 *C. aberti* and two *C. stegaria*), from Serb (2006), and 53 *C. stegaria* individuals from Grobler et al. (2011) (Table S1, Supporting Information).

**Mitochondrial DNA sequencing and analysis**

A ~900 base-pair fragment of the first subunit of the mitochondrial NADH dehydrogenase (ND1) gene was successfully amplified via PCR for 206 *Cyprogenia* (157 *C. aberti* and 49 *C. stegaria*) using primers described in Serb (2006). PCR was conducted using a 25 µL reaction volume, with 0.2 mM dNTPs, 1x Biolase buffer, 2.5 mM MgCl₂, 0.8 µM primers and 1.25 U Biolase Taq polymerase (Bioline Inc.). Cycling parameters included an initial denaturation at 95°C for 5 min, followed by 35 cycles of 94°C for 30s, 52°C for 30s, 72°C for 60s, and a final extension of 72°C for 4 min. PCR products were purified using ExoSAP-IT reagent (USB Corp.) and were bi-directionally sequenced on an ABI 3730x1 DNA Analyzer at the Iowa State University (ISU) DNA Facility. Thirty-eight *C. stegaria* ND1 haplotype sequences from Grobler et al. (2011) were downloaded from GenBank to obtain a total of 244 *Cyprogenia* sequences. In addition, ND1 sequences for 18 outgroup taxa were also obtained from GenBank and included in the phylogenetic analysis (Table S2, Supporting Information).

ND1 sequences were aligned using ClustalW and default parameters as implemented in the software Geneious Pro v.5.5.6 (Drummond et al. 2010). Sequences were translated into amino acids in order to check for stop codons, indels, and ambiguous nucleotides. MrModeltest v.2.3 (Nylander 2004) was used to perform hierarchical likelihood ratio tests to determine the appropriate nucleotide substitution models for Bayesian analysis. Bayesian inference was conducted in MrBayes v.3.2.1 (Ronquist and Huelsenbeck 2003) using
Markov chain Monte Carlo simulations of 22 million generations with tree sampling every 100 generations and a burn-in of 100,000 generations. The burn-in and length of MCMC simulations was determined by the stability of the mean standard deviation of split frequencies between two independent runs. Each run consisted of four chains (three hot and one cold) with a temperature difference of 0.2. A consensus tree was constructed by including all the post burn-in sampled trees, with nodal support indicated by posterior probabilities. Pairwise genetic distances were calculated within and between each clade to evaluate sequence dissimilarity using the Kimura 2-parameter model in program MEGA v.5.10 (Tamura et al. 2011).

Phylogenetic constraint analyses were performed to test whether two alternate tree topologies were as good or better fit to the data than the optimal trees obtained from the Bayesian analysis. The first alternative tree was constrained based on current taxonomy, by forcing all individuals east of the Mississippi River and all individuals west of the Mississippi River to form separate monophyletic groups (taxonomy). The second constraint tree forced individuals from the same hydrologic basins to form monophyletic groups (basins). Constrained trees were generated in MrBayes v.3.2.1 as above, but for 10 million generations and burn-in of 50,000. The Shimodaira-Hasegawa (1999) test implemented in PAUP* v.4.0b10 (Swofford 2002) was used to compare the likelihoods of these two constraint trees with the unconstrained tree that we had generated earlier with Bayesian inference to see which topology was better supported.

DNA sequences were converted into haplotypes using program DnaSP v.5.10.01 (Librado and Rozas 2009). Nucleotide diversity (π) and haplotype diversity (Hd) were estimated for each population. A haplotype network employing the optimality criteria of
parsimony was generated in the program Network v.4.613 and Network Publisher v.2.0.0.1 (Fluxus Technology, Ltd). For samples for which we were able to collect both mtDNA data and conglutinate color information, a Yule’s $Q$ contingency coefficient was calculated to determine the correlation between the mussel conglutinate colors and membership within the two mtDNA clades (Yule 1900). Neutrality of the ND1 sequences was examined to see if the gene is under selection as has been previously hypothesized. Deviations from neutrality were examined using the codon based Z-test of selection [$H_0: d_N = d_S$] (Nei and Gojobori 1986) and Tajima’s $D$ (Tajima 1989) implemented in MEGA v.5.10 (Tamura et al. 2011) and DnaSP v.5 (Librado and Rozas 2009) respectively.

**Microsatellite genotyping and analysis**

Ten microsatellite loci (Ecap1, Ecap2, Ecap4, Ecap5, Ecap6, Ecap7, Ecap8, Ecap10, PfaD06, LabD213) (Eackles and King 2002; Galbraith et al. 2011) were amplified for 216 individuals from 12 populations. An M13-tag (5’-AGGGTTTTCCAGTCACGACGTT-3’) was added to the 5’ end of the forward primer for all loci. For some microsatellite loci, an additional sequence (GTTTCTT) was added to the 5’ end of the reverse primer to promote adenylation and reduce one base pair stutter (Brownstein et al. 1996). Microsatellite reactions consisted of 0.2mM dNTPs, 1x Biolase buffer, 1.5mM MgCl$_2$, 0.2µM of M13 dye-labelled primer and non-tagged reverse primer, 0.02µM of M13-tagged forward primer, 0.25 U Biolase Taq polymerase, and 15ng template DNA in a 20 uL total reaction volume. PCR reactions were performed with initial denaturation at 95°C for 5 min, followed by 35 cycles of 94°C for 30s, 55°C annealing temperature for 30s (except for marker Ecap8, where we used 60°C), 72°C for 30s, and a final extension of 72°C for 4 min. Ten percent of the
samples were chosen randomly and replicated for each microsatellite locus to characterize and reduce genotyping errors as described in Meirmans (2015). PCR products for microsatellite genotyping were sent to the ISU DNA Facility and analyzed using an ABI 3730 DNA Analyzer.

Microsatellite alleles were scored using GeneMarker™ Software (Softgenetics, PA). Micro-Checker (Van Oosterhout et al. 2004) was used to examine the data from each marker for genotyping errors and presence of null alleles. Allelic richness, heterozygosity, and inbreeding coefficient were estimated for samples collected from the same locality using GenAlEx v.6.5 (Peakall and Smouse 2006; Peakall and Smouse 2012). Sampling sites with fewer than five individuals were not included in these analyses because allele frequencies could not be properly estimated due to low sample size. Deviation from Hardy-Weinberg equilibrium and linkage disequilibrium were tested for all loci in GENEPOP v.3.3 (Raymond and Rousset 1995). BOTTLENECK v.1.2 (Piry et al. 1999) was used to detect rapid changes in population size due to demographic factors. The two-phase model (TPM) with a fixed proportion of 70% single-step model (SSM) and 30% variance of geometric distribution was implemented with 1,000 iterations for sampling sites with more than eight individuals. We used STRUCTURE v.2.2 (Pritchard et al. 2000) to assign individuals into populations. A total of 100,000 MCMC replicates were performed using an admixture model with burn-in of 50,000 runs. The correlated allele frequencies model was selected to detect refined population structure. The most likely number of populations was estimated by determining the likelihood of K following Evanno et al. (2005) using STRUCTURE HARVESTER (Earl and vonHoldt 2012); the number of K tested ranged from K = 1 to K = 20. CLUMPP v.1.1.2 (Jakobsson and Rosenberg 2007) and DISTRUCT v.1.1 (Rosenberg 2004) were used to
construct the STRUCTURE barplot to infer the genetic structure of both species. We performed AMOVA of the genotype data in GenAlEx to determine whether partitioning of the genetic variation at the 10 microsatellite loci was consistent with the pattern obtained in the optimal mtDNA tree or either constrained tree (individuals constrained based on taxonomy, clades or hydrologic basins). The Akaike information criterion (AICc) was used to choose the best-supported hypothesis, following Halverson et al. (2008), which would select the model with the regional factor that contributed most to the total amount of genetic differentiation. Pairwise genetic differentiation between groups was estimated using F-statistics, standardized analogs ($F_{ST}$) and Jost’s $D$ (Weir and Cockerham 1984; Jost 2008; Bird et al. 2011; Meirmans and Hedrick 2011) using the program GenoDive v.2.0b23 (Meirmans and Van Tienderen 2004).

**Results**

*Mitochondrial DNA analyses*

The GenBank accession numbers for all new sequences including samples re-sequenced from Serb (2006) were KU687119-KU687320. Our phylogenetic analysis of the ND1 data resulted in a monophyletic *Cyprogenia* that consisted of two deeply diverged clades that were distributed sympatrically across most of our sampling sites (Fig. 2). These two clades differed by an average genetic distance of 15.9%, and are referred to as Clade A and Clade B following Serb (2006) (Fig. 3). Clades A and B consisted of 133 and 111 individual ND1 sequences, respectively. Both Clade A and Clade B included individuals that were morphologically identified as *C. aberti* and *C. stegaria*. Clade A is distributed across all our sampling locations (Fig. 2). Clade B co-occurs with Clade A at all sites, with the
exception of two sites in Kansas and one in Tennessee. Three subclades within each major
clade were identified using the posterior probabilities from the Bayesian inference and
genetic distances (Fig. 3). These subclades were estimated to be 2-4% divergent. The first
subclade within Clade A (Subclade A1) was the most widely distributed, with individuals in
seven drainages: Black, St. Francis, Spring (AR), Licking, Green, Salt, and Clinch rivers.
The second subclade (Subclade A2) included populations from the Black, Ouachita, Spring
(AR), Saline, and Caddo rivers. Members in the third subclade (Subclade A3) were found
only in the Fall and Spring (KS) rivers in Kansas. Clade B also consisted of three subclades
(Fig. 3). Subclade B1 included samples from the Black, St. Francis, Spring (AR), Current,
Buffalo, and Strawberry rivers. Subclade B2 included individuals from the Black, Ouachita,
Saline, Current, White, and Caddo rivers. Subclade B3 was limited to Kentucky populations
from the Licking, Green, and Salt rivers.

All mtDNA sequences included in this study were grouped into 71 haplotypes.
Overall nucleotide diversity (π) was 0.072 with high haplotype diversity Hₐ = 0.94. The
numbers of haplotypes, nucleotide and haplotype diversity are presented in Table 1. Despite
the co-occurrence of individuals from Clades A and B across the range of Cyprogenia, most
of the haplotypes were observed in a single drainage basin, with only two out of 71
haplotypes shared between the Ozark highlands (Black, St. Francis, and Spring AR rivers)
and Eastern highlands (Licking, Green, Salt, and Clinch rivers) (Fig. 4). Results of the
Shimodaira-Hasegawa (1999) test comparing the optimal mtDNA tree recovered from the
phylogenetic analysis with the trees constrained by either basins or taxonomy indicated that
the unconstrained tree was significantly better than either constrained topology (p < 0.001)
(Table 2).
Conglutinate color information was recorded for gravid female mussels during sampling. Due to the time of year when samples were collected, gravid female mussels were observed only when sampling the populations from the Ozark and Ouachita regions. Of the 180 samples that were collected, conglutinate color was observed for 53 individuals (29.4%). Of these, 21 individuals (39.6%) had red conglutinates, and 32 (60.4%) had brown conglutinates. No white conglutinates were observed during our sampling. 20 of the 21 females with red conglutinates were placed in clade A, and 31 of the 32 females with brown conglutinates were placed in clade B. Based on the Yule’s $Q$ contingency coefficient, mussel conglutinate colors and the mtDNA lineages (A vs. B) were strongly correlated ($Q = 0.997$).

Analysis of the ND1 sequences for evidence of deviation from neutrality using Tajima’s $D$ statistic ($D = 3.10$, $p < 0.001$) and in MEGA v.5.10 ($Z = -11.22$, $p < 0.001$) both rejected the null hypothesis of neutrality.

**Microsatellite genotyping analyses**

The results from examination of the microsatellite data using Micro-Checker (Van Oosterhout et al. 2004) indicated that locus Ecap1 may include a null allele. In order to assess the potential impact of including this locus in future analyses, we performed STRUCTURE and AMOVA analyses with and without including the Ecap1 locus. Inclusion of the Ecap1 locus did not alter the results of either analysis, and so this locus was retained in all further analyses. All loci were tested for Hardy-Weinberg equilibrium (HWE) using exact test and the default settings of GENEPOP (Guo and Thompson 1992). Hardy-Weinberg disequilibrium was detected at sites 3, 4, for locus LabD213, and at site 12 for locus Ecap1 after applying the Bonferroni correction ($p < 0.0004$). Gametic disequilibrium was not
evident in any pairs of loci ($p > 0.05$), and no evidence of a bottleneck was found in any population. Standard population genetics measures including allelic richness, genetic diversity, and inbreeding coefficient were estimated for each site (Table 1). The STRUCTURE analysis identified three clusters ($K = 3$) that corresponded to the hydrologic basins (Fig. 5). Only four of 208 individuals included in the study did not group according to their geographical region (Table 1). The first cluster (Ouachita) consisted of 49 individuals mainly from the Ouachita and Saline rivers. The second cluster (Ozark) was the largest, with 108 samples from the Black, St. Francis, and Spring (AR) rivers. The third cluster (Eastern) included 51 samples from four rivers: the Licking, Green, Salt, and Clinch rivers. Following Halverson et al. (2008), we compared the results from the AMOVA analysis to alternative hypotheses by grouping samples to reflect the same hypotheses tested for the mtDNA data (taxa, clades, or basins) to determine which contributed the most to the total amount of nuclear genetic differentiation. The Akaike information criterion (AICc) clearly indicated that grouping samples based on hydrologic basins was significantly better than the other models (Table 3).

Estimates of population structure indicated that the Ouachita cluster was genetically more similar to the Ozark cluster than the geographically more distant Eastern cluster. Pairwise $F_{ST}$ among the three distinct groups of mussels identified by the microsatellite data were estimated to be 0.09 to 0.15 among clusters (Table 4). However, $F_{ST}$ has been shown to provide underestimates of genetic differentiation when using highly polymorphic loci such as microsatellites (Jost 2008; Meirmans and Hedrick 2011). Unbiased estimators such as $F'_{ST}$ and Jost’s $D$ have been shown to provide more accurate measures of genetic differentiation when polymorphism is high. The values for $F'_{ST}$ and Jost’s $D$ (Table 4) indicated that the
pairwise genetic differentiation among clusters was dramatically higher (0.4-0.6), suggesting very limited recent gene flow between different hydrologic basins.

**Discussion**

*Discordance between mtDNA and nuclear markers*

Mitochondrial gene sequences have been a standard molecular marker for inferring phylogenetic relationships between species and phylogeographic patterns within species. However, in recent years, an increasing number of studies have documented incongruence between patterns produced by analysis of mtDNA and nuclear DNA (Lu et al. 2001; Sota and Vogler 2001; McCracken and Sorenson 2005; Rodríguez et al. 2010; Toews et al. 2014). In order to investigate the apparent incongruence further we conducted additional analyses on each data set. We tested the congruency of the results of the mtDNA data by using constraint analysis to impose the results of the microsatellite analysis onto the mtDNA data set, and we tested the microsatellite data by grouping the genotypes according to the results of the mtDNA analysis, conducting another AMOVA analysis, and comparing these results to the original analysis using the AICc. The results of the Shimodaira-Hasegawa test of the DNA sequences, and the AICc test for the microsatellite data clearly showed that the patterns recovered from the analysis of the mtDNA data and the microsatellite data were incongruent with each other. In a recent review of this topic, Toews and Brelsford (2012) identified 126 published cases of discordance between phylogeographic patterns produced by mtDNA and nuclear DNA markers. In the cases they reviewed, mito-nuclear discordance was reported by researchers to be due to several different factors, including incomplete lineage sorting, introgressive hybridization, and retention of ancestral polymorphisms (Ting et al. 2008;
An additional potential source of incongruence unique to some bivalve lineages is doubly uniparental inheritance (DUI) of mitochondria, in which sex-associated mitochondrial lineages are inherited maternally or paternally (Zouros et al. 1994; Liu et al. 1996). In DUI, male mussels contain both male and female mtDNA lineages, although the male mtDNA lineage is largely restricted to the gonads, and the female mtDNA lineage is found in the somatic tissue. Female mussels possess only the female mtDNA lineage in both gonads and somatic tissue. Incongruence between our mtDNA and microsatellite results could be obtained if the male mtDNA lineage was accidently amplified and sequenced for a subset of samples. The resulting mtDNA phylogeny would then consist of two divergent lineages (one male and the other female). We are confident that this is not the case in our study. We obtained samples for DNA extraction by collecting cells from the mantle and foot using cytology brushes, thus avoiding gonadal tissue. Additionally, male and female mtDNA lineages in unionid mussels have been shown to evolve at dramatically different rates, and differences between the two lineages can exceed 30% sequence divergence (Breton et al. 2007). A phylogenetic comparison of male and female mtDNA sequences in Cyprogenia (not shown) indicates all mtDNA sequences included in this study are from the female lineage.

Biased introgression in mtDNA can also result in the mito-nuclear discordance, and could be caused by sex-biased dispersal, assortative mating, and sexual selection (Chan and Levin 2005). Based on the biology of freshwater mussels, sex-biased dispersal is an unlikely explanation for the observed incongruence between the two markers. Like most freshwater mussel species, Cyprogenia are dioecious (Haag 2012), but there is no evidence that male and female glochidia larvae are dispersed different distances by their host fishes. Sex-biased
dispersal has also not been documented in adult mussels. Similarly, male freshwater mussels broadcast sperm, and no mechanisms whereby female mussels can “choose” between the sperm of different males have been proposed, which appears to eliminate sexual selection and assortative mating as explanations for the observed incongruence.

In the process of incomplete lineage sorting, the discordance between the patterns recovered for mtDNA and nuclear DNA may be explained by the different rates at which the two types of markers fix new mutations (Martinsen et al. 2001; Ballard and Whitlock 2004). Because of its smaller effective population size, mtDNA should fix new mutations and undergo lineage sorting faster than nuclear DNA (Ballard and Whitlock 2004). In *Cyprogenia* however, it appears that it is in fact the mtDNA that is exhibiting incomplete lineage sorting. One way to distinguish discordance arising from incomplete lineage sorting from discordance arising from other factors is that incomplete lineage sorting should not produce predictable biogeographic patterns (Funk and Omland 2003; Toews and Brelsford 2012). Our microsatellite data strongly indicated that there are three distinct nuclear DNA clusters within *Cyprogenia* that conform to biogeographic provinces (Ozark, Ouachita, and Eastern basins), with two divergent mtDNA lineages occurring within each of these clusters (Table 3 and 4). In contrast, the 71 mtDNA haplotypes did not display any strong biogeographic patterns (Fig. 4); therefore, we cannot rule out the possibility that the mito-nuclear discordance in *Cyprogenia* was due to incomplete lineage sorting of mtDNA.

An increasing number of studies have indicated that mtDNA appears to be under selection (e.g., Grant et al. 2006; Stewart et al. 2008), and it is conceivable that the disparate mtDNA lineages in *Cyprogenia* have been maintained in sympatry via selection. It was suggested previously (Barnhart et al. 2008, Serb and Barnhart 2008) that the mtDNA
lineages in *Cyprogenia* were somehow linked to the colors of the conglutinate lures, and the color polymorphism present in populations may maintained by negative frequency-dependent selection on conglutinates by the host fishes. The results of our more extensive examination of conglutinate color with a larger sample size concurred with Serb and Barnhart (2008) that mtDNA clade membership was highly correlated with the color of the conglutinates.

Furthermore, additional evidence from the codon based Z test and the Tajima’s test indicates that the two mtDNA clades are under selection. Positive Tajima’s $D$ values are consistent with balancing (negative frequency dependent) selection, but can also result from demographic effects, such as a recent bottleneck, or population subdivision (Maruyama and Fuerst 1985; Simonsen et al. 1995), demographic causes should leave their signature on the nuclear genome as well. Our tests of the nuclear microsatellite data for a recent bottleneck however, did not support a demographic explanation for the divergent haplotypes: tests for a bottleneck were negative, and population subdivision was minimal when specimens were grouped according to conglutinate color ($F_{ST} = 0.005$). Taken together, these results are consistent with the hypothesis that the divergent mtDNA lineages are being maintained in sympathy by negative frequency dependent selection imposed by host fish on the conglutinate color in *Cyprogenia*.

Our microsatellite data support recognizing three independent evolutionary lineages in genus *Cyprogenia* that correspond to the Ozark, Ouachita, and Eastern Highland regions of North America, respectively (Fig. 5). This same biogeographic pattern has been observed in a number of other aquatic taxa that are also highland endemics (e.g., Strange and Burr 1997; Ray et al. 2006; Berendzen et al. 2008). The Central Highlands of North America once consisted of a single region characterized by clear, cool high-gradient streams that
subsequently was fragmented by a series of glacial cycles (Thornbury 1965; Pflieger 1971). The highlands became isolated into three major areas separated by intervening lowlands, and today are composed of the Ozark and Ouachita highlands west of the Mississippi River, and the Eastern Highlands containing the Appalachian Mountains east of the Mississippi River (Mayden 1988). Gene flow of freshwater mussels is considered to depend largely on the dispersal ability of their host fishes (Watters 1992; Haag and Warren 1997). Host-fish dispersal in riverine ecosystems often can be limited by physiographic barriers such as natural features (falls, etc.) or unsuitable habitat. These barriers may create biogeographic islands by reducing gene flow among hydrologic basins, resulting in population structure that reflects hydrologic basins, as is seen in the western populations of anodontine freshwater mussels (Mock et al. 2010). The host fishes of Cyprogenia that have been identified via laboratory studies include: Cottus carolinae, Etheostoma blennioides, E. caeruleum, E. flabellare, E. spectabile squamosum, Percina caprodes, P. phoxocephala, and P. roanoka (Jones and Neves 2002; Eckert 2003), and all of these fishes occupy habitats that are typical of the highland regions inhabited by Cyprogenia. The restricted movement of host fishes for Cyprogenia between the three highland regions is supported by population genetic analyses of a number of species that indicate reduced gene flow between and within basins (i.e. Echelle et al. 1975; Turner et al. 1996; Turner and Trexler 1998; Ray et al. 2006; Haponski et al. 2009). Limited movement of host fishes between basins would restrict gene flow between mussels inhabiting these same basins, and additional evidence for limited gene flow between mussel populations in different basins is seen in the reduced suitability of allopatric vs sympatric host-fishes (Eckert 2003). In that study, fishes that were sympatric with the mussels tested transformed a higher proportion of mussel larvae to the juvenile stage than
fishes of the same species collected in different basins. The lack of shared mtDNA haplotypes between *Cyprogenia* inhabiting the three regions and the high degree of differentiation at the microsatellite loci are consistent with biogeographic scenarios that these regions became isolated during the late Miocene or Pliocene (Hoagstrom et al. 2014).

**Species delimitation and conservation implications**

Freshwater mussels are among the most endangered faunas in North America, and species delimitation is an important first step in understanding the significance of variation in conchological characteristics, reproductive strategies, habitats, and host fish requirements for these understudied organisms. Accurate delimitation of evolutionary lineages is important for the efficient use of conservation resources and the long-term preservation of biodiversity. Within *Cyprogenia* the Eastern Highland group identified in our study conforms to the existing range of *C. stegaria*. *Cyprogenia stegaria* is a federally endangered species (USFWS 1990) and reproducing populations are now restricted to the Licking, Green, and Salt rivers in Kentucky and the Clinch River in Tennessee and Virginia. Within the former range of *C. aberti*, two distinct clusters occur. The Ouachita cluster includes populations from the Ouachita and Saline rivers, whereas the Ozark cluster includes populations from Black, St. Francis, and Spring (AR) rivers. Harris et al. (2009) recommended that the status of *C. aberti* in Arkansas be changed from Threatened to Endangered, and our results indicate that a further re-examination of the conservation status of this species is warranted. At the present time, we recommend that Ozark and Ouachita clusters be treated as distinct evolutionarily significant units (ESUs) *sensu* Crandall et al. (2000) due to lack of ecological exchangeability as evidenced by apparent adaptation to local host fishes (Eckert 2003) and
genetic differentiation demonstrated in this study. Such a designation would recognize the ecological and genetic distinctiveness of these entities for management purposes until additional morphological and genetic data can be used to more directly test their status as distinct species. The genetic data generated for each of the sampling sites indicates that at present, the levels of genetic diversity as measured by allelic richness and expected heterozygosity are high, and there is an absence of substantial inbreeding at all sites sampled. Any plans to propagate and translocate individual *Cyprogenia* should not include transfer of individuals between these three distinct lineages and potentially risk introducing new alleles that may result in hybridization and out-breeding depression, which could have detrimental consequences.

**Conclusions**

A substantial number of phylogeographic and population genetic studies have been conducted on freshwater mussels using mitochondrial markers (e.g. King et al. 1999; Roe et al. 2001; Roe 2013; Zanatta and Harris 2013). A standard assumption is that mitochondrial genes are largely neutral markers and as such are well suited to reconstructing the evolutionary relationships of organisms (Avise et al. 1987). Our research provides another example that mitochondrial markers are not always neutral with respect to natural selection, and so may reflect a biased evolutionary history. Mitochondrial and nuclear markers in *Cyprogenia* revealed two very different geographic patterns, and our investigation indicates that the DNA sequences of the mitochondrial ND1 gene are highly correlated with the color of the conglutinate lures in *Cyprogenia* and tests we conducted are consistent with balancing selection. An investigation of the molecular basis for conglutinate colors and the impact of
conglutinate colors on host-fish choice have the potential to further test this hypothesis. Based on the analysis of ten microsatellite loci, we conclude that there are currently three independent evolutionary entities in *Cyprogenia* and we recommend that these are treated as a distinct species in the case of *C. stegaria*, and ESUs in the case of the entities in the Ozark and Ouachita basins. We are currently studying conchological variation in *Cyprogenia* shells using 3D morphometrics to compare shell morphology within and between the lineages defined by our genetic analyses. We are also investigating the basis of conglutinate colors and the relationship between conglutinate color and mitochondrial variation in hopes of improving our understanding of this fascinating system.

**Acknowledgements**

Funding for this research was provided by U.S. Fish and Wildlife Service and the National Institute of Food and Agriculture. We would like to acknowledge Chris Davidson, Bill Posey, Josh Seagraves, Stephen McMurray, Scott Faiman, David Hayes, and Jacob Culp for assisting with the field sampling collection. We would like to thank Jeanne Serb, Paul Grobler, and Jess Jones for sharing DNA samples with us. Sincere gratitude to John Nason and Alvin Alejandrino for their assistance with some of the analyses. We thank Chris Barnhart for sharing the conglutinate image. We are also grateful for Dean Adams, Julie Blanchong, Jeanne Serb, Chris Barnhart and other reviewers for their helpful comments on this manuscript.
Table 2-1. Summary of mtDNA ND1 gene and microsatellite diversity from *Cyprogenia* samples grouped according to current taxonomy or corresponding to the clusters resulting from the STRUCTURE analysis. Site corresponds to ID in Fig. 2 and Table S1 in Supporting Information. Sites with fewer than five individuals were not included in this table. MtDNA sequences of 194 samples were grouped into 71 haplotypes. The number of individuals (N<sub>seq</sub>), number of haplotypes (H), nucleotide diversity (π) and haplotype diversity (H<sub>d</sub>) are listed below. Microsatellite genotyping data presented here were collected from 208 individuals. The number of individuals (N<sub>msat</sub>), allele richness (A), observed heterozygosity (H<sub>O</sub>), Nei’s (1978) unbiased expected heterozygosity (H<sub>E</sub>), and inbreeding coefficient (F<sub>IS</sub>). The number of individuals from each sampling locality assigned to population clusters by STRUCTURE are listed in parentheses.

<table>
<thead>
<tr>
<th>Current taxonomy/clusters</th>
<th>Site</th>
<th>Region/Basin</th>
<th>MtDNA ND1 sequencing</th>
<th>Microsatellite Genotyping</th>
<th>STRUCTURE cluster</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>N&lt;sub&gt;seq&lt;/sub&gt;</td>
<td>H</td>
<td>π</td>
</tr>
<tr>
<td><em>C. aberti</em></td>
<td>3</td>
<td>Ozark</td>
<td>28</td>
<td>6</td>
<td>0.0315</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Ozark</td>
<td>28</td>
<td>9</td>
<td>0.0642</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Ozark</td>
<td>19</td>
<td>8</td>
<td>0.0561</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Ozark</td>
<td>8</td>
<td>6</td>
<td>0.0341</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Ozark</td>
<td>17</td>
<td>10</td>
<td>0.0626</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>Ouachita</td>
<td>27</td>
<td>13</td>
<td>0.0563</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>Ouachita</td>
<td>5</td>
<td>4</td>
<td>0.0758</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>Ouachita</td>
<td>13</td>
<td>7</td>
<td>0.0707</td>
</tr>
<tr>
<td><em>C. stegaria</em></td>
<td>15</td>
<td>Eastern</td>
<td>23</td>
<td>8</td>
<td>0.0571</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>Eastern</td>
<td>8</td>
<td>5</td>
<td>0.0685</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>Eastern</td>
<td>8</td>
<td>6</td>
<td>0.0549</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>Eastern</td>
<td>10</td>
<td>8</td>
<td>0.0046</td>
</tr>
</tbody>
</table>
### Table 2-1 continued

<table>
<thead>
<tr>
<th>Cluster</th>
<th>Range</th>
<th>Taxon</th>
<th>Value 1</th>
<th>Value 2</th>
<th>Value 3</th>
<th>Value 4</th>
<th>Value 5</th>
<th>Value 6</th>
<th>Value 7</th>
<th>Value 8</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ozark cluster</td>
<td>3-7</td>
<td>Ozark</td>
<td>100</td>
<td>27</td>
<td>0.065</td>
<td>0.865</td>
<td>108</td>
<td>20</td>
<td>0.753</td>
<td>0.794</td>
</tr>
<tr>
<td>Ouachita cluster</td>
<td>12-14</td>
<td>Ouachita</td>
<td>45</td>
<td>21</td>
<td>0.062</td>
<td>0.908</td>
<td>49</td>
<td>13</td>
<td>0.675</td>
<td>0.723</td>
</tr>
<tr>
<td>Eastern cluster</td>
<td>15-18</td>
<td>Eastern</td>
<td>49</td>
<td>24</td>
<td>0.055</td>
<td>0.940</td>
<td>51</td>
<td>17</td>
<td>0.760</td>
<td>0.805</td>
</tr>
</tbody>
</table>

- Ozark (106), Ouachita (1), Eastern (1)
- Ouachita (47), Ozark (1), Eastern (1)
- Eastern (51)
Table 2-2. The results of Shimodaira-Hasegawa Test (SH-test) on the likelihoods of constraint and unconstraint trees corresponding to three *a priori* models. The first model constrained individuals into *C. aberti* and *C. stegaria* based on current taxonomy identified by morphological and geographical locations. The second model was an unconstrained tree that identified two distinct mtDNA clades using optimal Bayesian analysis. The third model constrained individuals based on major hydrologic basins.

<table>
<thead>
<tr>
<th>Model</th>
<th>Tree constraint</th>
<th>-ln L</th>
<th>Diff in ln L</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>taxa (east, west)</td>
<td>7224</td>
<td>443</td>
<td>&lt; 0.001</td>
</tr>
<tr>
<td>2</td>
<td>mtDNA clades A and B (unconstrained)</td>
<td>6781</td>
<td>(best supported)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>hydrologic basins</td>
<td>7715</td>
<td>934</td>
<td>&lt; 0.001</td>
</tr>
</tbody>
</table>

Table 2-3. Results of AMOVA analysis performed on individuals grouped according to three *a priori* models. The Akaike information criterion (AICc) indicated that hydrologic basin was the regional factor that contributed the most to the total amount of genetic differentiation in microsatellite data.

<table>
<thead>
<tr>
<th>Model</th>
<th>Region</th>
<th>n</th>
<th>SSR</th>
<th>Est. Variance</th>
<th>% variance</th>
<th>$F'_{RT}$</th>
<th>P</th>
<th>AICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>taxa (east, west)</td>
<td>208</td>
<td>877.6</td>
<td>0.392</td>
<td>9%</td>
<td>0.484</td>
<td>0.001</td>
<td>307.65</td>
</tr>
<tr>
<td>2</td>
<td>mtDNA clades</td>
<td>192</td>
<td>910.4</td>
<td>0</td>
<td>0%</td>
<td>0</td>
<td>0.999</td>
<td>307.04</td>
</tr>
<tr>
<td>3</td>
<td>hydrologic basins</td>
<td>208</td>
<td>819.6</td>
<td>0.452</td>
<td>10%</td>
<td>0.480</td>
<td>0.001</td>
<td>293.43*</td>
</tr>
</tbody>
</table>

Table 2-4. Genetic differentiation estimators $F_{ST}$, $F'_{ST}$ and Jost's $D$ calculated for *Cyprogenia* populations among Ouachita, Ozark, and Eastern (Licking, Green, Salt, and Clinch rivers) clusters.

<table>
<thead>
<tr>
<th>Regional/Basin</th>
<th>$F_{ST}$ (Wright 1951)</th>
<th>$F'_{ST}$ (Meirmans 2006)</th>
<th>Jost's $D$ (Jost 2008)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Ouachita</td>
<td>Ozark</td>
<td>Eastern</td>
</tr>
<tr>
<td>Ouachita</td>
<td>49</td>
<td>0.00</td>
<td>0.09</td>
</tr>
<tr>
<td>Ozark</td>
<td>108</td>
<td>0.09</td>
<td>0.00</td>
</tr>
<tr>
<td>Eastern</td>
<td>51</td>
<td>0.14</td>
<td>0.11</td>
</tr>
</tbody>
</table>
Figure 2-1. Red and brown conglutinates produced by *Cyprogenia*. Conglutinates are clusters of mussel larvae and unfertilized eggs that some mussels produced to lure host fish. Photo credit: Chris Barnhart.
**Figure 2-2.** Map of *Cyprogenia* mtDNA lineage distribution in our sampling locations. *Cyprogenia aberti* (triangles) were collected from Arkansas, Missouri, and Kansas. *Cyprogenia stegaria* (circles) were collected from Kentucky and Tennessee. Drainages of sampling sites are labeled: (1) Fall, (2) Spring River, Kansas, (3) St. Francis, (4,5) Black, (6,7) Spring River, Arkansas, (8) Current, (9) Buffalo, (10) Strawberry, (11) White, (12) Ouachita, (13) Caddo, (14) Saline, (15) Licking, (16) Salt, (17) Green, (18) Clinch. *Cyprogenia* mtDNA Clade A (blue) occurred in all sampling sites. Clade B (red) co-occurred with Clade A in most sampling sites except in Kansas and Tennessee populations.
Figure 2-3. Bayesian inference gene tree based on the mitochondrial ND1 gene of two *Cyprogenia* species (-ln L = 6927.24). Two major mtDNA clades (A and B) differed by a genetic distance of 15.9%. Clade A includes 133 sequences and Clade B includes 111 sequences. Three subclades were identified in both Clade A and Clade B, with genetic distances ranged from 2~4% among subclades. Geographical distributions of samples for each subclade were listed by river drainage (State). Colors of clades are the same as in Figure 2.
Figure 2-4. Median-joining haplotype network for the ND1 sequences. Haplotypes are color-coded based on major hydrologic basins. Ouachita drainages (yellow), Ozark drainages (green), and Eastern drainages (blue) includes the Licking, Green, Salt, and Clinch rivers. Each node represents a unique haplotype, and node size indicates the number of individuals sharing the same haplotype. Black nodes represent inferred mutational events occurring between haplotypes. Star symbol indicates the separation point between mtNDA clades A and B. For n>4 mutational events, numbers next to interrupted lines are used to indicate the number of mutations.
Figure 2-5. Geographic distribution of Cyprogenia ESUs based on results of STRUCTURE analysis. Bar plot generated from the microsatellites data shows Cyprogenia individuals grouped into three clusters (Ouachita, Ozark, and Eastern) corresponding to the major hydrologic basin.
REFERENCES


Maruyama T, Fuerst PA (1985) Population bottlenecks and non-equilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. *Genetics*, **111**, 675-689.


Rodríguez F, Pérez T, Hammer SE, Albornoz J, Domínguez A (2010) Integrating phylogeographic patterns of microsatellite and mtDNA divergence to infer the evolutionary history of chamois (genus *Rupicapra*). *BMC Evolutionary Biology, 10*, 222.


Simpson CT (1914) *A descriptive catalogue of the naiades or pearly fresh-water mussels*. Bryant Walker, Detroit, Michigan.


Yule GU (1900) On the association of attributes in statistics: with illustrations from the material of the childhood societies, &c. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, 194, 257-319.


SUPPLEMENTAL INFORMATION

Supplemental Table 2-S1. Sample sizes from all Cyprogenia sampling locations included in this study. A total of 223 Cyprogenia individuals were included. DNA of 26 C. aberti and two C. stegaria (individuals from sites 1-14, and 18) were obtained from Serb (2006). DNA of 53 C. stegaria individuals (sites 15-18) were obtained from Grobler et al. (2011). The remaining 144 C. aberti individuals (sites 3-7, 12, 14) were collected by the authors.

<table>
<thead>
<tr>
<th>Species</th>
<th>Site ID</th>
<th>Drainage</th>
<th>State</th>
<th>Sample #</th>
<th>Source of samples</th>
</tr>
</thead>
<tbody>
<tr>
<td>C. aberti</td>
<td>1</td>
<td>Fall River</td>
<td>KS</td>
<td>2</td>
<td>Serb (2006)</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>Spring River</td>
<td>KS</td>
<td>1</td>
<td>Serb (2006)</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>St. Francis River</td>
<td>MO</td>
<td>27</td>
<td>This study, Serb (2006)</td>
</tr>
<tr>
<td></td>
<td>4</td>
<td>Black River</td>
<td>MO</td>
<td>28</td>
<td>This study, Serb (2006)</td>
</tr>
<tr>
<td></td>
<td>5</td>
<td>Black River</td>
<td>AR</td>
<td>17</td>
<td>This study, Serb (2006)</td>
</tr>
<tr>
<td></td>
<td>6</td>
<td>Spring River</td>
<td>AR</td>
<td>8</td>
<td>This study, Serb (2006)</td>
</tr>
<tr>
<td></td>
<td>7</td>
<td>Spring River</td>
<td>AR</td>
<td>17</td>
<td>This study, Serb (2006)</td>
</tr>
<tr>
<td></td>
<td>8</td>
<td>Current River</td>
<td>AR</td>
<td>1</td>
<td>Serb (2006)</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Buffalo River</td>
<td>AR</td>
<td>1</td>
<td>Serb (2006)</td>
</tr>
<tr>
<td></td>
<td>10</td>
<td>Strawberry River</td>
<td>AR</td>
<td>2</td>
<td>Serb (2006)</td>
</tr>
<tr>
<td></td>
<td>11</td>
<td>White River</td>
<td>AR</td>
<td>1</td>
<td>Serb (2006)</td>
</tr>
<tr>
<td></td>
<td>12</td>
<td>Ouachita River</td>
<td>AR</td>
<td>23</td>
<td>This study, Serb (2006)</td>
</tr>
<tr>
<td></td>
<td>13</td>
<td>Caddo River</td>
<td>AR</td>
<td>5</td>
<td>Serb (2006)</td>
</tr>
<tr>
<td></td>
<td>14</td>
<td>Saline River</td>
<td>AR</td>
<td>13</td>
<td>This study, Serb (2006)</td>
</tr>
<tr>
<td>C. stegaria</td>
<td>15</td>
<td>Licking River</td>
<td>KY</td>
<td>23</td>
<td>Grobler et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>Salt River</td>
<td>KY</td>
<td>8</td>
<td>Grobler et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>17</td>
<td>Green River</td>
<td>KY</td>
<td>8</td>
<td>Grobler et al. (2011)</td>
</tr>
<tr>
<td></td>
<td>18</td>
<td>Clinch River</td>
<td>TN</td>
<td>8</td>
<td>Grobler et al. (2011), Serb (2006)</td>
</tr>
</tbody>
</table>

Supplemental Table 2-S2. Outgroup taxa included in the Bayesian analysis with GenBank accession numbers:
- Actinonaias ligamentina (AY655085)
- Amblema plicata (AY158796)
- Dromus dromas (AY158750)
- Epioblasma brevidens (AY094378)
- Epioblasma capsaeformis (DQ208591)
- Lampsilis cariosa (EF446096)
- Lampsilis fasciola (DQ220721)
- Lampsilis higginsii (EF213061)
- Lampsilis ornate (AY158748)
- Lampsilis ovata (AY613797)
- Lampsilis siliquoidea (AY158747)
- Lemiox rimosus (AY655104)
- Ligumia recta (EF213055)
- Medionidus conradicus (AY158746)
- Obliquaria reflexa (AY158751)
- Potamilus alatus (AY655119)
- Ptychobranchus fasciolaris (AY655120)
- Villosa iris (DQ445185)
CHAPTER III

COMPARING THE GENE FLOW PATTERNS OF THE ENDANGERED SCALESHELL (LEPTODEA LEPTODON), THE WIDELY DISTRIBUTED FRAGILE PAPERSHELL (LEPTODEA FRAGILIS) AND THEIR HOST-FISH THE FRESHWATER DRUM (APLODINOTUS GRUNNIENS)

Manuscript prepared for submission to peer-reviewed journal

Jer Pin Chong and Kevin J. Roe

Abstract

The larvae of freshwater mussels in the order Unionoida are obligate parasites on fishes. Because adult mussels are infaunal and largely sessile, it has generally been assumed that the majority of gene flow between mussel populations relies on the dispersal of larvae by host fishes. Host specificity in parasites has been shown to often lead to congruence between the population structures of the parasite and its host. The objective of this study was to determine the degree of congruence between the population structure of two Leptidea species (Le. leptodon and Le. fragilis) and their fish host, Aplodinotus grunniens. Leptidea leptodon is a narrowly distributed, federally endangered species, whereas Le. fragilis is common and widely distributed. Assessing the congruence of genetic structure of Le. leptodon with its sister species Le. fragilis and their fish host is an important step in understanding the impact of host dispersal on structuring populations of Le. leptodon. We collected over 300 samples of the three species from the same five locations from the Gasconade, Bourbeuse, and Meramec rivers in Missouri. Both mtDNA and microsatellite data indicated that the population structures of all three species were not congruent with each other. Despite its imperiled status, Le. leptodon displayed levels of allelic richness similar to those exhibited by Le. fragilis. It also appears that Le. leptodon like Le. fragilis, exhibits
substantial gene flow among the populations examined. This study has provided insight into the cause of decline of the federally endangered *Le. leptodon* by ruling out barriers to gene flow as potential threats to the species. Instead, habitat specificity may play a role in the differences in population structures of the two mussel species.

**Introduction**

The direct or indirect interactions between species are the basis of the emerging properties of ecosystems and the varied ecological processes that occur within them. Interspecific interactions have evolutionary importance because interacting species affect not only each other’s ecological roles but also their evolutionary trajectories (i.e. the Red Queen hypothesis) (Lively et al. 1990; Morran et al. 2011). Interactions in which two or more species undergo reciprocal evolutionary changes represent cases that the participating species are more closely tied and are generally referred to as coevolutionary (Thompson 1999). Examples of some of these interactions include those between plants and pollinators (Grant and Grant 1965; Levin 1985), predators and prey (Brodie and Brodie 1990, 1991), and hosts and parasites (Dybdahl and Lively 1996; Gigord et al. 2001; Nason et al. 2002). Investigations into the population structure of parasites and their hosts have revealed a variety of patterns that appear to vary with such factors as dispersal ability of hosts and parasites (Blouin et al. 1995; McCoy et al. 2003), host specificity of the parasite (Johnson et al. 2002; Little et al. 2006; Dick and Patterson 2007), and the ecological requirements of the species involved (Campbell et al. 1974; Brooks et al. 2006). However, instances in which the host specificity of a parasite is high would tend to predict congruence between the population structures of the parasite and its host (Distel et al. 1994).
Freshwater mussels (Bivalvia: Unionoida) and their fish hosts represent a system that is underutilized in the study of the effects that hosts can have on the population structures of parasitic partners. Freshwater mussels are unique among bivalves in that they incorporate a parasitic stage into their life cycle by which the larvae, termed glochidia, must attach to a vertebrate host for a period of several weeks (Lefevre and Curtis 1912). Once attached to the appropriate host, the glochidia become encysted within the fish tissue (Coker et al. 1921; Telda and Fernando 1969). While embedded in a cyst, the mussel larva undergoes metamorphosis and develops the anatomy that allows it to become a filter-feeding juvenile mussel (Kat 1984; Roe et al. 1997). After metamorphosis is completed, the juvenile mussel ruptures the cyst and drops to the sediment to join the infaunal community. It has been considered that, compared to their largely sessile adult phase, the parasitic glochidium phase is largely responsible for the dispersal between populations as well as the establishment of new populations (Watters 1992). The distance of glochidial dispersal greatly depends on the mobility potential of the host fish. Some fish species like logperch may move only 30 m (Schwalb et al. 2011), while host fish like freshwater drum can travel up to 104 km (Funk 1955). Freshwater mussels can be broadly categorized as either host generalists, which can parasitize a number of different host species, or host specialists which can only parasitize a single host species (Haag and Warren 1997). If a mussel species infests only a single species of host fish, it is logical to assume that the genetic structure of the mussel populations should largely be congruent with the genetic structure of the host fish.

In this study, we compared the genetic structure of two freshwater mussel species, the federally endangered, *Leptodea leptodon* (Rafinesque 1820), its common congener, *Leptodeafragilis* (Rafinesque 1820) and their sole host fish (Barnhart et al. 1998), the freshwater drum
Aplodinotus grunniens (Rafinesque 1819). Leptodea leptodon was once widely distributed across central North America, found in 56 rivers across 13 states within the Mississippi River Basin. At present, it is restricted to only 13 streams and all known populations occur in Missouri, with the largest known populations occurring in the Gasconade, Bourbeuse, and Meramec rivers in Missouri (Oesch 1995; Szymanski 1998; USFWS 2010). In contrast, its sister species, Le. fragilis is widely distributed, ranging from Canada south to Mexico and from the Appalachian Mountains west to South Dakota, and is generally considered to be common throughout much of its range (Burch 1975; Clarke 1981).

The sole host fish for both species, Aplodinotus grunniens, displays the widest natural distribution of any freshwater fish in North America (Lee et al. 1980). Individual fish have been recaptured about 104 km from their tagged locality (Funk 1955), and appear to fit the established model of fishes being the primary dispersing agents for mussel larvae.

The objective of this study was to determine the degree of congruence between the population structure of the two Leptodea species and A. grunniens. Assessing the congruence of genetic structure of Le. leptodon with its sister species Le. fragilis and their fish host is an important step in understanding the impact of the host on structuring populations of Le. leptodon. We hypothesized that, if host dispersal is the primary determinant of population structure in freshwater mussels, then the population structure of both mussel species should be highly congruent, as they share the same host fish.

Materials and Methods

Sample Collection and DNA Extraction

We collected samples for DNA extraction from a total of 123 Le. leptodon, 104 L.
fragilis, and 126 A. grunniens from the same five locations: two locations in the Gasconade River (sites A and B), one location in the Bourbeuse River (site C), and two locations in the Meramec River (sites D and E) (Table 1, Fig. 1). Mussel samples for extraction of genomic DNA were collected non-lethally by swabbing the foot and mantle, following Henley et al. (2006). Aplodinotus grunniens specimens were captured with the assistance of the Missouri Department of Conservation using a Smith-Root VI-A Electrofisher powered by a 5,000-watt generator in a 16-foot jon boat. After the fish were landed, a 2-mm diameter fin clip was collected and stored in 95% EtOH. Total DNA of mussel species was extracted using the Puregene Buccal Cell DNA Kit (Qiagen), and DNA was extracted from the fin clips using the Qiagen DNeasy Blood and Tissue Kit (Qiagen). The concentration of extracted genomic DNA was measured using a Nanodrop ND1000 spectrophotometer.

**Mitochondrial DNA Amplification and Analysis**

We amplified ~870 base pairs (bp) of the mitochondrial ND1 gene of the two mussel species (117 Le. leptodon and 100 Le. fragilis) via polymerase chain reaction (PCR) using primers LEUuurF and LoglyR from Serb (2006). We amplified ~1100 bp of the mitochondrial cytochrome b gene from 125 A. grunniens using the Cytb F and Cytb R primers from Song et al. (1998). PCR reactions included initial denaturing at 95 °C for 5 min, followed by 35 cycles of 94°C for 30s, 52°C for 30s, 72°C for 60s, and a final extension of 72°C for 4 min. Each PCR reaction consisted of 25 µl of reaction mix, with 12.5 µl of MyTaq (Bioline Inc.), 1 µl of 10 µM primer, 1 µl of ~100 ng of genomic DNA, and 9.5 µl of ddH2O. PCR products for all three species was purified using the ExoSAP-IT (USB Corp.) before being bi-directionally sequenced at the Iowa State University DNA facility with an
ABI 3730xl DNA Analyzer (Applied Biosystems). DNA sequences were edited using the software Geneious Pro v.5.5.6 (Drummond et al. 2010). Sequences were aligned using ClustalW as implemented in Geneious using default settings and were trimmed to equal length. We translated the nucleotide sequences to amino acids to check for alignment errors and other point mutations such as insertions/deletions. Nucleotide sequences were then grouped into haplotypes using DnaSP v.5.10.01 (Librado and Rozas 2009). The number of haplotypes, nucleotide diversity, and haplotype diversity were estimated for each population. A median-joining haplotype network was generated for each species using the programs Network v.4.613 and Network Publisher v.2.0.0.1 (Fluxus Technology, Ltd) based on optimality criteria of parsimony to infer the genetic similarity among haplotypes.

**Microsatellite Genotyping and Analysis**

*Assessing genetic diversity*

We first assessed the nuclear gene diversity within each species separately. Ten microsatellite markers (Lele3, Lele7, Lele8, Lele9, Lele13, Lele16, Lele24, Lele30, Lele47, Lele48) from O’Bryhim et al. (2012) were used to genotype 122 *Le. leptodon* samples. The same 10 microsatellites markers (except for replacing Lele16 with Lele18) were used to genotype 102 *Le. fragilis* individuals. We optimized 11 microsatellite loci (Soc508, Soc509, Soc510, Soc521, Soc524, Soc536, Soc543, Soc551, Soc558, Soc588, Soc626) originally developed for red drum (Karlsson et al. 2008) to genotype the 125 *A. grunniens* samples. PCR reactions for all loci consisted of a total of a 20-µl volume mix comprising 0.2 mm of dNTPs, 1x Biolase buffer, 1.5 of mM MgCl2, 0.25 U of Biolase *Taq* polymerase, 0.2 µM of M13 dye-labelled primer and non-tagged reverse primer, 0.02 µM of M13-tagged forward
primer, and 15 ng template DNA. A standard annealing temperature of 55°C was used for all PCR reactions for microsatellite loci with the following exceptions, for which a 60°C annealing temperature was used: Lele3, Lele7, Lele8, Lele18 for *Le. leptodon* and Lele7, Lele8, Lele18, Lele48 for *Le. fragilis*.

Micro-Checker was used to assess the presence of null alleles in each marker (Van Oosterhout et al. 2004). A linkage disequilibrium test was performed on all microsatellite loci using GENEPOP v.3.3 (Raymond and Rousset 1995). GenAlEx v.6.502 (Peakall and Smouse 2006; Peakall and Smouse 2012) was used to calculate the allelic richness, observed and expected heterozygosity, and inbreeding coefficient for each sampling location within each species. We were also interested to know if the mussels and/or host fish populations experienced recent rapid changes in population size due to demographic factors. A recent bottleneck may significantly alter genetic structure and connectivity among local populations, thus affecting the congruency of population structures among the mussels and their host fish. A bottleneck test was conducted on all three species on each sampling site within each species using the BOTTLENECK v.1.2 program (Piry et al. 1999) with 1000 iterations of a fixed proportion of 70% single-step model and 30% variance of geometric distribution implemented in the two-phase model. Populations that have recently experienced bottleneck will exhibit a loss in allelic richness, which results in heterozygosity excess relative to the expected heterozygosity estimated as the mutation–drift equilibrium.

**Genetic clustering analysis and detection of first-generation migrants**

Next, we examined the population structures within each species using clustering analyses. Population structure was estimated using STRUCTURE v.2.2 software to cluster
individuals into populations based on the criteria that these groups are in Hardy–Weinberg and linkage equilibrium (Pritchard et al. 2000) and using the sample location as a prior (LOCPRIOR). An admixture model was used allowing individuals to have mixed ancestry, with 100,000 Markov chain Monte Carlo replicates and a burn-in of 50,000 runs. The correlated allele frequencies model was selected so that the refined population structure could be detected. The number of populations (K) was estimated following methods developed by Evanno et al. (2005), using STRUCTURE HARVESTER (Earl and vonHoldt 2012). A STRUCTURE bar plot was constructed using CLUMPAK (Kopelman et al. 2015).

Individuals were then grouped into genetic clusters according to the clustering results obtained from STRUCTURE. Genetic differentiation among genetic clusters was estimated using the analysis of molecular variation (AMOVA) (Excoffier et al. 1992) to obtain pairwise $F_{ST}$. The fixation index $F_{ST}$ is known to underestimate the genetic differentiation of highly polymorphic loci (such as microsatellites) because it is estimated based on the expected heterozygosity, which does not increase linearly when diversity increases (Jost 2008; Meirmans and Hedrick 2011). As a result, unbiased estimator $F_{ST}^\prime$ (Meirmans and Hedrick 2011) is used to provide more accurate measures of genetic differentiation among populations. However, estimates of $F_{ST}^\prime$ via AMOVA analysis in GenAlEx v.6.502 (Peakall and Smouse 2006; Peakall and Smouse 2012) currently does not produce a test statistic ($p$-value). In this study, pairwise linearized $F_{ST}$ (Slatkin 1995) and $F_{ST}^\prime$ values among genetic clusters were estimated with 10,000 permutations using GenAlEx v.6.502 (Peakall and Smouse 2006; Peakall and Smouse 2012). However, tests of statistical significance were only presented for the pairwise linearized $F_{ST}$ estimates.
Individual assignment tests and the likelihood of first-generation migrants were estimated using GENECLASS 2 (Piry et al. 2004) to determine the migration rate among the genetic clusters. Origins of individuals were considered to be correct if they were assigned to the genetic cluster identified in the STRUCTURE analysis. Individuals assigned to clusters that were different from the one they were collected from would indicate evidence of migration. The threshold of the assignment test was set to 5%. The likelihood of first-generation migrants ($L = \frac{L_{HOME}}{L_{MAX}}$) was calculated using Bayesian criteria described in Rannala and Mountain (1997). For both the assignment test and detection of first-generation migrants, a Monte Carlo resampling method of 10,000 simulations was applied following Paetkau et al. (2004).

**Congruence of interpopulation divergence among species**

After assessing the genetic diversities in each species, our final step of analyses was to evaluate whether the population structures and gene flow patterns are congruent among the mussel species and host fish. We grouped individuals based on their sampling locations and used Mantel tests as implemented in the Isolation By Distance Web Service v.3.23 (Jensen et al. 2005) to measure the correlation between genetic distance and geographic distance among individuals of each species. Pairwise genetic differentiation among sample sites was estimated using linearized $F_{ST}$ (Slatkin 1995) to construct a genetic distance matrix. Pairwise geographic distances among sites was measured using river miles in Google Earth (Google Inc.) to create a geographic distance matrix. Mantel tests were conducted with 10,000 iterations to demonstrate the relationship between genetic isolation and geographic distance.
Additional Mantel tests were also performed to examine the correlation of interpopulation divergence among both mussel species and the host fish. The null hypothesis of the Mantel test (Mantel 1967) states that the distances among objects in a matrix of variables (in this case distances among individuals of species A from different sampling locations) are not linearly correlated with distances among objects (populations of species B from different sites) in another matrix of variables. Genetic structure among populations can vary depending on the measure of genetic distance used (Dyer et al. 2010), so we conducted the Mantel tests using two types of genetic distances. We calculated the pairwise genetic distances among sampling sites within each species using linearized $F_{ST}$ (Slatkin 1995) and Cavalli-Sforza and Edward's (1967) chord distances (Cavalli-Sforza and Edwards 1967). Cavalli-Sforza and Edward’s chord distance was reported to be suitable for microsatellite data (Nei et al. 1983; Takezaki and Nei 2008). Multiple genetic distances were used in the analyses to ensure a better estimate of the gene flow pattern for each species. Mantel tests of the linearized $F_{ST}$ were performed in GenAlEx v.6.502 (Peakall and Smouse 2006) with 10,000 permutations. For the Cavalli-Sforza and Edward's chord distances, we used the Monte Carlo test implemented in the R package “ade4” to conduct the Mantel tests, with 10,000 permutations. Genetic distance matrices consisted of pairwise genetic distances among sampling sites. Mantel tests were performed to examine if the distance matrices of any pair of two species were linearly correlated ($\alpha = 0.05$). Since both mussel species shared the same host fish (the same glochidia dispersal mechanism), we predicted that the distance matrices between the two mussel species would be correlated, as would the distance matrices between mussel and host fish.
Results

*Leptodea leptodon* mtDNA and Genotyping Analysis

DNA sequencing resulted in 873 bp of sequence for the *ND1* gene for 117 *Le. leptodon* specimens. Populations of *Le. leptodon* exhibited low genetic diversity (Table 2). All *Le. leptodon* mtDNA sequences were grouped into four haplotypes (Fig. 2). Overall nucleotide diversity of the *ND1* gene was 0.0006 with moderate haplotype diversity (Hd = 0.473). All four haplotypes were only a single mutational step apart. Two haplotypes were widespread and abundant and were observed at all sample sites. One haplotype was found only in the Meramec and Bourbeuse river populations, and the other haplotype was restricted to the upper Gasconade River populations; both haplotypes were observed at lower frequencies (Fig. 2).

No null alleles were detected at any microsatellite locus based on the results from Micro-Checker (Van Oosterhout et al. 2004). We compared each pair of loci and found no evidence of genetic disequilibrium (p > 0.05). Hardy–Weinberg equilibrium was tested for all loci using the exact test based on the default settings of GENEPOP (Guo and Thompson 1992). Hardy–Weinberg disequilibrium was detected at locus Lele7 for the population from site A and locus Lele8p for the populations from sites A, B, C, and E after applying Bonferroni correction (p < 0.001). Allelic richness, heterozygosity, and inbreeding coefficients were estimated for each site (Table 2). Allelic richness was similar across populations from all sampling sites, ranging between 10 and 14 alleles. The inbreeding coefficient ranged from low (< 0.03) to intermediate (< 0.12), it was twice as high in the Bourbeuse population (site C) than in the Gasconade and Meramec populations (sites A, B, D, and E). The test for the occurrence of a recent bottleneck was positive in *Le. leptodon* (p =
0.005). However, genetic isolation distances and geographic distances were not linearly correlated in *Le. leptodon* (*r* = 0.05, *p* = 0.46) (Fig. 4).

The STRUCTURE analysis for *Le. leptodon* using LOCPRIOR (with sampling site as a prior) indicated that the appropriate value of *K* = 2 separated mussels from the Gasconade River from those in the Meramec/Bourbeuse rivers (Fig. 3). The AMOVA results showed that individuals from Gasconade (sites A and B) were significantly different from individuals from the Bourbeuse and Meramec rivers (sites C, D, and E) with a linearized *F*$_{ST}$ = 0.008, *p* < 0.001 (*F*$_{ST}'$ = 0.064). The significant AMOVA results coincided with the STRUCTURE analysis, suggesting a weak separation between the Gasconade and Meramec/Bourbeuse *Le. leptodon* clusters (Table 3). Of the *Le. leptodon* individuals, 95% were correctly assigned into the genetic clusters identified from STRUCTURE analysis. The migration rate between the two clusters was estimated to be 15% (18 first-generation migrants out of 123 individuals) at a 5% probability threshold (Table 4).

**Leptodea fragilis mtDNA and genotyping analysis**

An 864 bp fragment of the *ND1* gene was amplified for 100 *Le. fragilis* samples. All *Le. fragilis* mtDNA sequences were grouped into 22 haplotypes (Fig. 2). Overall nucleotide diversity was 0.0015 with high haplotype diversity (*Hd* = 0.759). There were two common haplotypes found in populations from all five sampling locations (Fig. 2). The remaining haplotypes occurred at low frequencies and tended to be restricted to particular sampling sites. Four rare haplotypes were only found in Gasconade River (sites A and B); six haplotypes were restricted to Bourbeuse River (site C), and another six haplotypes were located only in Meramec River (sites D and E).
No null alleles or linkage disequilibrium was detected in the *Le. fragilis* microsatellite data. Allelic richness ranged from 11 to 13 for each sampling location (Table 2). After employing Bonferroni correction, Hardy–Weinberg disequilibrium was detected at locus 8 for site C; locus 9 for sites A, B and C; and locus 13 for all sampling sites \((p < 0.001)\). A test for Hardy-Weinberg equilibrium could not be performed on marker F-Lele03p due to low allelic diversity. The inbreeding coefficient was relatively similar across all sites, ranging from 0.04 to 0.08. Unlike *Le. leptodon*, no evidence of a recent bottleneck was detected in *Le. fragilis* \((p = 0.42)\). Furthermore, genetic distance and geographic distance were linearly correlated in *Le. fragilis*, with \(r = 0.67, p = 0.046\) (Fig. 4).

The STRUCTURE analysis supported two clusters \((K = 2)\) in *Le. fragilis*, with the upper Gasconade River population (site A) forming a cluster separate from populations from the remaining sample sites (sites B, C, D, and E) (Fig. 3). The results of the AMOVA analysis on the genotypic data indicated that individuals from the upper Gasconade River were genetically different from individuals from the lower Gasconade, Bourbeuse, and Meramec rivers with linearized \(F_{ST} = 0.038, p < 0.001\) \((F'_{ST} = 0.15)\) (Table 3). The proportion of first-generation migrants between the upper Gasconade site and the rest of our sampling drainages was 5 of 104, a 5% migration rate (Table 4). All the remaining 93 *Le. fragilis* individuals from the outside the upper Gasconade River region were assigned correctly to the Meramec/Bourbeuse genetic cluster. At the upper Gasconade region (site A), only 64% \((4 \text{ of } 11)\) individuals were assigned correctly to the genetic cluster identified from the STRUCTURE analysis (Table 4). The results from the assignment test suggested that *Le. fragilis* from Gasconade River site A are genetically more related to *Le. fragilis* populations beyond the upper Gasconade region.
**Aplodinotus grunniens mtDNA and genotyping analysis**

DNA sequencing resulted in 1,086 base pairs of the *cytochrome b* gene for 126 individuals of *A. grunniens*. The overall nucleotide diversity of the mtDNA gene was 0.004. The DNA sequences grouped into 11 haplotypes, which formed two mitochondrial haplotype groups that differed by nine mutational steps. The two mtDNA haplotype groups co-occurred in all five sampling sites, with haplotype diversity equal to 0.73 (Fig. 2). Three haplotypes were common and widely distributed, whereas the remaining haplotypes occurred at low frequencies and generally were restricted to within river basins. A set of two rare haplotypes were each found restricted to Gasconade and Meramec rivers. No rare haplotypes were documented in the Bourbeuse River, i.e., Bourbeuse *A. grunniens* share all haplotypes with Gasconade and Meramec populations.

Null alleles or linkage disequilibrium was not evident at any locus. Allele richness among *A. grunniens* populations was very similar, ranging between 10 and 13 alleles (Table 2). The inbreeding coefficients were very low (*F*<sub>IS</sub> < 0.003) in all populations. Hardy–Weinberg disequilibrium was not detected at any locus or in any population after applying Bonferroni correction (*p* < 0.00091). There was no evidence of a recent bottleneck in *A. grunniens* (*p* = 0.52). Genetic isolation and geographic distance were not linearly correlated in *A. grunniens* (*r* = 0.16, *p* = 0.32) (Fig. 4).

**STRUCTURE** analysis of the microsatellite data indicated *K* = 1, suggesting that *A. grunniens* from our five sampling locations belonged to a single panmictic population maintained by high gene flow and connectivity. As shown in Figure 3, the **STRUCTURE** barplot of *K* = 2 clearly indicated no genetic differentiation among *A. grunniens* populations. Because *Le. leptodon* and *Le. fragilis* both showed some level of genetic differentiation
between the Gasconade and Meramec/Bourbeuse populations, we conducted an AMOVA analysis by grouping the *A. grunniens* into these two clusters. Population differentiation analysis estimated a linearized $F_{ST}$ of 0.002 ($F'_{ST} = 0.012$), but this was not significantly different from zero ($p = 0.07$), further confirming the results from the STRUCTURE analysis that *A. grunniens* from the Gasconade, Bourbeuse, and Meramec rivers can be considered as one population (Table 4).

*Mantel test analysis on the gene flow patterns*

Although both mussel species utilize the same host fish, the Mantel test results indicated that the population structures of all three species were not entirely congruent. We conducted the Mantel tests using two types of genetic distances: linearized $F_{ST}$ and Cavalli-Sforza and Edward's chord distances. Mantel test results based on the linearized $F_{ST}$ and Cavalli-Sforza and Edward's chord distances showed no significant correlations between population structure among *Le. leptodon*, *Le. fragilis*, and *A. grunniens*, with $p$-values ranging between 0.07 and 0.46 (Table 5). Correlation of the pairwise genetic distance linearized $F_{ST}$ ($F_{ST} / 1 - F_{ST}$) among populations of all three species was shown in Figure 5. Correlation of the pairwise genetic distance estimated in Cavalli-Sforza and Edward's chord distances yielded the same result, thought the figure was not included in this study. In summary, the population structures and gene flow patterns of three species are not correlated.
Discussion

*Population Structures and Gene Flow Patterns*

Genetic differentiation has been used to infer gene flow among populations (Slatkin 1987). In this study, our primary goal was to understand the gene flow pattern of extant *Le. leptodon* populations by comparing its population structures with its congener *Le. fragilis* and host fish *A. grunniens*. Examination of the mtDNA and microsatellite data for all three species indicates low genetic differentiation among sampling sites, that is, there is substantial gene flow between the populations residing in the Gasconade, Bourbeuse, and Meramec rivers. Mitochondrial diversity was low in *Le. leptodon* with only four haplotypes detected within our sampling region, though the allelic richness measured from microsatellite data indicated that nuclear gene diversity in *Le. leptodon* was similar to the common *Le. fragilis*. Some evidence of weak population structure was observed in the two freshwater mussel species studied. In particular, *Le. leptodon* seemed to display genetic differentiation between populations that were in different drainages (Gasconade vs Meramec + Bourbeuse). For *Le. leptodon* populations in the Meramec/Bourbeuse River system to maintain continuous gene flow with populations in the Gasconade River, host fishes that carry *Le. leptodon* glochidia must travel through portions of the Missouri and Mississippi rivers. The distance between site A (Gasconade river) to site C (Bourbeuse river) is approximately 383 river miles. This separation was supported based on the STRUCTURE analysis using the microsatellite genotypes. Despite the weak differentiation observed, *Le. leptodon* populations between the Gasconade and Meramec/Bourbeuse rivers appear to be connected by substantial gene flow (migration rate = 15%). However, the genetic and geographic distances among *Le. leptodon*
populations were not linearly correlated due to higher gene flow within drainages than among drainages.

Despite sharing the same host fish, the two closely related mussel species exhibit different population structures and gene flow patterns. Unlike *Le. leptodon*, distinct population differentiation was observed between *Le. fragilis* from upper Gasconade sample site (site A) and the remaining sites (Fig. 3). The Missouri and Mississippi rivers present fewer barriers to gene flow for *Le. fragilis* than for *Le. leptodon*, perhaps due to the fact that *Le. fragilis* populations are considered to be larger and more continuously distributed. The test of isolation by distance showed that the genetic distance and geographic distance among *Le. fragilis* populations were correlated, suggesting that *Le. fragilis* exhibits a gene flow pattern distinct from *Le. leptodon*. The migration rate of *Le. fragilis* between the upper Gasconade and the lower Gasconade (plus other drainages) was relative low ≈ 5%, maybe because the movement of host fish that carry *Le. fragilis* glochidia preferentially reach the upper Gasconade River less frequently compared to other drainages. An alternative explanation is there are less suitable habitats for *Le. fragilis* between the upper and lower Gasconade River.

Our Mantel test results showed that no significant correlations between population-level genetic matrices were detected among all three species. Coevolution of host and parasites often leads to local adaptation and cospeciation (Huyse and Volckaert 2005). For example, the ability of parasites to infest a host (infectivity) is directly linked to host-parasite interaction (Dybdahl and Storfer 2003), which depends on the spatial variation of their interactions (Thompson 1994). Often, the congruency between host and parasite trees are due to high interactions/specificity (Johnson et al. 2002; Nieberding et al. 2004). Freshwater
drum (A. grunniens) is the only host fish for Le. leptodon and Le. fragilis (Barnhart et al. 1998). Aplodinotus grunniens are known for their high mobility (Funk 1955) and are common throughout the study area. Thus, the high level of gene flow observed in A. grunniens was not unexpected. Although our samples in this study only represent a small subset of the species range of A. grunniens, gene flow among the fish populations was evidently higher than gene flow among mussel populations. Leptodea fragilis has larger, more continuously distributed populations, which means it may likely have more opportunities to interact with the host fish. That leads to higher gene flow between Le. fragilis populations in Gasconade and Bourbeuse/Meramec rivers. Our results indicated that Le. leptodon populations suffered a recent bottleneck that was not evident in Le. fragilis and the host fish. It is possible that the cause of the bottleneck (e.g. habitat degradation) may have affected all three species, but since Le. leptodon is an endangered species with smaller effective population size, bottleneck would presumably have larger impact on Le. leptodon than the two other widespread species. The effect of inbreeding and genetic drift on Le. leptodon populations that may be potentially more isolated (i.e., less opportunity to interact with host fish), resulted in a population structure and gene flow pattern that are not congruent with Le. fragilis and freshwater drum.

Prior investigations comparing the population structure of freshwater mussels and their host are few in number. The first study, conducted by Geist and Kuehn (2008), examined congruence between genetic diversity and population structure between the European pearl mussel (Margaritifera margaritifera) and its host, the brown trout (Salmo trutta m. fario). Zanatta and Wilson (2011) conducted the first comparison of mussel–host population structure and diversity in North America on the federally endangered snuffbox
mussel, *Epioblasma triquetra*, and its host, the common logperch, *Percina caprodes*. No significant correlation of genetic diversity was found between the snuffbox mussel and common logperch; however, comparison of pairwise genetic distance among sites indicated that the population structure of the mussel and host were largely congruent. Our findings are not entirely concordant with previous host–mussel co-evolution studies (Geist and Kuehn 2008; Zanatta and Wilson 2011), perhaps due to the different reproductive strategies pearl mussel, snuffbox, and *Le. leptodon* employ to infest their host fish. *Leptodea leptodon* is a host specialist like snuffbox, whereas the pearl mussel is a host generalist. Pearl mussels broadcast glochidia into the water column. Snuffboxes capture host fish with their shells in order to infest them with glochidia (Barnhart et al. 2008). The interaction between *Le. leptodon* and *A. grunniens* is not yet known. *Leptodea leptodon* does not possess a mantle lure or any visible phenotypical features to attract fish. *Aplodinotus grunniens* feeds on unionid bivalves, and perhaps become infested with glochidia after consuming gravid female *Le. leptodon*. In terms of host dispersal capability, *A. grunniens*, like brown trout, have higher mobility than does common logperch. Average dispersal distance for logperch was 30 m (Schwalb et al. 2011) while *A. grunniens* may travel up to 104 km (Funk 1955). Population isolation and genetic differentiation may be evident in both logperch and snuffbox populations, results in congruency of the population structures between mussels and host. Although glochidia in *Le. leptodon* presumably has greater dispersal distance than that of the snuffbox, the glochidia gene flow of *Le. leptodon* may still be significantly lower than the gene flow among freshwater drum populations because *Le. leptodon* has a smaller effective population size. This may explain why no correlation was found between population structures of *Le. leptodon* and *A. grunniens* in contrast to snuffbox and common logperch.
Conservation implications

Our study sampled all of the known extant populations of *Le. leptodon*, and our genetic analyses provided an overview of the population structure of this species. Comparing the genetic diversity of populations of related rare and common species of mussels has rarely been performed (Roe and Boyer 2015). In Roe and Boyer (2015) study, the rare species displayed substantially lower genetic diversity when compared to its common related species. We found lower mitochondrial genetic diversity in *Le. leptodon* with regards to *Le. fragilis*, but allelic richness measured by microsatellite loci was similar between these two species. Both mussel species are dependent on their host fish for dispersal of glochidia and maintenance of gene flow. Our results indicated that the rarer *Le. leptodon* has maintained nuclear gene diversity but lower population connectivity compared to its common sister taxon. Assisted gene flow will be beneficial in establishing *Le. leptodon* populations where habitats and host fish have met the requirements for a sustainable population (Kelly and Philips 2015). A genetic rescue (Frankham 2015) to translocate *Le. leptodon* to extirpated drainages may be helpful in restoring *Le. leptodon* populations to its historic range.

Our research goal is to eventually narrow down the possible causes of decline for the endangered *Le. leptodon*. Habitat specificity may be a possible reason why *Le. fragilis* is more abundant than *Le. leptodon*, despite sharing the same host fish. *Leptodea leptodon* prefer gravel and sand substrates in large rivers, while *Le. fragilis* are more tolerant of siltation of rivers and streams of all sizes (Cummings and Mayer 1992). Experience in the field collecting samples for this study indicated that *Le. leptodon* tends to remain ~7 cm below the surface of the substrate and, for this reason, may go undetected during surveys that do not include the excavation of sample quadrats. Habitat specificity may be affecting the
expansion of extant populations of *Le. leptodon*. A recent bottleneck was evident in *Le. leptodon* but not in *Le. fragilis* or *A. grunniens*, perhaps due to the fact that habitat niches preferred by *Le. leptodon* (sandy, gravel substrate) are often found near the river banks, which are easily influenced by catastrophic events such as floods and droughts. In other words, the population size of *Le. leptodon* may fluctuate dramatically each year based on the river conditions. Yet, as long as suitable habitats and host fish are present, the source population may restore the local shrinking populations through the high gene flow observed in this study.

**Conclusions**

No published examination of the genetic structure of *Le. leptodon* and *A. grunniens* had been conducted prior to this project. *Leptodea leptodon* and *Le. fragilis* share the same host fish and live in the same rivers in parts of Missouri. Yet, *Le. leptodon* is critically endangered and *Le. fragilis* is common and widely distributed. Understanding the factors threatening *Le. leptodon* may help in understanding why some unionid mussels are more vulnerable to environmental changes. Our study revealed that the population structures of all three species were not congruent. Although *Le. leptodon* is listed as a federally endangered species, *Le. leptodon* displayed substantial levels of allelic richness and gene flow between the extant populations comparable to the common *Le. fragilis*. A recent bottleneck was evident in *Le. leptodon* but not in *Le. fragilis* and host fish. Future research should focus on understanding the factors implicated in the contraction of the historic range of *Le. leptodon*. Attempts to quantify the differences in habitat preferences between these two bivalve species are recommended to aid in mussel translocation and restoration. Ecological factors that may
affect the expansion of mussel population (i.e., mussel–host interaction mechanism, competition of host fish) are also worth exploring to develop appropriate conservation strategies for the endangered *Le. leptodon*.

**Acknowledgements**

Funding for this project was provided by the Missouri Department of Conservation and AMEREN UE. Field assistance was provided by the Missouri Department of Conservation and the U.S. Fish and Wildlife Service. In particular we would like to acknowledge the assistance of Steven McMurray, Andy Roberts, Bryan Simmons, Scott Faiman, Josh Hundley, Phillip Pitts, Tim Grace, Nick Girondo, Jennifer Girondo, Ron Burke, Kevin Meneau, Josh Ward, and Rachel Haller for sample collection. We extend our gratitude to John Nason for his advice about the data analysis. We would also like to thank Morgan Sexton, Isaac Jepsen, and Sarah Studeman, undergraduates at Iowa State University, for their laboratory assistance in generating molecular data.

**Table 3-1.** Sampling locations and sample size for *Le. leptodon, Le. fragilis, and A. grunniens* populations.

<table>
<thead>
<tr>
<th>Population</th>
<th>Drainage</th>
<th>State</th>
<th><em>Le. leptodon</em></th>
<th><em>Le. fragilis</em></th>
<th><em>A. grunniens</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Gasconade River</td>
<td>MO</td>
<td>26</td>
<td>11</td>
<td>27</td>
</tr>
<tr>
<td>B</td>
<td>Gasconade River</td>
<td>MO</td>
<td>36</td>
<td>17</td>
<td>20</td>
</tr>
<tr>
<td>C</td>
<td>Bourbeuse River</td>
<td>MO</td>
<td>22</td>
<td>28</td>
<td>27</td>
</tr>
<tr>
<td>D</td>
<td>Meramec River</td>
<td>MO</td>
<td>22</td>
<td>23</td>
<td>23</td>
</tr>
<tr>
<td>E</td>
<td>Meramec River</td>
<td>MO</td>
<td>17</td>
<td>25</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td><strong>Total:</strong></td>
<td></td>
<td><strong>123</strong></td>
<td><strong>104</strong></td>
<td><strong>125</strong></td>
</tr>
</tbody>
</table>
Table 3-2. Summary of mtDNA ND1 gene and microsatellite diversity from five *Le. leptodon*, *Le. fragilis*, and *A. grunniens* populations. The number of individuals (N_{seq}), number of haplotypes (H), nucleotide diversity (π) and haplotype diversity (H_d) were listed below. The number of individuals (N_{msat}), allele richness (A), observed heterozygosity (H_O), Nei’s (1978) unbiased expected heterozygosity (H_E), and inbreeding coefficient (F_{IS}) were presented below.

<table>
<thead>
<tr>
<th>Taxonomy</th>
<th>Pop ID</th>
<th>Drainage</th>
<th>State</th>
<th>N_{seq}</th>
<th>H</th>
<th>π</th>
<th>H_d</th>
<th>N_{msat}</th>
<th>A</th>
<th>H_O</th>
<th>H_E</th>
<th>F_{IS}</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Leptodea leptodon</em></td>
<td>A</td>
<td>Gasconade River</td>
<td>MO</td>
<td>26</td>
<td>3</td>
<td>0.0006</td>
<td>0.446</td>
<td>12.800</td>
<td>0.818</td>
<td>0.886</td>
<td>0.059</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Gasconade River</td>
<td>MO</td>
<td>35</td>
<td>2</td>
<td>0.0007</td>
<td>0.057</td>
<td>13.700</td>
<td>0.824</td>
<td>0.862</td>
<td>0.030</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Bourbeuse River</td>
<td>MO</td>
<td>22</td>
<td>3</td>
<td>0.0007</td>
<td>0.567</td>
<td>10.500</td>
<td>0.745</td>
<td>0.871</td>
<td>0.122</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Meramec River</td>
<td>MO</td>
<td>18</td>
<td>3</td>
<td>0.0008</td>
<td>0.627</td>
<td>11.700</td>
<td>0.797</td>
<td>0.858</td>
<td>0.044</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>Meramec River</td>
<td>MO</td>
<td>16</td>
<td>3</td>
<td>0.0007</td>
<td>0.575</td>
<td>10.100</td>
<td>0.859</td>
<td>0.871</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td><em>Leptodea fragilis</em></td>
<td>A</td>
<td>Gasconade River</td>
<td>MO</td>
<td>9</td>
<td>3</td>
<td>0.001</td>
<td>0.417</td>
<td>6.100</td>
<td>0.565</td>
<td>0.667</td>
<td>0.084</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Gasconade River</td>
<td>MO</td>
<td>16</td>
<td>6</td>
<td>0.001</td>
<td>0.733</td>
<td>10.900</td>
<td>0.673</td>
<td>0.747</td>
<td>0.058</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Bourbeuse River</td>
<td>MO</td>
<td>28</td>
<td>11</td>
<td>0.002</td>
<td>0.825</td>
<td>13.000</td>
<td>0.705</td>
<td>0.760</td>
<td>0.047</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Meramec River</td>
<td>MO</td>
<td>22</td>
<td>8</td>
<td>0.001</td>
<td>0.771</td>
<td>11.500</td>
<td>0.694</td>
<td>0.750</td>
<td>0.043</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>Meramec River</td>
<td>MO</td>
<td>25</td>
<td>7</td>
<td>0.001</td>
<td>0.687</td>
<td>12.100</td>
<td>0.684</td>
<td>0.751</td>
<td>0.065</td>
<td></td>
</tr>
<tr>
<td><em>Aplodinotus grunniens</em></td>
<td>A</td>
<td>Gasconade River</td>
<td>MO</td>
<td>27</td>
<td>6</td>
<td>0.003</td>
<td>0.746</td>
<td>11.364</td>
<td>0.805</td>
<td>0.784</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>Gasconade River</td>
<td>MO</td>
<td>20</td>
<td>5</td>
<td>0.003</td>
<td>0.653</td>
<td>10.545</td>
<td>0.786</td>
<td>0.796</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C</td>
<td>Bourbeuse River</td>
<td>MO</td>
<td>28</td>
<td>6</td>
<td>0.003</td>
<td>0.762</td>
<td>11.273</td>
<td>0.778</td>
<td>0.793</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td></td>
<td>D</td>
<td>Meramec River</td>
<td>MO</td>
<td>23</td>
<td>7</td>
<td>0.005</td>
<td>0.826</td>
<td>12.545</td>
<td>0.771</td>
<td>0.790</td>
<td>0.003</td>
<td></td>
</tr>
<tr>
<td></td>
<td>E</td>
<td>Meramec River</td>
<td>MO</td>
<td>28</td>
<td>5</td>
<td>0.004</td>
<td>0.664</td>
<td>11.818</td>
<td>0.802</td>
<td>0.790</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>
**Table 3-3.** Analysis of molecular variance (AMOVA) among *Le. leptodon*, *Le. fragilis*, and *A. grunniens* from Gasconade (sites A and B), Bourbeuse (site C), and Meramec (sites D and E) rivers based on the genetic clusters identified from the microsatellite STRUCTURE analyses. * indicates value is significant (*p* < 0.05).

<table>
<thead>
<tr>
<th>Species</th>
<th>Cluster 1 (site)</th>
<th>Cluster 2 (site)</th>
<th>Linearized $F_{ST}$</th>
<th>Source of variation</th>
<th>$df$</th>
<th>Sum of squares</th>
<th>Estimated variance</th>
<th>% of variance</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Le. leptodon</em></td>
<td>A,B</td>
<td>C,D,E</td>
<td>0.008*</td>
<td>Among clusters</td>
<td>1</td>
<td>9.048</td>
<td>0.035</td>
<td>1%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Among mussels</td>
<td>121</td>
<td>572.708</td>
<td>0.358</td>
<td>8%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Within mussels</td>
<td>123</td>
<td>494.000</td>
<td>4.016</td>
<td>91%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td>245</td>
<td>1075.756</td>
<td>4.410</td>
<td>100%</td>
</tr>
<tr>
<td><em>Le. fragilis</em></td>
<td>A</td>
<td>B,C,D,E</td>
<td>0.038*</td>
<td>Among clusters</td>
<td>1</td>
<td>10.203</td>
<td>0.150</td>
<td>4%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Among mussels</td>
<td>102</td>
<td>439.004</td>
<td>0.479</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Within mussels</td>
<td>104</td>
<td>348.000</td>
<td>3.346</td>
<td>84%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td>207</td>
<td>797.207</td>
<td>3.975</td>
<td>100%</td>
</tr>
<tr>
<td><em>A. grunniens</em></td>
<td>A,B</td>
<td>C,D,E</td>
<td>0.002</td>
<td>Among clusters</td>
<td>1</td>
<td>5.597</td>
<td>0.011</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Among mussels</td>
<td>123</td>
<td>535.107</td>
<td>0.005</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Within mussels</td>
<td>125</td>
<td>542.500</td>
<td>4.340</td>
<td>100%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Total</td>
<td>249</td>
<td>1083.204</td>
<td>4.356</td>
<td>100%</td>
</tr>
</tbody>
</table>
Table 3-4. The assignment tests and detection of first-generation migrants for both mussel species *Le. leptodon* and *Le. fragilis*. Host fish *A. grunniens* was not included in this table because STRUCTURE analysis and results from AMOVA found no evidence of genetic differentiation among the five sampled *A. grunniens* populations. Sampling site in this table corresponded to the site designation in Figure 1. Genetic cluster was determined by the STRUCTURE analysis using the microsatellite data. N indicates the number of samples included in each genetic cluster. Individuals were considered to be correctly assigned if individuals collected from the sampling site column were assigned to the corresponded genetic cluster, and the % of individuals correctly assigned was presented below. m represents the number of first-generation migrants that traveled between the genetic clusters, the migration rate between clusters was presented as % m.

<table>
<thead>
<tr>
<th>Species</th>
<th>Sampling site</th>
<th>Genetic cluster</th>
<th>N</th>
<th>% individual correctly assigned</th>
<th>m</th>
<th>% m</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Le. leptodon</em></td>
<td>A, B</td>
<td>Gasconade</td>
<td>62</td>
<td>100%</td>
<td>9</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td>C, D, E</td>
<td>Meramec/Bourbeuse</td>
<td>61</td>
<td>90%</td>
<td>9</td>
<td>15%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>123</td>
<td>95%</td>
<td>18</td>
<td>15%</td>
</tr>
<tr>
<td><em>Le. fragilis</em></td>
<td>A</td>
<td>Gasconade</td>
<td>11</td>
<td>64%</td>
<td>3</td>
<td>27%</td>
</tr>
<tr>
<td></td>
<td>B, C, D, E</td>
<td>Meramec/Bourbeuse</td>
<td>93</td>
<td>100%</td>
<td>2</td>
<td>2%</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Total</td>
<td>104</td>
<td>96%</td>
<td>5</td>
<td>5%</td>
</tr>
</tbody>
</table>

Table 3-5. Results of Mantel tests showed the correlations among population structures of *Le. leptodon*, *Le. fragilis*, and *A. grunniens*. Below diagonal are p-values calculated from Mantel tests conducted with linearized $F_{ST}$. Above diagonal are p-values estimated from Cavalli-Sforza and Edward's chord distances.

<table>
<thead>
<tr>
<th></th>
<th><em>Le. leptodon</em></th>
<th><em>Le. fragilis</em></th>
<th><em>A. grunniens</em></th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Le. leptodon</em></td>
<td>--</td>
<td>0.46</td>
<td>0.07</td>
</tr>
<tr>
<td><em>Le. fragilis</em></td>
<td>0.18</td>
<td>--</td>
<td>0.39</td>
</tr>
<tr>
<td><em>A. grunniens</em></td>
<td>0.11</td>
<td>0.20</td>
<td>--</td>
</tr>
</tbody>
</table>
Figure 3-1. Sampling map for *Le. leptodon*, *Le. fragilis*, and *A. grunniens* in Missouri. Sites A and B are from Gasconade River. Site C is in Bourbeuse River. Sites D and E are in Meramec River.
Figure 3-2. Mitochondrial median-joining haplotype networks constructed based on ND1 gene fragment of *Le. leptodon* and *Le. fragilis*, and cytochrome b gene fragment of *A. grunniens* from five sampling locations in Missouri. Each haplotype is represented by a node. Node size represents the number of individuals in our collection that shared the same haplotype. Black nodes (smallest node) are inferred mutational events differentiating the haplotypes. In the *A. grunniens* network, the number next to the interrupted lines indicates the number of mutational events greater than 4.
Figure 3-3. STRUCTURE barplots for *Le. leptodon*, *Le. fragilis*, and *A. grunniens* when K = 2. Individuals of each species were clustered into two groups (black and white). The bottom of the barplot was labeled with population ID and sampling drainage.

Figure 3-4. Isolation by distance figures for *Le. leptodon*, *Le. fragilis*, and *A. grunniens* species. Independent variable (x-axis) indicates the pairwise differences of the geographic distance (in river miles) between all sampling sites. Dependent variables (y-axis) indicates the pairwise genetic distances linearized $F_{ST} (F_{ST}/1-F_{ST})$ among populations.
Figure 3-5. Pairwise genetic distance linearized $F_{ST}$ ($F_{ST}/1-F_{ST}$) estimated among populations of *Le. leptodon*, *Le. fragilis*, and *A. grunniens*. 
REFERENCES


CHAPTER IV

ASSESSING SPERM AND ZYGOTE-MEDIATED GENE FLOW IN FRESHWATER MUSSELS (BIVALVIA: UNIONIDAE)

Manuscript prepared for submission to peer-reviewed journal

Jer Pin Chong and Kevin J. Roe

Abstract

Gene flow among populations can occur at different life stages through dispersal of gametes, zygotes, and adult progeny. It is important to understand how different forms of gene flow contribute to the genetic differentiation and adaptation of local populations. Contributions to gene flow from various life stages is an under-studied research topic in animals that lack internal fertilization. Here, we present an approach to estimate gamete-mediated gene flow (sperm dispersal) relative to zygote-mediated gene flow (larval dispersal) in broadcast spawning freshwater molluscs using both maternally inherited and nuclear genetic markers. We tested our approach on three freshwater mussel species Leptodea leptodon, Le. fragilis, and Lampsilis abrupta. Our preliminary results indicate that sperm dispersal in these three species is equal to or exceeds larval gene flow. For conservation management purposes, i.e., to determine where to establish newly translocated populations in a river to maintain sufficient gene flow with the extant populations, it would be crucial to include estimates of both sperm and larval gene flow to overall estimates of gene flow among populations. We hope that our study serves the purpose of initiating additional studies investigating the contributions of sperm gene flow in maintaining population connectivity in aquatic animals.
**Introduction**

Gene flow is defined as the movement of genetic material from one population to another (Slatkin 1985). When gene flow occurs, it varies across the different life stages exhibited by organisms. In some species, it tends to take place during the adult phase (Koenig *et al.* 1996; Peakall *et al.* 2003), while in others it may occur during either the gametic or zygotic phases (McCauley 1997; Sork *et al.* 1999; Heuberger *et al.* 2010). Some organisms (such as birds and mammals) with internal fertilization have no dispersal of gametes (either sperm or eggs) and gene flow is restricted to the dispersal of progeny or adults, and research on such organisms is focused on examining dispersal of progeny or adults, i.e., zygote-mediated gene flow (Hagenblad *et al.* 2009; Jacobsen & Omland 2011). In such instances, the contribution of gametic gene flow to overall gene flow can safely be ignored. One of the more common examples of investigations of zygote-mediated gene flow is the study of the bias in dispersal between individual males and females. Sex-biased dispersal in progeny is well documented in birds and mammals, and results from intra-sexual competition or as a mechanism for avoidance of inbreeding (Pusey 1987). Female-biased dispersal and male philopatry are common in birds, while mammals tend to have male-biased dispersal (Greenwood 1980). Tables 1 and 2 in Handley and Perrin (2007) summarize the examples of male-biased and female-biased dispersal in mammals. In contrast, other organisms may exhibit high gametic dispersal ability either through pollen (such as plants) (Faegri & Pijl 1971) or through sperm (e.g. aquatic invertebrates and fishes) (Cosson 2004; Bishop and Pemberton 2006). Understanding the magnitude of both gamete-mediated and zygote-mediated gene flow can aid in understanding factors that contribute to the population structure of organisms (Loiselle *et al.* 1995; Sebenn *et al.* 2011), predict genetic responses
of populations to climate change (Aitken & Whitlock 2013), and support development of management policies appropriate for species conservation (Sork et al. 1999).

Direct estimation of gametic dispersal is challenging in natural environments, particularly when paternally inherited markers (e.g. Y chromosome) are not available (Hedrick et al. 2013). Paternal gene flow can be indirectly estimated by comparing nuclear and maternally inherited markers (Ennos 1994; Hamilton & Miller 2002; Hedrick et al. 2013; Hedrick et al. 2015). Under the assumption of migration-drift equilibrium, gene flow among populations can be interpreted as $N_e m \approx 1 / 4(1/F_{ST} - 1)$, where $N_e$ refers to the effective population size and $m$ indicates the migration rate among populations (Wright 1951). For organisms that have the potential for high gametic gene flow such as plants, aquatic invertebrates and fishes, gametic gene flow can provide an estimate of male gene flow while nuclear markers provide an overall estimate of the total amount of gene flow including both gametic and zygotic gene flow. The relative contributions of each of these life history stages to gene flow can be estimated by comparing the biparentally (nuclear genome) and uniparentally (mitochondrial or chloroplast) inherited genomes (Ennos 1994; Birky 1995; Hamilton & Miller 2002). Ennos (1994) equated the nuclear and organelle markers to seed and pollen gene flow, respectively, and proposed an innovative approach for indirectly estimating the average amount of pollen and seeds migrating among populations. This approach has been applied in a variety of studies, including estimates of pollen dispersal in plants (Sork et al. 1999; Sebbenn et al. 2011) and migration patterns of male terrestrial animals (Seddon et al. 2005; Hagenblad et al. 2009; Borner & Reinsch 2010). Equations from Ennos (1994) were later modified for dioecious plant species (McCauley 1997). This
approach was adopted by Hedrick et al. (2013, 2015) to indirectly estimate male gene flow in dioecious animals by comparing maternally inherited and nuclear genetic markers.

Reproduction in freshwater mussels

Unlike animals with internal fertilization, gametic gene flow in aquatic animals may make significant contributions to impact the overall gene flow among populations. Freshwater mussels are ideal organisms for such a study because both sperm and larvae of freshwater mussels have the potential to disperse substantial distances (Watters 1992; Ferguson et al. 2013), and in some ways can be considered analogous to the dispersal of pollen and seeds in plants. The reproductive cycle of freshwater mussels includes a parasitic larval stage in which the larval freshwater mussels, called glochidia are obligate ectoparasites on fishes (Lefevre & Curtis 1912). Prior to attachment of the larvae on a host fish, male mussels release sperm into the water column and female mussels filter the sperm from the water and fertilize their eggs, which are retained inside modified portions of their gills called marsupia. Freshwater mussels are not strictly considered as internal fertilization organisms because fertilization of eggs occurs at marsupia outside of reproductive tract (Mackie 1984). Fertilized eggs develop into mature glochidia, at which time they must parasitize a host fish. Once attached and encysted in the tissue of a host fish, the glochidia metamorphose into juvenile mussels inside the cyst. After a variable period of time of up to several weeks, the juvenile mussels rupture the cyst and eventually drop off and grow into adult mussels (Jirka & Neves 1992). Most freshwater mussel species are gonochoristic, although a few species exhibit monoecy and hermaphroditism in some populations (e.g. Anodonta imbecillus, Carunculina parva, Margaritifera falcata, Lasmigona compressa) (Tepe 1943; Heard 1970;
van der Schalie 1970). Gonochoristic mussels have similar gene flow potential as dioecious plant species, while gene flow of hermaphroditic mussels resembles the monoecious plant species.

After many years of intensive study on the reproduction of freshwater mussels, a substantial understanding of gamete maturation and spawning mechanisms has been obtained (Haag 2012). Individual sperm are generally viable in water for only a few minutes, but a taxonomically diverse group of mussel species has been observed to release sperm aggregates in a spherical structure called a spermatozeugmata, in which thousands of sperm are embedded in a thin spherical membrane. These structures appear to be an adaptation for increased reproductive success (Coe 1931; Barnhart & Robert 1997; Waller & Lasee 1997) as the individual sperm in the spermatozeugmata exhibit synchronous swimming which would allow them to travel greater distances than individual sperm. Spermatozeugmata also provide an additional layer of protection to sperm (Ishibashi et al. 2000) and increases the longevity of sperm up to 48 hours in water (Falese et al. 2011). Because sperm in spermatozeugata can survive in water longer than individual sperm, they therefore have an increased opportunity to fertilize female mussels at greater distances. In Lampsilis cardium, parentage analysis has shown that spermatozeugata successfully fertilized females 16.2 km (about 10 river miles) downstream (Ferguson et al. 2013).

Based on their biology, gene flow between populations of freshwater mussels is attributable to three possible sources: the movement of sperm (gamete-mediated gene flow), movement of glochidia by the host, and movement of juvenile mussels (zygote-mediated gene flow). Freshwater mussels are generally considered to be sessile organisms that tend to remain in the same general location for their entire lives (Amyot & Downing 1997). It is
possible, however, that juvenile or adult mussels may be washed downstream during flood events (Hastie et al. 2001). For the purposes of this study, we will assume that freshwater mussels are sessile organisms, and that the movement of juveniles and adults is extremely rare so that the zygotic gene flow is equivalent to the dispersal of glochidia. Glochidia dispersal is greatly facilitated by the movement of host fishes, which exhibit much higher mobility compared to that of freshwater mussels. The gene flow patterns based on glochidia may be highly variable among mussel species (and even within a species) as it would seem to be dependent on host fish number and behavior. For instance, some host fishes such as the logperch (Percidae) generally only move within a 30 m range (Schwalb et al. 2011), whereas freshwater drum (Sciaenidae) may travel distances up to 104 km (Funk 1955). Thus, mussels with less mobile host fishes (i.e., minnows, darters, and sculpins, etc.) may exhibit a reduced amount of zygote-mediated gene flow compared to mussels that utilize host fishes with higher mobility (e.g. freshwater drum, bass, salmonids, etc.). Another major difference is that glochidia gene flow, unlike sperm gene flow, is not limited to movement in a downstream direction. Finally, the viability of glochidia outside of the female mussel is longer than sperm, and ranges from 2 to 14 days, depending on the species of mussel. Larval viability is also extended in colder water temperatures (Zimmerman & Neves 2002).

Our aim in this paper is to investigate the relative contributions of gamete-mediated and zygote-mediated gene flow of a non-plant organism using comparative analysis of nuclear and organelle genomes. We hypothesize that zygote-mediated (glochidia) gene flow will be the dominant form of gene flow among mussel populations because glochidia are able to travel greater distances on host fishes (Watters 1992) and are not limited to downstream movement. This study represents the first attempt to investigate the contribution of sperm-
mediated gene flow to the genetic structure of freshwater mollusk populations using maternally inherited and nuclear markers.

**Materials and Methods**

*Sperm gene flow estimates*

At a first glance, there is little obvious similarity between the reproductive biology of freshwater mussels and plants; however, on closer inspection the analogy is quite striking. Both organisms are sessile as adults. Male gametes in both types of organisms can disperse: male mussels release sperm into the water where it can be carried by currents, plants release pollen that travel with wind or pollinators. Eggs are fertilized and retained in the female reproductive structures of plants and female mussels until they are mature. The glochidia of mussels must attach to a host fish that can potentially disperse at a great distance; in many plant seeds are dispersed by animals. In summary, freshwater mussels have sperm (gamete) and glochidia (zygote) gene flow, corresponding to pollen and seed gene flow in plants, respectively. Ennos (1994) developed a method for comparing pollen and seed gene flow among plant populations using maternally inherited markers ($F_{ST(f)}$) and nuclear markers ($F_{ST(n)}$). The equations Ennos (1994) developed were for monoeocious plants and since freshwater mussels are dioecious organisms, we adapted them following McCauley (1997) to be able to estimate sperm and glochidia gene flow in freshwater mussels.

Assuming that $F_{ST(f)}$ = the genetic differentiation of a haploid maternally inherited gene (mtDNA), $F_{ST(n)}$ indicates the genetic differentiation of nuclear genes at equilibrium. When seed gene flow is small, for dioecious plants
Pollen flow/Seed flow = \((1/F_{ST(n)} - 1) - 4(1/F_{ST(f)} - 1)\) / \(2(1/F_{ST(f)} - 1)\) (Equation 2 in McCauley 1997)  

Substituting the variables with sperm and glochidia gene flow, let \(m_s\) = the rate of sperm gene flow and \(m_g\) = interpopulation glochidia dispersal, indicating glochidia disperse from another population and become established in the reference population, 

Sperm flow/Glochidia flow \((m_s/m_g) = [(1/F_{ST(n)} - 1) - 4(1/F_{ST(f)} - 1)] / 2(1/F_{ST(f)} - 1)\)  

Equations 1 was developed under the assumption that seed flow among populations is small (Ennos 1994). As described in Ennos (1994), let \(y\) = the proportion of established seeds that are being produced by the reference population. The pollen/seed flow ratio for dioecious organisms without prior assumption is 

Pollen flow/Seed flow = \([(1/F_{ST(n)} - 1) - 4(1/F_{ST(f)} - 1)] / 2y(1/F_{ST(f)} - 1)\)  

In theory, pollen tends to have greater dispersal distance than seeds, as demonstrated in Figure 6 of McCauley (1997). When interpopulation seed flow is small, \(y\) is close to 1, and Equation 3 can be simplified to produce Equation 1. 

In freshwater mussels, it is generally believed that glochidia are largely responsible for the movement of individuals among populations (Watters 1992), and that sperm dispersal distance is constrained to downstream movement, and is positively correlated with sperm longevity (Andre & Lindegarth 1995) and velocity of the river current. 

We let \(g\) = the proportion of glochidia produced by the reference population that develop into juvenile mussels and remain within the reference population. 

Since interpopulation glochidia flow = \(m_g\), then \(g = 1 - m_g\). Equation 2 can then be modified to include the interpopulation glochidia flow as follows 

Sperm flow/Glochidia flow \((m_s/m_g) = [(1/F_{ST(n)} - 1) - 4(1/F_{ST(f)} - 1)] / 2g(1/F_{ST(f)} - 1)\)
Sperm flow/Glochidia flow \((m_s/m_g) = \left[\frac{1}{F_{ST(n)}} - 1\right] - 4\left(\frac{1}{F_{ST(f)}} - 1\right)\right] / 2 \left(1 - m_g\right) \left(1/F_{ST(f)} - 1\right)\)

(4)

Estimating the interpopulation glochidia flow of freshwater mussels is extremely challenging in the natural environment because it requires tracking the movement of host fishes that are infested with glochidia during the mussel reproductive season, and manually checking each host fish periodically to determine whether the juveniles are being released within or outside of the reference population. Since we cannot directly measure the interpopulation glochidia flow, we will make the following assumptions. When glochidia dispersal among populations is limited, i.e., when \(m_g \approx 0\), the ratio of sperm and glochidia gene flow in mussel populations is close to the value obtained from Equation 2. However if glochidia dispersal among mussel populations is larger, \(m_g > 1\), the sperm flow/glochidia flow ratio calculated from Equation 2 will underestimate the true ratio.

**Comparing sperm flow with glochidia flow**

To compare the rate of sperm or glochidia flow in a population, we established a null hypothesis that sperm gene flow and glochidia gene flow are equal (adopted from Hamilton and Miller 2002 with the null hypothesis of pollen = seed gene flow). When \(m_s = m_g\), Equation 2 can be modified to produce

\[
\left(\frac{1}{F_{ST(n)}} - 1\right) - 4\left(\frac{1}{F_{ST(f)}} - 1\right) / 2 \left(1/F_{ST(f)} - 1\right) = 1
\]

\[
(1/F_{ST(n)} - 1) - 4\left(1/F_{ST(f)} - 1\right) = 2 \left(1/F_{ST(f)} - 1\right)
\]

\[
[(1 - F_{ST(n)}/F_{ST(f)}) - [(4 - 4F_{ST(f)})/F_{ST(f)}] = (2 - 2F_{ST(f)})/F_{ST(f)}
\]

\[
[F_{ST(f)}(1 - F_{ST(n)})/ F_{ST(n)}F_{ST(f)}] - [F_{ST(n)}(4 - 4F_{ST(f)})/ F_{ST(n)}F_{ST(f)}] = F_{ST(n)}(2 - 2F_{ST(f)})/ F_{ST(n)}F_{ST(f)}
\]

\[
F_{ST(f)} - F_{ST(n)}/F_{ST(f)} - 4F_{ST(n)} + 4 F_{ST(n)}F_{ST(f)} = 2F_{ST(n)} - 2 F_{ST(n)}F_{ST(f)}
\]
\[ F_{ST(f)} - 4F_{ST(n)} + 3F_{ST(n)}F_{ST(f)} = 2F_{ST(n)} - 2F_{ST(n)}F_{ST(f)} \]

\[ F_{ST(f)} + 5F_{ST(n)}F_{ST(f)} = 6F_{ST(n)} \]

\[ F_{ST(f)} = \frac{6F_{ST(n)}}{1 + 5F_{ST(n)}} \]  \hspace{1cm} (Equation in Yu et al. 2010)  \hspace{1cm} (5)

Under the assumption that glochidia flow is limited among populations, sperm gene flow is greater than glochidia gene flow when \( F_{ST(f)} > \frac{6F_{ST(n)}}{1 + 5F_{ST(n)}} \). Sperm gene flow is less than glochidia gene flow when \( F_{ST(f)} < \frac{6F_{ST(n)}}{1 + 5F_{ST(n)}} \). To test the null hypothesis that sperm flow equals glochidia flow, the dataset of nuclear loci were bootstrapped to generate a 95% confidence interval (CI) of \( F_{ST(n)} \). The upper and lower limits of \( F_{ST(n)} \) are applied to Equation 5 to create a 95% CI of expected \( F_{ST(f)} \) for the mitochondrial marker. The observed \( F_{ST(f)} \) estimated from maternally inherited marker will be compared to the expected \( F_{ST(f)} \) estimated from nuclear data. If the observed \( F_{ST(f)} \) is outside the 95% CI range of the expected \( F_{ST(f)} \), null hypothesis is rejected, sperm flow is significantly different from glochidia flow. If the observed \( F_{ST(f)} \) is within the 95% CI range of the expected \( F_{ST(f)} \), the null hypothesis cannot be rejected, and sperm flow is similar to glochidia flow.

**Correction with unbiased \( F'_{ST} \) estimator**

The estimates of gene flow can be further improved by applying unbiased \( F_{ST} \) estimators. For multiallelic loci, \( F_{ST} \) (Weir & Cockerham 1984) and its analogs are known to underestimate genetic differentiation among populations because of their dependency on within-population heterozygosity (Jost 2008; Meirmans & Hedrick 2011). Heterozygosity doesn’t scale linearly with increasing diversity and thus may not be suitable for estimating population diversity (Jost 2008). For better estimation of inter-population differentiation,
Meirmans and Hedrick (2011) developed unbiased estimators ($F'_ST$) to adjust $F_{ST}$ by dividing it by a possible maximum $F_{ST}$ calculated from the observed allele frequencies.

After applying the adjusted estimate of $F_{ST}$, for nuclear markers,

$$N_{m_g} \approx \frac{(1 - F'_{ST(n)})}{4F_{ST(n)}}$$

(Meirmans and Hedrick 2011)

For maternally inherited markers in a dioecious species,

$$N_{m_f} \approx \frac{(1 - F'_{ST(f)})}{F_{ST(f)}}$$

Substituting the variables into Equation 2,

$$\text{Sperm flow/glochidia flow} = \frac{[(1 - F'_{ST(n)})/ F_{ST(n)} - 4(1 - F'_{ST(f)}) / F_{ST(f)}]}{[2(1 - F'_{ST(f)}) / F_{ST(f)}]}$$

(6)

**Sampling Methods and Molecular Data Generation**

We compared the ratio of sperm gene flow and glochidia gene flow in three mussel species *Leptodea leptodon*, *Leptodea fragilis*, and *Lampsilis abrupta*. 117 *Le. leptodon* and 100 *Le. fragilis* samples were collected from the same five sampling sites in the Bourbeuse, Meramec, and Gasconade rivers in Missouri (Table 1, Fig. 1). These sampling locations comprise the entirety of the current range for the endangered *Le. leptodon*. About 870 base pairs of the maternally inherited ND1 gene were generated and ten microsatellite loci were amplified for each species to estimate nuclear gene variation. A total of 121 *La. abrupta* individuals were collected from four drainages: Gasconade, Meramec, and Osage rivers in Missouri, and the Tennessee River in Tennessee (Table 2, Fig. 2). The same region of the ND1 gene (~850 bps) was amplified for 116 *La. abrupta* samples. Ten different microsatellite markers were optimized for 113 *La. abrupta* samples from five populations.
using primers from Eackles and King (2002). DNA sample collection methods, molecular data generation and analyses were described in Chong et al. (2016).

We performed two types of analyses to estimate sperm and glochidia gene flow. First, we grouped individuals based on their geographical locations according to their drainages. *Le. leptodon* and *Le. fragilis* were grouped into three populations: Gasconade, Meramec, and Bourbeuse rivers. *La. abrupta* were grouped into four populations: Gasconade, Meramec, Osage, and Tennessee rivers. In the second type of analysis we grouped individuals based on the clustering pattern suggested by the STRUCTURE analysis, described in Chapter III. The STRUCTURE analysis indicated weak population structure ($K = 2$) and separated the Gasconade *Le. leptodon* individuals from the Meramec and Bourbeuse individuals. Based on this information, *Le. leptodon* samples were grouped into two populations, Gasconade (sites A and B) and Meramec/Bourbeuse (sites C, D, and E) populations. *Leptodea fragilis* samples were grouped into an upper Gasconade (site A) population and another population that consisted of the rest of the individuals (from sites B, C, D, and E). For *La. abrupta*, the STRUCTURE analysis indicated four distinct populations ($K = 4$) corresponding to the river drainages. GenAlEx v.6.502 (Peakall & Smouse 2006; Peakall & Smouse 2012) was used to estimate pairwise $F_{ST(n)}$ for nuclear gene marker (microsatellites data) among populations of each species. SPAGeDi v1.5 (Hardy & Vekemans 2002) was used to jackknife across the loci and produce 95% CI of $F_{ST(n)}$ (mean $\pm 1.96 \times$ s.e.). The upper and lower limits of $F_{ST(n)}$ were imported into Equation 5 to calculate 95% CI of expected $F_{ST(f)}$. When $F_{ST(n)}$ or $F_{ST(f)}$ values estimated to be 0 or negative values, we assumed $F_{ST}$ to be a very small value ($\approx 0.0001$) so that the equations could be solved. The observed $F_{ST(f)}$ was estimated with Arlequin v.3.5 (Excoffier & Lischer 2010) from maternally inherited marker (mtDNA data), and then
compared with the expected $F_{ST(f)}$ estimated from microsatellite data to see if sperm flow is significantly different from glochidia flow. The ratio of sperm and glochidia flow was estimated for populations with observed $F_{ST(f)}$ beyond the estimated 95% CI range of expected $F_{ST(f)}$. The unbiased $F_{ST}$ estimator, $F'_{ST}$ (Meirmans & Hedrick 2011), was calculated by dividing the estimated $F_{ST}$ by maximum possible $F_{ST}$, which was estimated with GenAlEx v.6.502. The $F_{ST}$ and $F'_{ST}$ from nuclear and mtDNA markers were then applied to Equation 6 to calculate sperm and glochidia flow ratio adjusted with $F_{ST}$ correction.

**Results**

The F-statistics for mtDNA $F_{ST(f)}$, nuclear $F_{ST(n)}$, and the calculated sperm flow/glochidia flow ratios for all three species were presented in Table 3. The null hypothesis of sperm flow equals glochidia flow was tested for pairs of populations grouped either by drainages or by STRUCTURE clusters. When the observed $F_{ST(f)}$ between a pair of populations was within the 95% CI range of expected $F_{ST(f)}$ estimated from $F_{ST(n)}$, sperm gene flow was equal to glochidia gene flow, and thus it was not necessary to calculate the sperm flow/glochidia flow ratio. Sperm flow was significantly different from glochidia flow in 6 out of 14 comparisons made in this study (Table 3). For these 6 pairs of populations, $F_{ST(n)}$, $F'_{ST(n)}$, $F_{ST(f)}$, and $F'_{ST(f)}$ were estimated and applied to Equation 6 to calculate the ratio of sperm flow/glochidia flow. The $F_{ST(f)}$ and $F'_{ST(f)}$ were very similar due to low allelic richness observed in the mtDNA data, thus only $F_{ST(f)}$ was presented in Table 3.

When grouping based on river drainages, sperm flow/glochidia flow ratios for *Le. leptodon* were 18.1 and 21.2 between Gasconade and Bourbeuse populations, and between
Gasconade and Meramec populations respectively. The same pattern was observed when the
*Le. leptodon* individuals were grouped based on the STRUCTURE clusters. Sperm flow
greatly exceeding glochidia flow between Gasconade and Meramec/Bourbeuse populations,
with a ratio of 19.2. Sperm gene flow and glochidia gene flow were not significantly
different from each other for *Le. fragilis* when individuals were grouped according to
drainages. However, when clustered based on the STRUCTURE analysis, sperm gene flow
was less than glochidia flow between upper Gasconade and the rest of the *Le. fragilis*
populations. The estimated sperm and glochidia flow ratio was -0.84. Although a ratio should
not be a negative value, this is one of the limitations in our current approach. When \((1 – F'_\text{ST(n)})/F_{\text{ST(n)}}\) is less than \(4(1 – F'_\text{ST(f)}) / F_{\text{ST(f)}}\), the estimated sperm/glochidia ratio yields a
negative result.

In *La. abrupta*, we recorded the largest sperm gene flow/glochidia gene flow ratio
between the Gasconade and Osage populations (~ 510) due to low nuclear gene
differentiation \(F_{\text{ST(n)}}\). Between the Gasconade/Osage and the Meramec rivers, and between
the Meramec/Osage and the Tennessee rivers, levels of glochidia gene flow were similar to
sperm gene flow in *La. abrupta*. Between the Gasconade and Tennessee population, sperm
gene flow once again was higher than the glochidia gene flow, although the ratio of
sperm/glochidia flow was comparatively low (~ 2.7).

**Discussion**

*Estimating the ratio of sperm gene flow and glochidia gene flow using empirical data*

We applied the Ennos (1994) and McCauley (1997) approaches to estimate sperm
flow in three freshwater mussel species *Le. leptodon, Le. fragilis*, and *La. abrupta.*
Contradicting previous widely held assumptions for freshwater mussels, we found that in the majority of cases, sperm gene flow contributes as much or more to the maintenance of connectivity among mussel populations than glochidia gene flow. This result is similar to those observed in plant species, in which pollen flow has been generally been shown to be greater than seed flow. Perhaps more surprisingly, the magnitude of the difference between sperm and glochidia gene flow can sometimes be greater in mussels than the difference between pollen and seed gene flow in plants. In plants, the ratio of pollen gene flow to seed gene flow ranges between 4 - 196 (Table 2 in Ennos 1994), whereas in our data, the estimated ratio of sperm gene flow and glochidia gene flow ranged between -0.86 to 500+, with only one instance in which zygote-mediated gene flow exceeded gamete-mediated gene flow. This study has demonstrated that sperm gene flow, or the contribution of males to maintaining connectivity between populations of freshwater mussels, has been consistently under-estimated.

Despite sharing the same host fish, *Le. leptodon* and *Le. fragilis* exhibit different patterns of gene flow between the Gasconade and Meramec populations when grouped according to the STRUCTURE results. The sperm gene flow was significantly higher than glochidia gene flow in *Le. leptodon*, but the completely opposite pattern was observed for *Le. fragilis*. Based on the differences in dispersal range of spermatozeugmata and glochidia, it is reasonable to assume that sperm gene flow will be dominant in geographically close populations while glochidia may be more prominent in connecting more isolated populations. This also seems to be true between Gasconade and Osage populations of *La. abrupta*. However, our results in general showed that the ratio of sperm flow and glochidia flow was not always based on the distance between populations. For example, Bourbeuse and
Meramec populations of *Le. Leptodon* and *Le. fragilis* are adjacent to one another (~ 25 river miles apart). In theory, sperm flow should exceed glochidia flow, but our results showed that sperm and glochidia gene flow were equal.

In *La. abrupta*, the gene flow pattern was more complicated among populations within Missouri (Gasconade, Meramec, and Osage rivers) and between the Missouri population and the Tennessee population. Fishes in the genus *Micropterus* are known to be the host for *La. abrupta* (Dodd *et al.* 2005), and are documented to travel distances of a few kilometers (Wilde 2003). The Gasconade population is geographically closest to the Osage population. $F_{ST(n)}$ estimated between the two populations was close to 0, indicating that individuals from the two locations can be considered to be a panmictic population. It appeared that sperm gene flow is the major mechanism of gene flow between *La. abrupta* in the Gasconade and Osage rivers, resulting in the highest sperm/glochidia flow ratio estimated. However, sperm flow is also greater than glochidia flow between Gasconade and Tennessee populations. These two sampling sites are so far apart that it is unlikely that sperm could to one population to another within the same generation. This observation can perhaps be explained as representing the overall male contribution to the gene flow among intermediate populations between the Gasconade and Tennessee rivers having exceeded the contribution of glochidia to overall gene flow. One of the unexpected discoveries observed when we applied the equations with real-world data, was that gene flow between adjacent populations (Meramec and Bourbeuse in *Le. leptodon* and *Le. fragilis*) was not always dominated by sperm gene flow. We observed higher sperm gene flow among populations of both *Leptodea* species that are separated by greater distances. This phenomenon might be due to the movement behavior of the host fish. Both mussel species share the same and only host
fish species, freshwater drum (*A. grunniens*). Although freshwater drum is a fish known for its high mobility and capable of traveling great distances (Funk 1955), fish infested with glochidia may not always travel very far and instead remain in the same region of the river where they encountered the gravid female mussel reducing the glochidia dispersal distance. The second possible explanation for this phenomenon is that the approach we employed underestimates sperm gene flow. Our current approach assumes that glochidia gene flow is limited among populations and thus greatly underestimates the sperm gene flow among populations. Taking this into account, sperm gene flow may be higher than glochidia flow among all populations.

**Conservation Implication**

Many aquatic animals have two types of gene flow: gamete-mediated (usually referred to as sperm dispersal) and zygote-mediated gene flow (movement of progeny). Freshwater mussels are interesting organisms for this study because sperm from some species like *L. cardium* can remain viable 16.2 km downstream (Ferguson *et al.* 2013). Their progeny dispersal is associated with the movement of the parasitic glochidia, moving either upstream or downstream depending on the host fish mobility. Our research has provided an estimation of sperm gene flow in mussel and aquatic species where paternally inherited markers are generally lacking. Our approach can be modified for hermaphroditic mussels (e.g. *Margaritifera falcata*) and adjusted using unbiased $F_{ST}$ estimators.

Our results indicated that sperm gene flow contributes significantly toward maintaining genetic connectivity among populations and has previously been underestimated. Sperm gene flow significantly exceeded glochidia gene flow between populations in *Le.*
Leptodon and La abrupta. In one instance, sperm gene flow was 500 times greater than glochidial flow between La. abrupta from Gasconade and Osage rivers. This ratio might be underestimated if the assumption of limited glochidial flow among populations was violated, which means sperm gene flow can be higher between these populations than the current estimation. Male contribution to the overall gene flow among populations should be re-evaluated in freshwater mussels and aquatic organisms as having genetically diverse males are greatly beneficial to maintain diversity among populations.

Understanding the importance of male-mediated gene flow has important implications for conservation management. Our research fills in a gap by addressing the importance of sperm gene flow in population structure of freshwater mussels. Our study was focused only on a relatively small region in Missouri and Tennessee. The approach should be tested on freshwater mussels with a larger distribution range and more diverse population structure. This approach may be underestimating sperm flow, but this study should serve the purpose of initiating more studies focusing on sperm gene flow of freshwater mussels. Future research should focus on understanding sperm longevity and survival in freshwater mussels.

Conclusions

We developed a population genetic approach to estimate sperm gene flow in freshwater mussels and aquatic dioecious organisms using nuclear and maternally inherited markers. We improved the approaches developed by Ennos (1994) and McCauley (1997) by adjusting the equations with unbiased $F_{ST}$ estimator ($F'_{ST}$). Our preliminary results have shown that sperm gene flow is predominant in maintaining connectivity among populations with compared to glochidial gene flow. Our study provides an alternative approach to
estimating sperm dispersal in aquatic invertebrates when paternally inherited markers are not available.

**Acknowledgements**

This project was funded by the Missouri Department of Conservation and AMEREN UE. We would like to thank John Nason for his thoughtful advice and contribution to the analysis. We are grateful to Steven McMurray, Andy Roberts, Bryan Simmons, Scott Faiman, Josh Hundley, and Morgan Sexton for their assistance in sample and molecular data collection. We also thank Jeanne Serb for reviewing this manuscript.

**Table 4-1.** Sampling locations and sample size for *Le. leptodon* and *Le. fragilis* populations.

<table>
<thead>
<tr>
<th>Population</th>
<th>Drainage</th>
<th>State</th>
<th><em>Le. leptodon</em></th>
<th><em>Le. fragilis</em></th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Gasconade River</td>
<td>MO</td>
<td>26</td>
<td>11</td>
</tr>
<tr>
<td>B</td>
<td>Gasconade River</td>
<td>MO</td>
<td>36</td>
<td>17</td>
</tr>
<tr>
<td>C</td>
<td>Bourbeuse River</td>
<td>MO</td>
<td>22</td>
<td>28</td>
</tr>
<tr>
<td>D</td>
<td>Meramec River</td>
<td>MO</td>
<td>22</td>
<td>23</td>
</tr>
<tr>
<td>E</td>
<td>Meramec River</td>
<td>MO</td>
<td>17</td>
<td>25</td>
</tr>
<tr>
<td>Total:</td>
<td></td>
<td></td>
<td>123</td>
<td>104</td>
</tr>
</tbody>
</table>

**Table 4-2.** *Lampsilis abrupta* samples collected from each sampling location.

<table>
<thead>
<tr>
<th>Population ID</th>
<th>Drainage</th>
<th>State</th>
<th>Samples Collected</th>
</tr>
</thead>
<tbody>
<tr>
<td>AGC</td>
<td>Gasconade River</td>
<td>MO</td>
<td>8</td>
</tr>
<tr>
<td>AGF</td>
<td>Gasconade River</td>
<td>MO</td>
<td>26</td>
</tr>
<tr>
<td>AMX</td>
<td>Meramec River</td>
<td>MO</td>
<td>25</td>
</tr>
<tr>
<td>AOH</td>
<td>Osage River</td>
<td>MO</td>
<td>22</td>
</tr>
<tr>
<td>ATR</td>
<td>Tennessee River</td>
<td>TN</td>
<td>40</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td>121</td>
</tr>
</tbody>
</table>
Table 4-3. Individuals of *Le. leptodon*, *Le. fragilis*, and *La. abrupta* were grouped into populations according to their sampling locations or STRUCTURE results. For each pair of populations within species, $F_{ST(n)}$, 95% CI for $F_{ST(n)}$, $F_{ST(n)}$, 95% CI for expected $F_{ST(f)}$, max $F_{ST(n)}$ and $F'_{ST(n)}$ were estimated from microsatellite and mtDNA data. $F'_{ST(n)}$ were not included in this table because they were very similar to $F_{ST(f)}$. If observed $F_{ST(f)}$ is within the 95% CI of expected $F_{ST(f)}$, sperm flow is similar to glochidia flow and the ratio was not calculated.

<table>
<thead>
<tr>
<th>Species</th>
<th>Populations</th>
<th>$F_{ST(n)}$</th>
<th>95% CI for $F_{ST(n)}$</th>
<th>95% CI for Expected $F_{ST(f)}$</th>
<th>$F_{ST(f)}$</th>
<th>Observed $F_{ST(f)}$ outside the range of expected $F_{ST(f)}$</th>
<th>max $F_{ST(n)}$</th>
<th>$F'_{ST(n)}$</th>
<th>Sperm flow /Glochidia flow</th>
<th>Sperm flow &gt; Glochidia flow</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grouped by Sample sites</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Le. leptodon</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gasconade</td>
<td>Bourbeuse</td>
<td>0.012</td>
<td>0.005 - 0.021</td>
<td>0.028 - 0.113</td>
<td>0.348</td>
<td>Yes</td>
<td>0.125</td>
<td>0.096</td>
<td>18.10</td>
<td>Yes</td>
</tr>
<tr>
<td>Gasconade</td>
<td>Meramec</td>
<td>0.007</td>
<td>0.002 - 0.012</td>
<td>0.012 - 0.065</td>
<td>0.256</td>
<td>Yes</td>
<td>0.126</td>
<td>0.056</td>
<td>21.20</td>
<td>Yes</td>
</tr>
<tr>
<td>Meramec</td>
<td>Bourbeuse</td>
<td>0.002</td>
<td>0 - 0.007</td>
<td>0 - 0.043</td>
<td>0.0001</td>
<td>No</td>
<td>0.126</td>
<td>0.016</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><em>Le. fragilis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gasconade</td>
<td>Bourbeuse</td>
<td>0.005</td>
<td>0 - 0.012</td>
<td>0 - 0.068</td>
<td>0.061</td>
<td>No</td>
<td>0.228</td>
<td>0.023</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Gasconade</td>
<td>Meramec</td>
<td>0.013</td>
<td>0.004 - 0.018</td>
<td>0.026 - 0.100</td>
<td>0.055</td>
<td>No</td>
<td>0.236</td>
<td>0.054</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td>Meramec</td>
<td>Bourbeuse</td>
<td>0.0004</td>
<td>0 - 0.005</td>
<td>0 - 0.029</td>
<td>0.0001</td>
<td>No</td>
<td>0.231</td>
<td>0.002</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Grouped by STRUCTURE results</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Le. leptodon</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gasconade</td>
<td>Meramec + Bourbeuse</td>
<td>0.008</td>
<td>0.003 - 0.013</td>
<td>0.017 - 0.076</td>
<td>0.266</td>
<td>Yes</td>
<td>0.125</td>
<td>0.064</td>
<td>19.20</td>
<td>Yes</td>
</tr>
<tr>
<td><em>Le. fragilis</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Upper</td>
<td>Gasconade</td>
<td>0.038</td>
<td>0.024 - 0.063</td>
<td>0.127 - 0.286</td>
<td>0.094</td>
<td>Yes</td>
<td>0.258</td>
<td>0.147</td>
<td>-0.84</td>
<td>No</td>
</tr>
<tr>
<td>Lower</td>
<td>Gasconade</td>
<td>0.0001</td>
<td>0 - 0.004</td>
<td>0 - 0.023</td>
<td>0.093</td>
<td>Yes</td>
<td>0.17</td>
<td>0.001</td>
<td>510.17</td>
<td>Yes</td>
</tr>
<tr>
<td>+ Meramec</td>
<td>Lower</td>
<td>0.038</td>
<td>0.008 - 0.068</td>
<td>0.044 - 0.306</td>
<td>0.317</td>
<td>Yes</td>
<td>0.159</td>
<td>0.237</td>
<td>2.66</td>
<td>Yes</td>
</tr>
<tr>
<td>+ Meramec</td>
<td>Tennessee</td>
<td>0.012</td>
<td>0 - 0.034</td>
<td>0 - 0.175</td>
<td>0.024</td>
<td>No</td>
<td>0.185</td>
<td>0.066</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>Osage</td>
<td>0.041</td>
<td>0.018 - 0.067</td>
<td>0.099 - 0.300</td>
<td>0.167</td>
<td>No</td>
<td>0.172</td>
<td>0.240</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>Tennessee</td>
<td>0.031</td>
<td>0.004 - 0.060</td>
<td>0.025 - 0.278</td>
<td>0.149</td>
<td>No</td>
<td>0.159</td>
<td>0.197</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td><strong>Grouped by STRUCTURE results</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>La. abrupta</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Gasconade</td>
<td>Meramec</td>
<td>0.014</td>
<td>0.004 - 0.027</td>
<td>0.021 - 0.143</td>
<td>0.059</td>
<td>No</td>
<td>0.183</td>
<td>0.075</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>Osage</td>
<td>0.0001</td>
<td>0 - 0.004</td>
<td>0 - 0.023</td>
<td>0.093</td>
<td>Yes</td>
<td>0.17</td>
<td>0.001</td>
<td>510.17</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Tennessee</td>
<td>0.038</td>
<td>0.008 - 0.068</td>
<td>0.044 - 0.306</td>
<td>0.317</td>
<td>Yes</td>
<td>0.159</td>
<td>0.237</td>
<td>2.66</td>
<td>Yes</td>
</tr>
<tr>
<td></td>
<td>Osage</td>
<td>0.012</td>
<td>0 - 0.034</td>
<td>0 - 0.175</td>
<td>0.024</td>
<td>No</td>
<td>0.185</td>
<td>0.066</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>Tennessee</td>
<td>0.041</td>
<td>0.018 - 0.067</td>
<td>0.099 - 0.300</td>
<td>0.167</td>
<td>No</td>
<td>0.172</td>
<td>0.240</td>
<td>--</td>
<td>--</td>
</tr>
<tr>
<td></td>
<td>Osage</td>
<td>0.031</td>
<td>0.004 - 0.060</td>
<td>0.025 - 0.278</td>
<td>0.149</td>
<td>No</td>
<td>0.159</td>
<td>0.197</td>
<td>--</td>
<td>--</td>
</tr>
</tbody>
</table>
Figure 4-1. Sampling map for *Le. leptodon* and *Le. fragilis* in Missouri. Sites A and B are from Gasconade River. Site C is in Bourbeuse River. Sites D and E are in Meramec River.

Figure 4-2. Map showing sampling locations of all *La. abrupta* samples. *La. abrupta* were collected from Osage River (AOH), Gasconade River (AGC, AGF), Meramec River (AMX) in Missouri, and Tennessee River (ATR) from Tennessee.
REFERENCES


CHAPTER V
GENERAL CONCLUSIONS

Dispersal ability and reproductive success can determine the potential of whether a threatened or endangered species is capable to recover from bottlenecks or catastrophic population decline. In this dissertation, I have demonstrated how molecular genetic techniques can be effectively applied in wildlife management, i.e., by profiling individuals, examining genetic diversity at individual, population, and species levels, determining population structure and migration among populations, and predicting the evolutionary trajectory of populations under selection, etc. The objective of this dissertation was to understand how population connectivity (i.e. gene flow) influences species viability and adaptability to environmental changes in freshwater mussels. I first assessed the genetic variation of sampled individuals to delineate the number of evolutionary lineages or entities present within our study area. Then I examined the genetic relatedness among individuals to identify population and/or management units. Individuals with higher probability of mating with one another (i.e. higher genetic similarity and relatedness) than with individuals from other groups are considered to be a population. Understanding the genetic connectivity among populations allows wildlife biologists to identify whether populations are fragmented and isolated, whether populations recently underwent a bottleneck event, and whether populations are currently suffering from inbreeding and genetic drift. I then partitioned the gene flow into two types: gamete-mediated gene flow contributed through sperm dispersal, and zygote-mediated gene flow contributed through glochidia dispersal via movement of host fish. This is the first study attempts to measure the gamete-mediated gene flow (male gene flow) in freshwater mollusk using both maternally inherited and nuclear genetic marker.
Delimiting evolutionary entities (e.g. species, population units) is an important first step in wildlife conservation and management. In Chapter II, I focused on resolving taxonomic confusion in freshwater mussel genus *Cyprogenia*. Currently two *Cyprogenia* species are recognized within this genus, with *C. aberti* reside in west of the Mississippi river while *C. stegaria* (a federally endangered species) occur east of the Mississippi river. As reported in previous molecular studies, two deeply divergent mtDNA lineages found to co-exist in most of our sampling regions did not concur with the current species designation of *Cyprogenia*. With the addition of nuclear microsatellite data of *C. aberti* not included in previous studies, I found evidence of mito-nuclear discordance in *Cyprogenia*. The microsatellite data supported three allopatric clades corresponding to the major hydrologic drainages. Three evolutionarily significant units were designated for *Cyprogenia*: Ozark, Ouachita, and Ohio. Our data suggested that *C. stegaria* forms its own cluster in the Ohio River Basin. West of the Mississippi river, Ouachita *C. aberti* are genetically different from *C. aberti* from the Ozark region. My study also showed that the mtDNA gene in *Cyprogenia* is highly correlated with conglutinate pigmentation of conglutinate and may be subject to frequency-dependent selection imposed by host fish.

Freshwater mussels are among the most threatened fauna in North America. To better understand factors causing the decline of mussel species, in Chapter III I compared the gene flow patterns and population structures of an endangered mussel species *Leptodea leptodon* and its common congener *Le. fragilis* with their shared host fish, the freshwater drum *Aplodinotus grunniens*. No genetic studies have been conducted on *Le. leptodon* and *A. grunniens* prior to this research. The purpose of this study was to determine whether host dispersal has a similar impact on the genetic structures of both an endangered and a common
mussel species. Theoretically both mussel species should share the same population structures as the host fish. However, my results indicated that the genetic structures of all three species were not congruent. *Leptodea leptodon* individuals were grouped into Gasconade and Meramec/Bourbeuse populations, with Missouri and Mississippi rivers forming a barrier of gene flow between the two populations. In *Le. fragilis*, upper Gasconade population was separated from the rest of populations (lower Gasconade + Meramec + Bourbeuse), perhaps due to lack of mussel habitat or limited host fish movement between upper and lower Gasconade River. Freshwater drum from all three rivers can be considered as a panmictic population. Despite the imperiled status of *Le. leptodon*, allelic richness was similar in *Le. leptodon* and the widespread *Le. fragilis*. Gene flow between the extant populations of *Le. leptodon* is also substantial and comparable to the *Le. fragilis* populations. This study has provided insight into conservation management by ruling out barriers to gene flow or lack of available host fish as potential threats to the federally endangered *Le. leptodon*. Habitat specificity and recent bottleneck may result in the differences in population structures of the two mussel species.

In Chapter IV, I partitioned gene flow among mussel populations into two types, gamete-mediated (sperm dispersal) and zygote-mediated gene flow (glochidia dispersal) to further our understanding on the biological factors that influenced the population structures of freshwater mollusk. Sperm gene flow is a research topic in aquatic freshwater mollusks that is currently under-studied due to the general lack of paternally inherited genetic marker. I developed a novel approach to indirectly estimate sperm flow using maternally inherited markers (i.e., mtDNA marker) and nuclear markers. I estimated the sperm vs glochidia flow ratio for three mussel species (*Le. leptodon*, *Le. fragilis*, and *La. abrupta*) and discovered that
the contribution of sperm gene flow to population connectivity has been underestimated. Sperm gene flow was significantly higher in *Le. leptodon* than *Le. fragilis* between Gasconade and Meramec populations. However, glochidia gene flow was higher between the adjacent Meramec and Bourbeuse populations in both mussel species. In *La. abrupta*, sperm gene flow was evidently higher between Gasconade and Osage/Tennessee populations. Sperm and glochidia gene flow was relatively similar among Meramec, Bourbeuse, and Tennessee populations. This approach can be applied to other aquatic organisms that produce gametes with high dispersal distances and perform external fertilization. I think my study has an important implication by filling in the gap in knowledge for freshwater mussel research. More importantly, this study serves a purpose of initiating additional research focusing on the sperm gene flow of freshwater mussels, a currently under-studied field.