2013

A Population Genetic Study of the Fish *Rhinichthys cataractae* in Northeast Iowa Using Microsatellite Genotype Data

Haley Ann Rinehart
University of Northern Iowa

A POPULATION GENETIC STUDY OF THE FISH *RHINICHTHYS CATARACTAE* IN NORTHEAST IOWA USING MICROSATELLITE GENOTYPE DATA

A Thesis Submitted

in Partial Fulfillment

of the Requirements for the Designation

University Honors

Hayley Ann Rinehart

University of Northern Iowa

May 2013

This Study by: Hayley Ann Rinehart

Entitled: A Population Genetic Study of the Fish *Rhinichthys cataractae* in Northeast Iowa Using Microsatellite Genotype Data

has been approved as meeting the thesis or project requirement for the Designation University Honors.

5/1/13
Date

Dr. Peter Berendzen, Honors Thesis Advisor, Department of Biology

5/10/13
Date

Dr. Jessica Moon, Director, University Honors Program

ACKNOWLEGEMENTS

I would like to first and foremost thank my wonderful advisor, Pete Berendzen, for his optimism, endless patience, encouragement, and expertise. His exemplary knowledge and passion for this field of research is what motivated me to become involved in the project. Next, I would like to express my undying gratitude to Brigid Wieman, whose guidance in the laboratory has helped me to understand and apply the techniques described in this thesis. Thirdly, I would like to extend my thanks to all members of the UNI Fish Lab, especially Corinthia Black, Megan Merner, Erica Scullin, Courtney Calhoun, and Nikki Miller. You are all intelligent, talented, and hardworking individuals and I feel extremely lucky to share in your love of science every day. Next, I would like to thank those who took the time to help me in the revision process, such as Dr. Jessica Moon, Dr. James Demastes, and Dr. Carl Thurman. Lastly, I could not have completed this research without the love and support of my friends and family. To my parents Mick and Karen, sisters Katia and Jenna, friends Nicole Peyton, Mallory Bazan, Becca Saladin, Ashley Richardson, and my boyfriend, Josh Faley, I cannot thank you enough.

INTRODUCTION

The assemblage of fishes found in northeast Iowa is as distinguishable as the region's geographical history. The last glacial maximum ended roughly 15,000 years ago (YA) at the close of the Pleistocene epoch. The topography of northeast Iowa was molded by the unique pattern of glacial advances during this time. The region was covered by pre-Illinoian glaciers, but was completely bypassed by both the Illinoian and Wisconsinian ice sheets (Figure 1). The distinct history of this region of interest is evident in its steep bluffs and deep valleys otherwise absent in the rest of Iowa (Hobbs, 1999). Likewise, streams in northeast Iowa have distinctive characteristics when compared to those found elsewhere in the state, such as high gradients and cool water (Rowe et al., 2004).

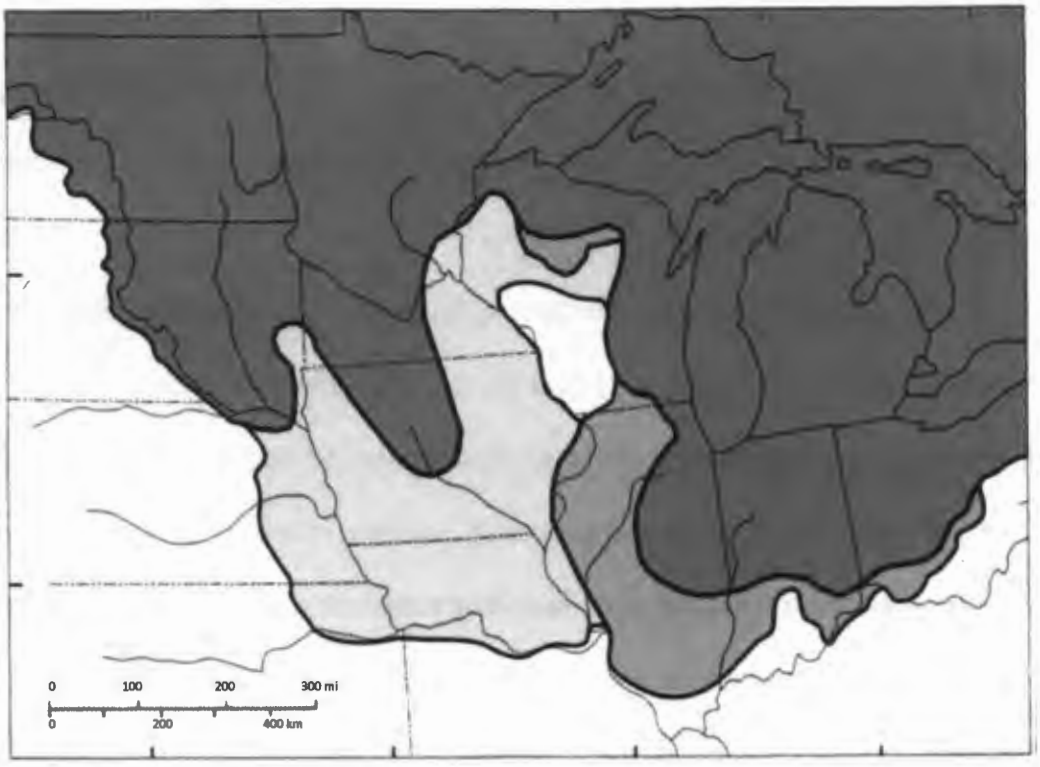

Figure 1. Map of estimated glacial cover by the pre-Illinoian (light grey), Illinoian and Wisconsinan ice sheets (dark grey) (Berendzen et al., 2010).

The qualities of the landscape in northeast Iowa have greatly influenced the ichthyofauna of the region. Previous studies have shown that repeated glacial advances in North America have altered the diversity and distribution of freshwater fishes (Berendzen et al., 2010). Ancient events such as glaciations are known to contribute to the currently observed assemblage and genetic variation of fish populations due to the marked topographical changes that occurred during glacial advance and retreat. The range shift caused by glacial advances may result in population expansion, reduction, and/or fragmentation into several geographically isolated refugia, which experience restricted gene flow (Berendzen et al., 2003). Glacial retreat has the potential to stimulate re-colonization of previously glaciated areas, allowing for the colonization of new ecological niches and the establishment of genetic populations (Burkhead, 2012). A mass of literature has confirmed that glacial patterns have a profound effect on aquatic communities, contemporary distributions of fishes, and the genetic variation within these distributions (Hewitt, 2000; Mayden, 1985, 1988; Strange & Burr, 1997; Near et al., 2001; Ray et al., 2006; Berendzen et al., 2008).

In addition to historical climatic and geological events, the Iowa landscape has been extensively transformed for agriculture since the mid-19th century. As a consequence, the landscape surrounding streams has been severely altered at spatial scales ranging from in-stream, to surrounding bank, to the entire drainage. According to a report published in 2012 by the Environmental Working Group, the water quality of Iowa streams is chronically poor and worsening due to farm pollution (Cox & Hug, 2012). This, in conjunction with other physical habitat variables, was found to have strong correlations with the assemblage of fish, overall species richness, and health of Iowa wadeable streams (Rowe et al., 2009b). Furthermore, the landscape surrounding these streams was shown to directly influence physical habitat variables,

and by extension, the health and biotic integrity of the fish species residing there (Rowe et al., 2009a).

Unfortunately, many northeast Iowa fish species are on the decline as a result of habitat destruction from human activities (Iowa DNR, 2007). A recently published study revealed that extinction rates of freshwater fishes have been on the rise over the past century and this trend is predicted to continue (Burkhead, 2012). Hence, the current and future state of aquatic biodiversity is extremely vulnerable. Researchers confirmed that "a fundamental concern among biologists is that contemporary rates of extinction due to human activities are orders of magnitude greater than background rates evidenced in the fossil record" (Burkhead, 2012, p. 798).

Therefore, the pervasiveness of agriculture in northeast Iowa provides an impetus for the conservation of one of its freshwater fishes, *Rhinichthys cataractae*. Commonly referred to as longnose dace, this species is listed as a Species of Greatest Conservation Need in the Iowa Wildlife Action Plan (Iowa DNR, 2007). This plan was designed to serve as a comprehensive blueprint for the conservation of all threatened wildlife species in the state.

Figure 2. The minnow *Rhinichthys cataractae* (North American Native Fishes.org, 2010).

Rhinichthys cataractae (Figure 2) is a non-game, native fish species. It is a member of the Cyprinidae family, which includes the minnows and carps. Reaching an adult average length of six centimeters, this fish has a maximum life span of five years. They have a long, round body, triangular head, and characteristic overhanging snout (COSEWIC, 2007). Longnose dace spawn multiple times during the summer months, are generally found near the stream substrate bottom, and feed on

aquatic insect larva. *Rhinichthys cataractae* inhabits and has high fidelity to fast-moving riffles in cool water streams (Evans et al., 2012).

Rhinichthys cataractae has a wide distribution across the country and can be found in the majority of temperate habitats across North America (Girard & Angers, 2011). Similar to other species that inhabit northeast Iowa streams, its present-day distribution has been heavily influenced by Pleistocene glaciations (Girard & Angers, 2006). Corresponding to the distinct landscape and glacial history of northeast Iowa, the longnose dace is distributed almost exclusively in the northeast corner of Iowa (Figure 3). Investigation of the genetic diversity of this species will help clarify what steps need to be taken to ensure their survival and success, and may reveal the processes that led to the current distribution observed in Iowa. This study used genetic techniques to explore the diversity of longnose dace to contribute to future conservation efforts.

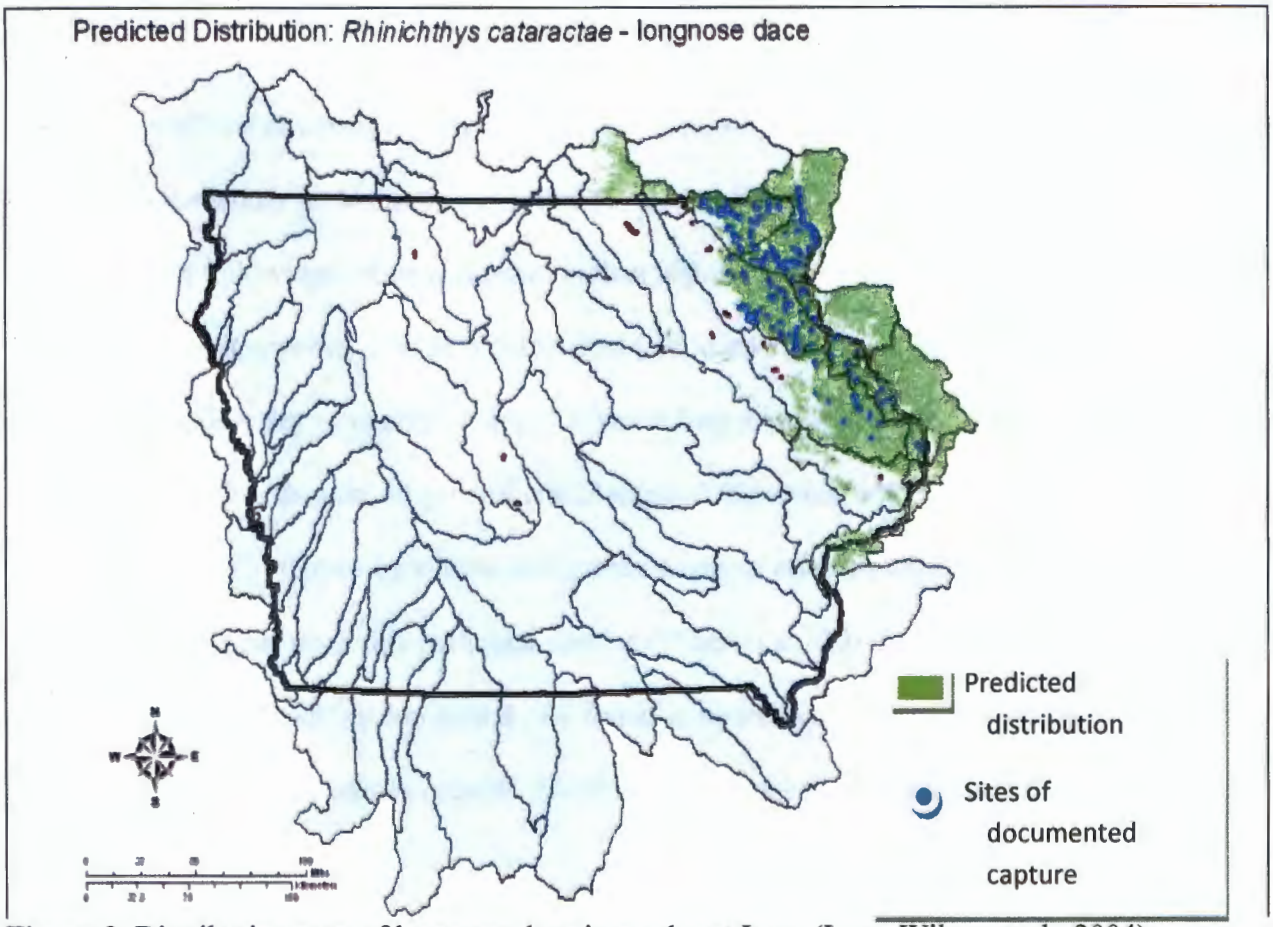

Figure 3. Distribution map of longnose dace in northeast Iowa (Loan-Wilsey et al., 2004).

LITERATURE REVIEW

Genetic structure and diversity is impacted by several intertwining factors (Frankham et al., 2005). These include, but are not limited to: neutral and non-neutral evolutionary processes, short term demographic events such as genetic drift and gene flow, and the long-term effects of mutation rate and geographic isolation. The level of genetic variation resulting from these processes can be used to assess the role of glaciation in shaping current distributions. For example, colonization events and species redistribution following glacial retreat are expected to lead to reduced genetic diversity in recently colonized areas (Berendzen et al., 2010). A study using neutral and non-neutral genetic markers confirmed a potentially threatening decline in the diversity of longnose dace populations in the Quebec province of Canada, an area subject to the sculpting and often reductive effects of both Pleistocene glacial history and human interference (Girard & Angers, 2011).

Along with the impacts of ancient phenomena, assessments of genetic diversity can uncover the affects of contemporary landscape features on natural populations. The imperiled Red Hills salamander in Alabama was found to be more heavily influenced by habitat fragmentation following human colonization than historical events, illustrating the obstacles presented by contemporary habitat modification (Apodaca et al., 2012). Contemporary influences on diversity were also observed in wood frog populations in southeast Michigan, evident even within the past 50 generations (Zellmer & Knowles, 2009).

Studies of longnose dace have verified the negative effects of contemporary human agricultural practices upon this particular species (Evans et al., 2012). Chemical contaminates such as estrogen-mimicking compounds are found in aquatic environments resulting from water treatment plants and agricultural runoff. These substances are known to interfere with biological

processes ranging from the molecular to organismal level. In *Rhinichthys cataractae*, endocrine-disrupting water contaminates were found to create female-biased sex ratios. These changes could result in the variance of spawning activities and future population viability (Evans et al., 2012).

Climate change is an imminent threat to stream fishes, thus presenting another obstacle to the persistence of this species. Lyons et al. (2010) predicted the effects of climate warming on 50 stream fish species in Wisconsin. Changes in species survival and distribution were predicted to be the most severe in streams containing cold and cool-water species, including longnose dace. The habitats of longnose dace in Wisconsin are comparable to those in northeast Iowa and can therefore be predicted to transform similarly in response to climate change—further stressing the need for effective conservation.

The fields of phylogeography and population genetics focus on studying genetic diversity and distribution, seeking to reveal the processes that lead to currently observed patterns (Avise, 2000). Studies of this kind are practical for describing patterns of migration, survival, and extirpation of fish populations in response to natural events. Although the factors affecting genetic variability within natural population is complex, reduced diversity generally increases the likelihood of extirpation of local populations and species-wide extinction (Amos & Harwood, 1998). A variety of potential threats are now better understood due to population genetics, including reduced variability, inbreeding, and bottleneck events. These threats cause deviations from the equilibrium expected of a healthy population, leading to reduced ability for species adaptation to the environment (Amos & Balmford, 2001).

In evaluating the health of longnose dace populations in northeast Iowa, population genetics is a useful tool. Vital to an analysis of this kind is the distribution and frequency of

alleles, which are variations of a particular gene or genetic marker that are present for a given population or species. A maximum of two different alleles can be present for a particular genetic marker in each individual and the identity of those alleles constitutes what is known as the individual's genotype. The genotypes of all individuals in a population contribute to the gene pool for that marker in the population. In this pool, as the number of distinct alleles possible for a genetic marker increases, the genetic health of that population also presumably increases. This relationship can be assumed because genetic variability enables populations to reproduce successfully, withstand the pressures of natural selection, and give future generations improved chances to adapt to their constantly changing environments. (Frankham et al., 2005)

Allele frequency and distribution can also provide insight into how strongly populations are related and connected to each other. Shared alleles between populations are an indicator of gene flow, while alleles unique to a particular population indicate genetic distinctiveness and isolation (Frankham et al., 2005). If the number of distinct genetic populations of longnose dace in northeast Iowa is determined, conservation planning can be molded to specifically cater to these individual groups, improving the chances for conservation success. Imperative to the protection of threatened species is the definition of populations, groups of populations, or species that can be managed as distinct units. Conservation units act as a valuable guide for the implementation of protective efforts (DeSalle & Amato, 2004). Thus, the population genetic structure of longnose dace can serve as the foundation for which conservation efforts are applied and how they are carried out.

One type of genetic marker commonly utilized in the determination of population structure is the microsatellite. Microsatellites are short fragments of DNA ubiquitous within the genome. They consist of repetitive units of nucleotides that occupy a certain location, or locus,

on the DNA strand. Microsatellite alleles differ in the number of times the nucleotide unit is repeated. A striking characteristic of microsatellites is their polymorphism, meaning they have a high mutation rate that can result in a myriad of different alleles for a particular locus. Allele frequency and distribution of these markers can therefore be used as a sensitive measurement of the genetic variability, connectivity, and health of study populations (Frankham et al., 2005).

Genetic techniques involving microsatellite markers have become heavily utilized in applied conservation since the late 1980s. Microsatellites are useful for population analyses because they are presumed to be selectively neutral, ensuring that observed variations are the result of demographic processes and not evolutionary ones (Chistiakov et al., 2006). Population structure analysis using these markers can detect lower degrees of variation within and between populations, allowing for proactive and effective definition of conservation units (Frankham et al., 2005).

The efficacy of microsatellite genotyping data for applied conservation has been observed in a host of population genetic studies. Jehle & Arntzen (2002) demonstrated that microsatellites are useful for the determination of effective population size in amphibians and the assignment of individuals to the population they best fit. A study done on the watercress darter found in the southeastern United States uncovered low genetic variability and small population size, indicating an elevated risk of extinction and the need for population-specific protection (Fluker et al., 2010). Microsatellite analysis of the Okaloosa darter in northwest Florida helped to reclassify the threatened status of some populations, while clarifying other populations that remain at a higher level of risk (Austin et al., 2011).

OBJECTIVES

A project utilizing microsatellite genotyping data for the analysis of population structure has not yet been attempted for *Rhinichthys cataractae* in northeast Iowa. The use of multiple microsatellite loci can generate sufficient data for clarifying levels of connectivity and variability of longnose dace, making it possible to aid in the conservation of this unique fish species. This study of longnose dace provides population structure data that in combination with future habitat and climate modeling will identify genetic threats to this species, allowing for the determination of conservation units and effective conservation actions.

The purpose of this study was to determine the population structure of longnose dace populations found in four river drainages of northeast Iowa: the Maquoketa, Yellow, Volga, and Upper Iowa. This was accomplished by the collection and analysis of microsatellite genotype data. A study of this kind provides original research beneficial to the conservation of a species of concern and the biodiversity of northeast Iowa streams.

MATERIALS AND METHODS

Sampling

In order to complete genetic analyses, tissue samples were collected from a total of 111 individuals of *Rhinichthys cataractae* inhabiting the four river drainages under study. A total of 13 localities (Table 1, Figure 4) were chosen based on the distribution of longnose dace reported by the Iowa Aquatic Gap Fish Atlas and Iowa DNR Northeast District Office (Figure 3). Specimens were collected using standard backpack electrofishing and seining techniques. A small tissue sample was clipped from the pectoral fins of each individual before their release back to the water. The samples were preserved in 95% ethanol for transport back the University of Northern Iowa for DNA extraction and analysis.

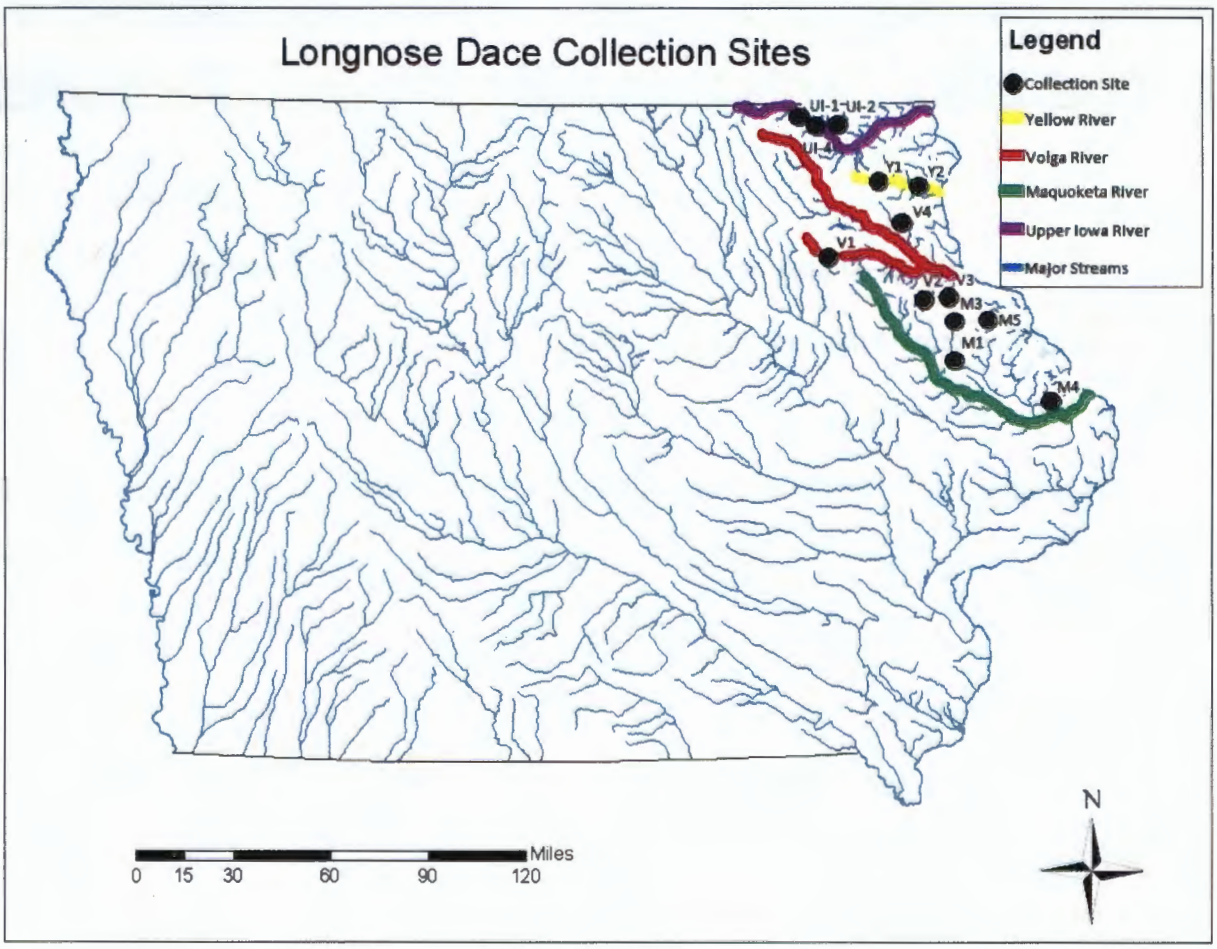

Figure 4. Sample collection sites labeled by locality ID.

Table 1. *Rhinichthys cataractae* tissue samples were collected from 13 localities across four major river drainages in northeast Iowa.

Drainage	Locality ID	# Individuals	Collection Date	Lat/Long
Volga	V1	2	Sept 25, 2011	42.81050 N/ 91.87944 W
	V2	4	Oct 2, 2011	42.61739 N/ 91.30699 W
	V4	3	June 8, 2012	42.95916 N/ 91.41955 W
Yellow	Y1	14	Oct 16, 2011	43.14113 N/ 91.57006 W
		***Total:* 41**		
Maquoketa	M1	4	May 10, 2012	42.34542 N/ 91.12314 W
	M3	2	May 18, 2012	42.52662 N/ 91.12823 W
	M5	5	Sept 11, 2012	42.55780 N/ 90.95532 W
Upper Iowa	UI-1	11	May 29, 2012	43.44114 N/ 92.03708 W
	UI-2	20	June 12, 2012	43.40186 N/ 91.80769 W
	UI-4	2	Aug 28, 2012	43.47419 N/ 92.07595 W

Genetic Data Collection

Whole genomic DNA was isolated from the fin tissue samples by the following extraction process: fins were digested overnight at an incubation temperature of 55° C in 300 µL TNES urea buffer (10 mM Tris, 125 mM NaCl, 10 mM EDTA-2Na, 0.5% SDS, 4 M Urea) and 2 µL Proteinase K. Next, DNA was isolated with 100 µL protein precipitation solution (4 M guanidine thiocyanate, 100 mM Tris-Cl), 1.5 µL RNAse, 300 µL 100% isopropanol, and a 70% ethanol wash.

Three microsatellite loci (LCO3, Seat405, and Ca13) were chosen due to their variability and quality of amplification (Table 2). These loci were then amplified by polymerase chain reaction (PCR) for all collected individuals to allow for allele identification. Initial denaturation

was done at 95° C for 2 minutes and final extension was done at 72°C for 10 minutes. Primer sets and individual thermal profiles can be found in Table 2. The PCR products for all individuals at each locus were sent to the Iowa State University DNA Facility for genotyping, producing a set of raw genotyping data. This data was scored in order to determine the true size of each allele using GENE MARKER version 1.3 (SoftGenetics, 2004) and related scoring techniques described in Guichoux et al. (2011).

Table 2. One tetranucleotide microsatellite marker (Seat405) and two dinucleotide microsatellite markers (LCO, Ca13) were amplified for all collected DNA samples. PCR was completed using an initial melting temperature of 95°C for 2 minutes, followed by the cycles indicated above and a final extension temperature of 72° C for ten minutes.

Locus	F-M13 primer sequence	R primer sequence	M13 tag sequence	Thermal Profile	
LCO3 Motif: [TG] (Turner et al., 2004)	5'-CACGACGTT GTAAAACGACG CAGGAGCGAAA CCATAAAT-3'	5'-AAACAG GCAGGACAC AAAGG-3'	5'- /5HEX/CAC GACGTTGTAAAA CGAC-3'	95° 20s 55.8° 20s 72° 30s 10x	95° 20s 48° 20s 72° 30s 30x
Seat405 Motif: [TTCA] (Skalski and Grose, 2006)	5'-CACGACGTT GTAAAACGACG GCCTCTGGTAA AGGAAACTAA-3'	5'CAGTCAGTC CGTCCATCCA T-3'	5'-/5HEX/CAC GACGTTGTAAA CGAC-3'	95° 20s 56.6° 20s 72° 30s 10x	95° 20s 48° 20s 72° 30s 30x
Ca13 Motif: [CA] (Dimsoski et al., 2000)	5'-CACGACGT TGTAAAACGAC GATCATTGATC CGCATGTCTC-3'	5'-CTCCCTG ACAGCAGCG ACC-3'	5'- /56-FAM/CAC GACGTTGTAAAA CGAC-3'	95° 20s 60.7° 20s 72° 30s 10x	95° 20s 48° 20s 72° 30s 30x

Analysis of Genotype Data

Once official genotyping data was collected, the data set was analyzed using a series of softwares (Table 3). First, the data was verified using MICROCHECKER version 2.2.3 (Van

Oosterhaut et al., 2004). This analysis tests raw data for scoring errors due to stutter peaks, large allele dropout, and null alleles. GENEPOP version 4.2 (Rousset, 2008) was used to test the data for two conditions: deviations from Hardy-Weinberg equilibrium (HWE) and linkage disequilibrium (LD). HWE assumes that in ideal populations, genotype frequency remains constant. Natural populations differ from these equilibrium conditions due to selective pressures and other processes. Linkage disequilibrium indicates that alleles at two or more loci are linked in some way and are not inherited independently. For both HWE and LD analysis, Markov chain method parameters were kept at the default values of 10,000 for dememorization, 20 for the number of batches, and 5,000 iterations per batch.

Descriptive statistics such as alleles per locus, number of effective alleles, number of private alleles, expected heterozygosity, and observed heterozygosity were obtained using GenAlEx version 6.5 (Peakall & Smouse, 2006). Data was entered into this software twice, once as a single, large population, and once with four population distinctions associated with drainage of origin. Allelic richness, which is a measure of allelic diversity corrected for population size, was obtained through FSTAT version 2.9.3 (Goudet, 2001).

Pairwise comparisons of genotypes between drainages were completed using ARLEQUIN version 3.5.1.3 (Excoffier et al., 2005). The values resulting from this analysis are termed F_{ST} and serve as a measurement of similarity (gene flow) or difference (isolation) between populations. Default settings of 110 permutations and a significance level of $P<.05$ were used to calculate F_{ST} values. ARELEQUIN was also used to complete a standard AMOVA test (Excoffier et al., 1992), which measured whether the observed genetic variance is primarily between drainages or within drainages. For the AMOVA test, the default setting of 1000 permutations was used.

Contemporary migration rates between drainages within the past few thousand generations were calculated using BAYESASS version 3.0 (Wilson & Rannala, 2003). This program estimates the probability of migration using a statistical method called Bayesian inference. Parameters were set to the default 3,000,000 iterations with a sampling frequency of 2000.

The population structure was determined using the program STRUCTURE 2.3 (Pritchard et al., 2000). This software determines the most probable number of genetic clusters or populations (denoted by the value K) present in the total pool of individuals. It assigns individuals to populations without need to identify an individual's drainage of origin, relying strictly on genotype to group individuals into populations. This method is known as Bayesian population assignment. Parameters included an independent allele frequency model (Pritchard et al., 2000), burn-in period of 1×10^5, and 1×10^6 MCMC iterations under the admixture model of ancestry; 20 iterations of K values ranging from 1 to 14 were completed. The most probable K value was determined by the likelihood and ΔK methods outlined in Pritchard et al. (2000) and Evanno et al. (2005), respectively.

Table 3. Genotype data analyses, listed in the order completed.

Program	Analysis	What it shows
MICROCHECKER version 2.2.3	Checks for null alleles, scoring of stuttering peaks, large allele dropout	• scoring errors
GENEPOP version 4.2	Exact test of HWE	• significant deviations from HWE
	Linkage disequilibrium	• loci that are inherited non-independently
GenAlEx version 6.5	Descriptive statistics	• no. alleles per locus • no. effective alleles • no. private alleles/drainage • heterozygosity
FSTAT version 2.9.3	Quantification of allelic diversity, corrected for population size	• allelic richness
ARLEQUIN version 3.5.1.5	Exact test of population differentiation	• level of similarity between pairs of drainages
	AMOVA	• how variation is distributed
BAYESASS version 3.0	Migration rate via Bayesian inference	• contemporary migration rates between drainages
STRUCTURE version 2.3	Bayesian population assignment	• most probable number of genetic clusters/populations

RESULTS

With all individuals pooled as a single population, MICROCHECKER did not detect any scoring errors due to stuttering, large allele dropout, or null alleles for both LCO3 and Ca13. However, null alleles were suspected in locus Seat405 due to the general excess of homozygous individuals. Historical processes such as population expansion have been known to lead to these genetic signatures. Therefore, Seat405 was kept in the study under the caveat that there is

homozygous excess. Seat405 was therefore expected to deviate from the normal genotype frequencies of HWE.

GENEPOP revealed significant deviations from Hardy-Weinberg equilibrium ($P<0.05$) in both Seat405 ($P=0$) and LCO3 ($P=0$). The locus Ca13 did not have any significant deviations ($P=.0634$). Although the MICROCHECKER and Hardy-Weinberg tests serve as important checkpoints to ensure the appropriateness and quality of the data, all three loci were maintained for further analysis because the deviations found could be explained by historical processes or the limited number of individuals in the data set. The other analysis completed in GENEPOP did not reveal any evidence of linkage disequilibrium between the three microsatellite loci under study. No paired loci had a P value of significance ($P<0.05$). These results indicated that the markers were indeed independently inherited and usable for further analysis.

Employment of GenAlEx produced descriptive statistics for the single, pooled population (Table 4A). Observed heterozygosities ranged from 0.477-0.685. As expected, the two loci that failed to meet HWE conditions in the GENEPOP test showed the greatest discrepancy between observed and expected heterozygosity. Seat405 had the greatest level of polymorphism with a total number of 11 alleles across all individuals, while LCO3 and Ca13 had moderate levels of polymorphism with 5 and 6 alleles, respectively. Descriptive statistics were utilized to reveal the observed and expected heterozygosities by drainage (Table 4B). Mean observed heterozygosities ranged from 0.515-0.783. The Maquoketa River drainage had more heterozygous individuals than expected for two of the three loci (LCO3, Ca13). In contrast, the Yellow and Upper Iowa river drainages were found to have lower heterozygosity than expected for all three loci, indicating lower levels of diversity.

Table 4. A) Allele size, number, and heterozygosity values for all individuals pooled as a single population. Expected heterozygosity (H_e) and observed heterozygosity (H_o) were quite similar for all loci- the greatest differences were seen in the two loci that failed HWE equilibrium in the GENEPOP analysis (denoted by an asterisk).

Locus	Allele Size Range (bp)	No. Alleles	No. Effective Alleles	H_e	H_o
***LCO3**	260-268	5	2.128	0.530	0.477
***Seat405**	204-332	11	6.760	0.852	0.685
Ca13	165-175	6	2.802	0.643	0.622
				Mean=0.675	Mean= 0.595

Table 4. B) H_e and H_o calculated for each drainage (Volga=V, Yellow=Y, Maquoketa=M, Upper Iowa= UI). The Yellow and Upper Iowa drainages showed evidence of decreased heterozygosity.

	Expected Heterozygosity (H_e)				**Observed Heterozygosity (H_o)**			
Locus	V	Y	M	UI	V	Y	M	UI
LCO3	0.540	0.471	0.574	0.405	0.588	0.390	0.750	0.364
Seat405	0.682	0.867	0.848	0.792	0.706	0.610	0.750	0.727
Ca13	0.680	0.633	0.605	0.606	0.706	0.610	0.850	0.455
mean	0.634	0.657	0.675	0.601	0.667	0.537	0.783	0.515

The FSTAT calculation of allelic richness was identical to the number of alleles found for the single, pooled population. By drainage (Table 4C), allelic richness values were similar to, but lower than the number of alleles per locus for the Yellow, Maquoketa, and Upper Iowa drainages. The Upper Iowa drainage showed the most difference between these two values for all three loci. One private allele was found in three of the four drainages (Volga, Yellow, Upper Iowa), while the Maquoketa drainage had no private alleles across the three loci.

Table 4. C) GenAlEx statistics for allelic richness, number of alleles, and number of private alleles by drainage. Only the Maquoketa drainage lacked private alleles. Letter codes for drainages are cross-referenced with Table 4B.

	Allelic Richness				No. Alleles				No. Private Alleles			
Locus	**V**	**Y**	**M**	**UI**	**V**	**Y**	**M**	**UI**	**V**	**Y**	**M**	**UI**
LCO3	4.000	3.549	3.000	4.355	4	4	3	5	0	0	0	0
Seat405	8.000	8.934	8.940	5.889	8	10	9	6	1	1	0	0
Ca13	4.000	3.946	4.700	4.026	4	4	5	5	0	0	0	1

Pairwise comparisons between drainages completed by ARLEQUIN resulted in the F_{ST} values found in Table 5. All values were significant except for one (Maquoketa v. Yellow). Of the significant values, genetic differentiation was quite low; F_{ST} values ranged from 0.01831-0.14230, indicating high levels of gene flow between the four river drainages.

Table 5. F_{ST} values associated with pairwise comparisons between drainages. Statistically significant values ($P<0.01$) are shaded. Letter codes for drainages are cross-referenced with Table 4B.

	V	Y	M	UI
V	-			
Y	0.08515	-		
M	0.01831	0.01359	-	
UI	0.1423	0.02886	0.0489	-

The AMOVA (Analysis of Molecular Variance) test revealed that the majority of the total variation in the data set was among individuals (86.41%). At the drainage level, a higher percentage of the variation was seen within drainages (8.49%) rather than between them (5.10%).

Contemporary migration rates between drainages were estimated using BAYESASS (Table 6). The largest rates of migration were from the Maquoketa to the Volga (m= 0.191471),

and from the Yellow to the Maquoketa (m=0.184204). All other drainage pairings were found to have migration rates less than 0.01.

Table 6. Mean contemporary migration rate estimates (m) and confidence intervals found by BAYESASS. Highest migration was found from Maquoketa to Volga and from Yellow to Maquoketa. Letter codes for drainages are cross-referenced with Table 4B.

From	To	Mean Migration Rate (m)	95% confidence interval
Y	V	0.0432132	(0.000252485, 0.172837)
M	V	0.191471	(0.000849474, 0.316643)
UI	V	0.0224089	(0.000160786, 0.0937191)
V	Y	0.0512478	(0.000797722, 0.132212)
M	Y	0.0365428	(0.000234965, 0.211226)
UI	Y	0.032852	(0.000243163, 0.15027)
V	M	0.0558899	(0.000450195, 0.241679)
Y	M	0.184204	(0.0166958, 0.306394)
UI	M	0.0624203	(0.000642969, 0.203912)
V	UI	0.00907734	(2.46274e-05, 0.0416332)
Y	UI	0.0547827	(4.99566e-05, 0.25079)
M	UI	0.0174815	(2.08867e-05, 0.202259)

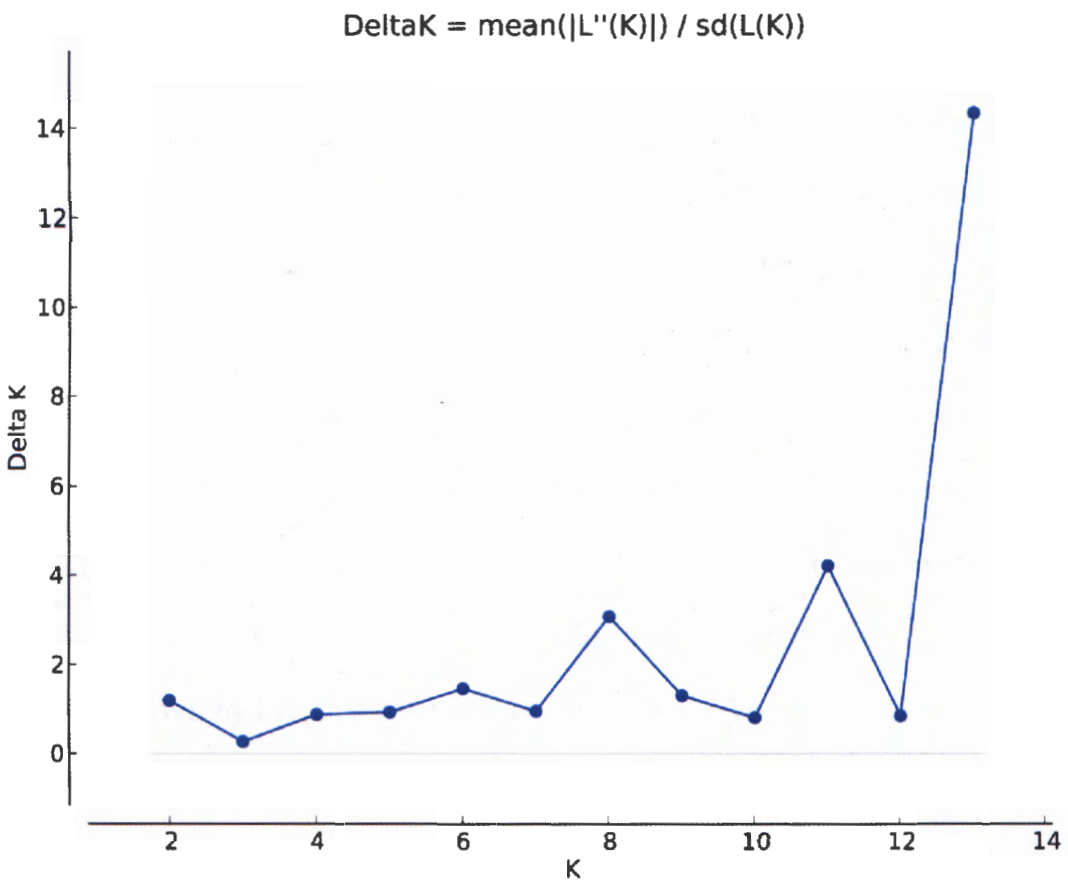

Figure 5. The most probable number of genetic populations was predicted by STRUCTURE to be 13 (equal to the number of sample localities).

Surprisingly, the results from STRUCTURE showed a much higher level of differentiation in data set. A total 13 distinct genetic clusters were predicted, as evidenced by the maximum reached by the the ΔK plot (Figure 5). Contrary to all previous analyses, there appeared to be population distinctions within the same river drainage.

DISCUSSION

Population structure quantifies the consequences of demographic processes in natural groups of organisms. For selectively neutral microsatellite markers, the allele and genotype frequencies should be in line with Hardy-Weinberg equilibrium, which assumes no interference from evolutionary events. For longnose dace populations in northeast Iowa, it is unlikely that new alleles have arisen due to mutation since the last glacial maximum. If a mutation even has taken place, not enough time has passed for significant signal to be evident, due to the deep coalescence of microsatellites (Zink & Barrowclough, 2008). Thus, deviations from Hardy-Weinberg equilibrium for longnose dace, a species known to be of concern, can be expected to imply a loss of genetic diversity. In this genetic study, two loci (Seat405 and LCO3) were found to deviate from HWE by having a lower number of heterozygous individuals than expected, confirming the hypothesis of the reduced genetic variation.

However, there are alternate explanations for these deviations. Population expansion following bottleneck events can lead to this signature without necessarily a reduction in population health. This pattern is most likely explained by the shaping effects of ancient glacial retreat, rather than contemporary, human-induced landscape modifications. Also, given the parameters of the study and the relatively small number of individuals and localities sampled, deviations of this kind can be expected to be present simply due to incomplete sampling and the low number of loci used.

Based on the difference between expected and observed levels of heterozygosity, the Upper Iowa and Yellow river drainages appear to have lower levels of diversity than the other drainages. Increases in the number of homozygous individuals, no matter how slight, can be a signal of future population vulnerability. The high mutation rate of microsatellites makes them highly sensitive to changes in diversity. Thus, increased homozygosity detected at microsatellite

loci could indicate the potential for similar changes in regions of the genome that could undergo selection and reductive effects (Chistiakov et al., 2006).

Allelic richness has been proposed to be an accurate measure of diversity due its normalization of population size (Leberg, 2002). In the Upper Iowa drainage, the difference observed between the absolute number of alleles and allelic richness value confirm that population size is a factor affecting diversity. The allelic richness value can be interpreted several ways. For example, the limited number of samples from this drainage could be causing the analyses to reveal a lower level of diversity than is truly there. Conversely, low genetic diversity could be contributing to small population size through destructive phenomena such as inbreeding.

Despite the differences between drainages detected by GenAlEx and FSTAT, pairwise comparisons completed with ARLEQUIN uncovered the opposite. F_{ST} values were statistically significant across the majority of the data set and low enough to suggest extensive gene flow between river drainages. Well-connected populations have a greater probability of sustaining themselves in an ever-changing environment. Therefore, longnose dace in northeast Iowa appear to constitute one large population that could be resilient to future fragmentation and extirpation. The AMOVA results also support the high connectivity of northeast Iowa's longnose dace populations. The vast majority of the molecular variance was found to be among individuals. Furthermore, differences observed between drainages contributed the least to the overall diversity of the genotypes. These results strongly suggest the presence of one large genetic population in northeast Iowa.

The BAYESASS results indicate that recent migration between drainages is low and rather consistent. Although this may seem contrary to the high connectivity suggested by

previous analyses, it is important to note that the migration rates calculated were on a contemporary scale. Longnose dace are known to have high fidelity to fast moving riffles. Because the river drainages under study are not currently connected and have not been for a long time, it is intuitive that recent migration rates are low. In future analyses, historical migration rates can be expected to be much higher than contemporary ones.

Contemporary migration rates can perhaps explain why the Maquoketa River drainage did not have private alleles. The results from BAYESASS indicated two pairings with a noticeably higher migration rate: Maquoketa to Volga, and Yellow to Maquoketa. Because the relative migration to and from the Maquoketa River appears to be higher than the other drainages, it is less likely that longnose dace in the Maquoketa would possess alleles that are exempt from gene flow with other rivers.

The results from STRUCTURE were in clear opposition to those of ARLEQUIN, which suggested one large population across northeast Iowa. Instead of revealing one large genetic cluster incorporating all four rivers, 13 population distinctions were made. The detection of populations within the same river drainage does not seem to be in line with the relatively homogenous genotype data. The most likely explanation for this discrepancy comes from the limitation of the data. Some collection localities yielded as few as 2 individuals, making it unlikely that a single locality could constitute a unique population. Along with this, the limited variance in the data could have hindered the analysis by failing to produce significant signal of the true genetic structure.

CONCLUSION

Rhinichthys cataractae is a Species of Greatest Conservation Need in a geographically and ecologically unique region. Examining the health of longnose dace populations in northeast Iowa is important for ensuring their conservation. The population genetic analysis completed by this study has taken some of the first steps towards understanding the structure, diversity and health of longnose dace populations in northeast Iowa. Microsatellites were used as a sensitive and selectively neutral tool to observe the demographic processes of interest.

Even though the number of microsatellite loci and incomplete sampling were some limiting factors of this study, some light was shone on the diversity of longnose dace in northeast Iowa. The results of this study indicate populations of longnose dace inhabiting the Yellow, Maquoketa, Volga, and Upper Iowa drainages most likely constitute one large genetic population. Evidence for this is seen in both the low F_{ST} values and the low variance between drainages revealed by AMOVA. However, the vulnerability of this species was confirmed by the slight deviations from HWE and low allelic richness values found by GENEPOP, GenAlEx and FSTAT. Results from STRUCTURE were inconclusive due to the low number of samples.

To provide a more comprehensive picture, more samples and microsatellites need to be included in the data set and further analyses should be carried out. Utilization of the programs MIGRATE-N (Beerli, 2008; Beerli & Palczewski, 2010), BARRIER (Manni et al., 2004), and BOTTLENECK (Piry et al., 1999) would serve to further evaluate the genotype data, revealing how many individuals contribute offspring to future generations, historical migration rates, barriers to gene flow between populations, and evidence of decline in population size. This future research will add additional insight by detecting and quantifying some of the mechanisms that can lead to changes in the evolutionary success of a species.

Within the scope of this study, the results suggest that populations of longnose dace in northeast Iowa have little potential for genetic isolation, fragmentation, and extirpation in the near future. However, this study is very preliminary in the journey to understanding the health status of longnose dace. At this point, it is difficult to know how the results of this study compare to the real genetic structure of longnose dace due to limited sample size and low number of loci. With more samples and microsatellites, a clearer picture of the genetic structure of northeast Iowa populations will emerge. Overall, this study provides insight into one of northeast Iowa's declining species. The results, in conjunction with a larger study currently in progress by Dr. Berendzen's lab examining habitat and climate modeling, will someday provide a comprehensive look at what is threatening *Rhinichthys cataractae* and serve as a guide for how best to manage conservation efforts.

Literature Cited

Amos, W., & Harwood, J. (1998). Factors affecting levels of genetic diversity in natural populations. *Phil. Trans. R. Soc. B*, 353, 177-186.

Amos, W., & Balmford, A. (2001). When does conservation genetics matter? *Heredity*, 87, 257-265.

Apodaca, J. J., Rissler, L. J., & Godwin, J. C. (2012). Population structure and gene flow in a heavily disturbed habitat: implications for the management of the imperiled Red Hills salamander (*Phaeognathus hubrichti*). *Conservation Genetics*, 13, 913-923.

Austin, J. D., Jelks, H. L., Tate, B. T., Johnson, A. R., & Jordan, F. (2011). Population genetic structure and conservation genetics of threatened Okaloosa darters (Etheostoma okaloosae). *Conservation Genetics*, 12, 981-989.

Avise, J. C. (2000). *Phylogeography, The History and Formation of Species*. Harvard University Press, Cambridge.

Beerli, P. (2008). MIGRATE documentation (version 3.0). Technical Report. http://popgen.sc.fsu.edu.

Beerli, P., & Palczewski, M. (2010). Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*, 185, 313-326.

Berendzen, P. B., Simons, A. M., & Wood, R. M. (2003). Phylogeography of the northern hogsucker, *Hypentelium nigricans* (Teleostei: Cypriniformes): genetic evidence for the existence of the ancient Teays Rivers. *Journal of Biogeography*, 30, 1139-1152.

Berendzen, P. B., Gamble, T., & Simons, A. M. (2008). Phylogeography of the bigeye chub *Hybopsis amblops* (Teleostei: Cypriniformes): early Pleistocene diversification and post-glacial range expansion. *Journal of Fish Biology*, 73, 2021-2039.

Berendzen, P. B., Dugan, J. F., & Gamble, T. (2010). Post-glacial expansion into the Paleozoic Plateau: evidence of an Ozarkian refugium for the Ozark minnow *Notropis nubilus* (Teleostei: Cypriniformes). *Journal of Fish Biology*, 77, 1114-1136.

Burkhead, N. M. (2012). Extinction rates in North American freshwater fishes, 1900-2010. *BioScience*, 62, 798-808.

Chistiakov, D. A., Hellemans, B., & Volckaert, F. A. M. (2006). Microsatellites and their genomic distribution, evolution, function, and applications: A review with special reference to fish genetics. *Aquaculture*, 255, 1-29.

COSEWIC (2007). COSEWIC assessment and update status report on the Nooksack dace *Rhinichthys cataractae* s.s.p. in Canada. Committee on the Status of Endangered Wildlife in Canada. Ottawa. vii + 27 pp. (www.sararegistry.gc.ca/status/status_e.cfm).

Cox, C., & Hug, A. (2012). Murky Waters: Farm Pollution Stalls Cleanup of Iowa Streams. *Environmental Working Group,* 1-52.

DeSalle, R., & Amato, G. (2004). The expansion of conservation genetics. *Nature*, 5, 702-712.

Dimsoski, P., Toth, G. P., & Bagley, M. J. (2000). Microsatellite characterization in central stoneroller *Campostoma anomalum* (Pisces: Cyprinidae). *Molecular Ecology,* 9, 2155-2234.

Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, 14, 2611-2620.

Evans, J. S., Jackson, L. J., Habibi, H. R., & Ikonomou, M. G. (2012). Feminization of longnose dace (Rhinichthys cataractae) in the Oldman River, Alberta, (Canada) provides evidence of widespread endocrine disruption in an agricultural basin. *Scientifica*, 1-11.

Excoffier, L., Smouse, P. E., & Quattro, J. M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes, application to human mitochondrial DNA restriction data. *Genetics*, 131, 479-491.

Excoffier, L., Laval, G., & Schneider, S. (2005). Arelquin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online,* 1, 47-50.

Fluker, B. L., Kuhajda, B. R., Lang, N. J., & Harris, P. M. (2010). Low genetic diversity and small long-term population sizes in the spring endemic watercress darter, *Etheostoma nuchale. Conservation Genetics*, 11, 2267-2279.

Frankham, R., Ballou, J. D., & Briscoe, D. A. (2005). *Introduction to Conservation Genetics.* Cambridge: Press Syndicate of the University of Cambridge.

Girard, P., & Angers, B. (2006). The impact of postglacial marine invasions on the genetic diversity of an obligate freshwater fish, the longnose dace (*Rhinichthys cataractae*), on the Quebec peninsula. *Canadian Journal of Fisheries & Aquatic Sciences*, 63, 1429-1438.

Girard, P., & Angers, B. (2011). The functional gene diversity in natural populations over postglacial areas: the shaping mechanisms behind genetic composition of longnose dace (*Rhinichthys cataractae*) in northeaster North America. *Journal of Molecular Evolution*, 73, 45-57.

Guichoux, E., Lagache, L., Wagner, S., Chaumeil, P., Leger, P., Lepais, O., Lepoittevin, C., . . . Petit, R. J. (2011). Current trends in microsatellite genotyping. *Molecular Ecology*, 11, 591-611.

Goudet, J. (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). Available from http://www2.unil.ch/popgen/softwares/fstat.htm. Updated from Goudet (1995).

Hewitt, G. (2000). The genetic legacy of the Quaternary ice ages. *Nature*, 405, 907–913.

Hobbs, H. (1999). Origin of the driftless area by subglacial drainage- a new hypothesis. *Geological Society of America,* Special Paper, 337, 93-102.

Iowa Department of Natural Resources. (2007). *Iowa Wildlife Action Plan-Securing a Future for Fish and Wildlife: a Conservation Legacy for Iowans*. Chapter 3, p.48. Retrieved from http://www.iowadnr.gov/Environment/WildlifeStewardship/IowaWildlifeActionPlan.aspx

Jehle, R., & Arntzen, J. W. (2002). Review: Microsatellite markers in amphibian conservation genetics. *Herpetological Journal*, 12, 1-9.

Lehberg, P. L. (2002). Estimating allelic richness: effects of sample size and bottlenecks. Molecular Ecology, 11, 2445-2449.

Loan-Wilsey, A., McNeely, R. L., Brown, P. D., Kane, K. L., & Pierce, C. L. (2004). A comprehensive biological inventory database for the Iowa Aquatic Gap Analysis Project. Gap Analysis Bulletin No.12:2-3. USGS/BRD/Gap Analysis Program, Moscow, Idaho.

Lyons, J., Stewart, J. S., & Mitro, M. (2010). Predicted effects of climate warming on the distribution of 50 stream fishes in Wisconsin, U.S.A. *Journal of Fish Biology*, 77, 1867-1898.

Manni, F. E., Guerard, E., & Heyer, E. (2004). Geographical patterns of (genetic, morphologic, linguitc) variation: how barriers can be detected by 'Monmonier's algorithm'. *Human Biology*, 76, 173-190.

Mayden, R. L. (1985). Biogeography of Ouachita Highland fishes. *The Southwestern Naturalist*, 30, 195-211.

Mayden, R. L. (1988). Vicariance biogeography, parsimony, and evolution in North American freshwater fishes. *Systematic Zoology*, 37, 329-355.

Near, T. J., Page, L. M. & Mayden, R. L. (2001). Intraspecific phylogeography of *Percina evides* (Percidae: Etheostomatinae): an additional test of the central highlands pre-Pleistocene vicariance hypothesis. *Molecular Ecology*, 10, 2235-2240.

North American Native Fishes, Official Gallery. (2010). *Rhinichthys cataractae.* JPEG file. http://gallery.nanfa.org/v/members/blakemarkwell/cyprinidae/rhinichthys/cataractae01.html?g2_imageViewsIndex=1

Peakall, R., & Smouse, P. E. (2006). Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, 6: 288–295.

Piry, S., Luikart, G., & Cornuet, J. M. (1999). BOTTLENECK: a computer program for detecting recent reduction in the effective population size using allele frequency data. *Journal of Heredity*, 90, 502-503.

Pritchard, J. K., Stephens, M., & Donnely, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.

Ray, J. M., Wood, R. M. & Simons, A. M. (2006). Phylogeography and post-glacial colonization patterns of the rainbow darter, *Etheostoma caeruleum* (Teleostei: Percidae). *Journal of Biogeography*, 33, 1550-1558.

Rousset, F. (2008). GENEPOP'007: a complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, 8, 103-106.

Rowe, K. C., Heske, E. J., Brown, P. W., & Paige, K. N. (2004). Surviving the ice: northern refugia and postglacial colonization. *Proceeding of the National Academy of Sciences*, 101, 10355-10359.

Rowe, D. C., Pierce C. L., & Wilton T. F. (2009a). Physical Habitat and Fish Assemblage Relationships with Landscape Variables and Multiple Spatial Scales in Wadeable Iowa Streams. *North American Journal of Fisheries Management*, 29, 1333-1351.

Rowe, D. C., Pierce C. L., & Wilton T. F. (2009b). Fish Assemblage Relationships with Physical Habitat in Wadeable Iowa Streams. *North American Journal of Fisheries Management*, 29, 1314-1333.

Skalski, G. T., & Grose, M. J. (2006). Characterization of microsatellite loci in the creek chub (*Semotilus atromaculatus*). *Molecular Ecology Notes,* 6, 1240-1242.

SoftGenetics, State College, Pennsylvania. (2004). GeneMarker Operation Manual version 1.3.

Strange, R. M., & Burr, B. M. (1997). Intraspecific phylogeography of North American highland fishes: A test of the Pleistocene vicariance hypothesis. *Evolution*, 51, 885-897.

Van Oosterhaut, C., Hutchinson, W. F., Wills, D. P. M., & Shipley, P. (2004). MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4, 535-553.

Turner, T. F., Dowling, T. E., Broughton, R. E., & Gold, J. R. (2004). Variable microsatellite markers amplify across divergent lineages of cyprinid fishes (subfamily Leusicinae). *Conservation Genetics,* 5, 279-281.

Wilson, G. A., & Rannala, B. (2003). Bayesian inference of recent migration rates using multilocus genotypes. *Genetics*, 163, 1177-1191.

Zellmer, A. J., & Knowles, L. L. (2009). Disentangling the effects of historic vs. contemporary landscape structure of population genetic divergence. *Molecular Ecology*, 18, 3593-3602.

Zink, R. M., & Barrowclough, G. F. (2008). Mitochondrial DNA under siege in avian phylogeography. *Molecular Ecology*, 17, 2107-2121.