Population genetics and phylogeography of *Aplodinotus grunniens* (freshwater drum)

A thesis presented to

The Faculty of Graduate Studies

of

Lakehead University

by

Amber Jarvinen

In partial fulfillment of requirements
for the degree of
Master of Science in Biology

December 13, 2011

© Amber Jarvinen, 2011

Abstract

The purpose of this study is to determine the population genetic and phylogeographic structure of *Aplodinotus grunniens* (freshwater drum) populations through analysis of the NADH dehydrogenase subunit 2 (ND2) region and the control region of mitochondrial DNA. Hypotheses were formed regarding genetic variation in freshwater drum, identification of regions where freshwater drum survived glaciation, and source populations and colonization routes. Primers specific to the ND2 and control regions of freshwater drum were developed and genetic analysis performed. Genetic data were analyzed to assess relationships within and between populations. The contemporary population structure of sampled freshwater drum populations was analyzed. I determined there to be significant relationships between populations, in some cases these relationships were correlated with geographic proximity and common drainage basins as observed in Lake Winnipeg and Nelson River (Hudson Bay drainage) and Lake Erie and Lake Ontario (Great Lakes drainage). Phylogenetic analyses using both parsimony and maximum likelihood methods revealed genetic relationships among populations. This data was further used to infer phylogeography and historical relationships of freshwater drum. Populations within the Mississippi, Missouri and Ohio Rivers were determined to be the primary colonization sources in the freshwater drum population. My results suggest that freshwater drum populations diverged prior to the Pleistocene and were isolated in separate refugia. The distribution pattern of freshwater drum appears to be one of pre-Pleistocene vicariance in refugia followed by postglacial dispersal and vicariance.

Lay Summary

The mission statement of the Lakehead University Department of Biology is as follows: "Faculty and students in the Department of Biology are bound together by a common interest in explaining the diversity of life, the fit between form and function, and the distribution and abundance of organisms." My study focuses on the first and third of these themes. The contribution of my study in advancing knowledge of the diversity of life is based on analysis of genetic variation within the freshwater drum population. I examined the genetic composition of freshwater drum in various sample locations, analyzing genetic diversity to explain the relationships within and among subpopulations. This data was then analyzed to determine the origins and factors which contributed to the distribution and abundance of the freshwater drum species. I proposed theories of source populations of freshwater drum and the colonization of each sampled population, including glacial refugia and post-glacial colonization routes. In my study I have been able to advance understanding of the diversity and distribution of freshwater drum, which also enables further understanding of other taxa, particularly North American freshwater fish species.

Acknowledgements

I thank Dr. Stephen Hecnar for being a committee member for this project. Special thanks to Dr. Carney Matheson for being on the committee of this project and for providing exceptional support and assistance. I thank my external examiner Dr. Anoindo Choudhury for his insightful questions. Thank you to Stephen and Renée Fratpietro and the staff of the Lakehead University Paleo DNA Laboratory for all of their help and advice. Thank you to NSERC RCD and Lakehead University Senate Research for providing funding for this project. Thank you to my family and friends for their encouragement and patience. Lastly, I would like to thank thesis advisor Dr. Joseph Carney for providing me with the opportunity to pursue this study and supporting and advising me throughout.

Table of Contents

List of Tables v

List of Figures ix

List of Appendices xxiii

Introduction 1

Population genetics 2

Phylogeography 8

Mitochondrial DNA 11

Aplodinotus grunniens (freshwater drum) 16

Glacial geography of North America 18

Purpose and hypotheses 38

Methods 41

Materials and Methods 41

Control region – initial amplification 43

Amplification with primers specific to freshwater drum 46

ND2 region 49

Sequence analysis 52

Intra-population analysis 56

Inter-population analysis 59

Results 62

Control region – intra-population analysis 62

Control region – inter-population analysis 66

ND2 region – intra-population analysis 81

ND2 region – inter-population analysis 85

Discussion 99

Contemporary population analysis 120

Historical relationships and postglacial dispersal of freshwater drum........................149

Future studies 165

Literature cited..168

List of Tables

Table 1. AMOVA (analysis of molecular variance) table.

The formulas used to determine each partition within the AMOVA table are displayed. 60

Table 2. Nucleotide diversity of the control region of freshwater drum.

Nucleotide diversity was calculated based on the number of differences between all pairs of haplotypes in each sample. Samples with higher values of π are more diverse. Samples marked with an asterisk* are significant. 63

Table 3. Shared control region haplotypes among sampled populations.

Number of haplotypes shared among each population sampled. .... .64

Table 4. Control region mismatch distribution.

Column 2 illustrates the distribution of differences between pairs of haplotypes. Columns 3 and 5 illustrate tau values under the demographic or spatial expansion models, and columns 4 and 6 illustrate goodness of fit for these models. P < 0.05 is statistically significant and marked in bold. 65

Table 5. Control region analysis of Tajima's D test of neutrality.

P < 0.05 are statistically significant are marked in **bold**. Negative Tajima's D values indicate population size expansion, positive selection or a recent bottleneck. 67

Table 6. Control region AMOVA analysis of populations.

(a) Analysis of the variation among and within populations; (b) Analysis of the variation among and within drainage basins. 68

Table 7. Population pairwise F_{ST} values based on control region analysis.

0.00000 indicates no differentiation between populations, 1.00000 indicates complete differentiation. Negative values should be considered as 0.00000. Theoretically, all values >0.25000 indicate significant differentiation and are highlighted in **bold**. 69

Table 8. Control region pairwise F_{ST} P values.

Mean probability plus or minus standard deviation. P values <0.005 are significantly different and are marked in **bold**. 70

Table 9. Haplotype frequency based on Phylip analysis of the control region.

Frequency of haplotypes with each clade as determined by parsimony analysis using Phylip. This table corresponds to Figure 24. 72

Table 10. Haplotype frequency based on Bayesian analysis of the control region.

Frequency of haplotypes with each clade as determined by Bayesian maximum likelihood analysis. This table corresponds to Figure 26. 75

Table 11. Bayesian model of best fit based on control region analysis.

The Bayesian model of best fit is highlighted in bold. 79

Table 12. Nucleotide diversity of the ND2 region of freshwater drum.

Nucleotide diversity was calculated based on the number of differences between all pairs of haplotypes in each sample. Samples with higher values of π are more diverse. Samples marked with an asterisk* are significant. 82

Table 13. Shared ND2 region haplotypes among sampled populations.

Number of haplotypes shared among each population sampled. 83

Table 14. ND2 region mismatch distribution.

Column 2 illustrates the distribution of differences between pairs of haplotypes. Columns 3 and 5 illustrate tau values under the demographic or spatial expansion models, and columns 4 and 6 illustrate goodness of fit for these models. $P < 0.05$ is statistically significant and marked in **bold**. 84

Table 15. ND2 region analysis of Tajima's D test of neutrality.

$P < 0.05$ are statistically significant are marked in **bold**. Negative Tajima's D values indicate population size expansion, positive selection or a recent bottleneck. 86

Table 16. ND2 region AMOVA analysis of populations.

(a) Analysis of the variation among and within populations; (b) Analysis of the variation among and within drainage basins. 87

Table 17. Population pairwise F_{ST} values based on ND2 region analysis.

0.00000 indicates no differentiation between populations, 1.00000 indicates complete differentiation. Negative values should be considered as 0.00000. Theoretically, all values >0.25000 indicate significant differentiation and are highlighted in bold... 88

Table 18. ND2 region pairwise F_{ST} P values.

Mean probability plus or minus standard deviation. P values <0.005 are significantly different and are marked in bold. 89

Table 19. Haplotype frequency based on Phylip analysis of the ND2 region.

Frequency of haplotypes with each clade as determined by Phylip parsimony analysis. This table corresponds to Figure 29. 90

Table 20. Haplotype frequency based on Bayesian analysis of the ND2 region.

Frequency of haplotypes with each clade as determined by Bayesian maximum likelihood analysis. This table corresponds to Figure 31. 94

Table 21. Bayesian model of best fit based on ND2 region analysis.

The Bayesian model of best fit is highlighted in bold. 97

List of Figures

Figure 1: Forces of evolution.

Mutation alters an allele, and is the source of all genetic change. Genetic drift affects the distribution of alleles in a population by random chance, it is not adaptive. Selection rids the population of deleterious alleles and preserves beneficial alleles. Migration introduces alleles to a population, altering allelic frequency. These forces, individually or working in concert, affect the evolution of a population. 4

Figure 2. A phylogenetic tree.

A phylogenetic tree, the length of each branch is correlated with the amount of genetic change. 10

Figure 3. The mitochondrial genome.

The mitochondrial genome is a circular genome encoding for 37 genes used in phylogeographic studies. 12

Figure 4. Distribution range of freshwater drum.

Present day distribution of freshwater drum. The dispersal range stretches from the mouth of the Nelson River at Hudson Bay to the southern United States. The species exists in all the Great Lakes, with the exception of Lake Superior. 17

Figure 5. North America in the last glacial maximum.

Wisconsin glaciation in North America, the greatest extent of the ice sheet dating approximately 18000 years ago. The Great Lakes and Lake Winnipeg, although shown, did not exist. 19

Figure 6. Preglacial drainage systems of North America in the Pliocene.

Pliocene drainage patterns (approximately 2.5 million years ago) are superimposed in dark lines over modern day drainage systems. Drainage systems noted in this study are in bold. (1) Plains Stream; (2) Old Red River; (3) Old Ouachita River; (4) Old Arkansas River; (5) White River; (6) Old Grand-Missouri River; (7) Ancestral Iowa River; (8) Old Mississippi River; (9) Old Teays-Mahomet River; (10) Old Kentucky River; (11) Old Licking River; (12) Old Big Sandy River; (13) Kanawha River; (14) Kaskasia River; (15) Wabash River; (16) Green River; (17) Old Ohio River; (18) Old Cumberland River; (19) Old Duck River; (20) Old Tennesee River; (21) Appalachian River; (22) Old Tallapoosa River; (23) Mobile Basin; (24) Hudson Bay drainage; (25) St. Lawrence River. 21

Figure 7. Major North American drainage systems in the Tertiary.

Figure A displays the major drainage systems in the Pliocene. Figure B illustrates major drainage systems in the early Pleistocene. Figure C displays drainage systems in the Holocene. The southern course of the Mississippi River is indicated by the arrow. 23

Figure 8. The ancient Teays River.

Drainage systems differed from the present day prior to the Wisconsin glaciation. The Teays River drained much of the eastern-central United States into the Gulf of Mexico. The Erigan River, hypothesized to drain into the Teays River, existed in what is now the Lake Erie basin and drained into the St. Lawrence. These systems were altered by the advance and retreat of the ice sheet, the drainage system of the Teays River being altered to become the present day Ohio River. 25

Figure 9. Glacial Lake Tight.

Glacial Lake Tight was formed by damming of the Teays River in Ohio. It is hypothesized that water spilled into the Licking River valley (dashed line) allowing for dispersal of freshwater fishes into the old Kentucky River and old Lower Ohio River 26

Figure 10. The Niobrara River Valley.

The Niobrara River was a glacial refugium in northern Nebraska which displays great diversity of plant and animal species. 28

Figure 11. Glacial Lake Wisconsin.

Glacial Lake Wisconsin came into existence approximately 1.8 million years ago. Black `lines in the main figure, and dashed lines in the inset, indicate the leading edge of the ice. 29

Figure 12. Glacial Great Lakes and Lake Agassiz.

Proglacial lakes which provided dispersal routes for freshwater fishes. A. The maximum extent of glacial lakes in northeastern North America. Light stippling indicates salt water, dark stippling indicates the maximum extent of freshwater lakes. AG, Lake Agassiz; CS, Champlain Sea; ER, Lake Erie; HU, Lake Huron; MI, Lake Michigan; OB, Barlow-Ojibway; ON, Lake Ontario; SU, Lake Superior; TS, Tyrell Sea. Figures B-D illustrate glacial outlets at various periods throughout the Pleistocene. Al, Allegheny; AT, Au Train; Bu, Buffalo; Ch, Chicago; FW, Fort Wayne; GB, Green Bay; PH, Port Huron; Sh, Sheyenne; SL, St.Lawrence; SM, Straits of Mackinac; Sq, Susquehanna; Wa, Warren. Arrows indicate the direction of flow. 31

Figure 13. Potential dispersal routes to northeastern North America.

Outlets are numbered as follows: (1) Sheyenne; (2) Warren; (3) Chicago; (4) Fort Wayne; (5) Lower peninsula of Michigan. 34

Figure 14. Drainage basins of North America.

Three major river drainage systems are applicable to present day Freshwater Drum habitat (identified by stars): the Mississippi, St. Lawrence and Nelson-Saskatchewan River basins..37

Figure 15. Freshwater drum sample collection sites.

Eleven sites were sampled across the distribution range. (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB. 42

Figure 16. Polyacrylamide gel of freshwater drum control region sequences using red drum primers.

465 base pair control region sequence using red drum primers. Bands are faint and non-specific. Amplified region is circled. Base pairs corresponding to bands are marked on the right of the figure. 45

Figure 17. Analysis of control region primers.

Primers were tested using the Operon analysis tool to verify suitability. MW, molecular weight; GC, guanine/cytosine; TM, melting temperature. 48

Figure 18. Polyacrylamide gel of freshwater drum control region sequences.

438 base pair control region sequence using primers designed for freshwater drum. The size marker on the right of the gel is a 50bp ladder. Base pairs corresponding to bands are marked on the right of the figure. 50

Figure 19. Analysis of ND2 region primers.

Primers were tested using the Operon analysis tool to verify suitability. MW, molecular weight; GC, guanine/cytosine; TM, melting temperature. 51

Figure 20. Agarose gel of freshwater drum ND2 sequences.

1 kbp ladder is not shown. Sequences are 500 bp, lane 3 is a negative control. 53

Figure 21. Alignment and analysis of sequences.

Sequences were aligned and analyzed using Gap4 software. Polymorphisms were identified and checked by eye. Illustrated is a partial section of a control region sequence. A base substitution is displayed in the black rectangle – the fourth sequence from the top has substituted adenine for guanine. A base substitution of thymine for cytosine is depicted in the lowest 2 sequences, identified by the black line. An insertion of thymine is illustrated inside the black circle. 54

Figure 22. Comparison of *Aplodinotus grunniens* and *Sciaenops ocellatus* mitochondrial control region sequences.

Overall maximal identity was determined to be 90% between sequences. Polymorphisms are highlighted in yellow. Query is the amplified sequence of *Aplodinotus grunniens* from this study, Sbjct refers the mitochondrial control region of *Sciaenops ocellatus* at GenBank accession number EU363526.1. 55

Figure 23. Control region cladogram based on parsimony analysis.

Cladogram based on control region analysis using Phylip. Haplotypes are indicated by "h___", capital letters refer to individual clades. 73

Figure 24. Haplotype frequency based on parsimony analysis of the control region.

Frequency of haplotypes within each clade as determined by parsimony analysis. Clade A = blue; clade B = red; clade C = green; clade D = black; clade E = yellow; clade F = orange. This figure corresponds to Table 9. 74

Figure 25. Control region cladogram based on Bayesian analysis.

Cladogram based on control region analysis using Mr.Bayes. Haplotypes are indicated by "h___", capital letters refer to individual clades......... 76

Figure 26. Haplotype frequency based on Bayesian analysis of the control region.

Frequency of haplotypes within each clade as determined by Bayesian maximum likelihood analysis. Clade A = blue; clade B = red; clade C = green; clade D = black; clade E = yellow; clade F = orange. This figure corresponds to Table 11. 77

Figure 27. Control region minimum spanning network.

Black circles represent haplotypes that were not sampled or no longer exist. Haplotypes are noted as four digit numbers, the numbers below each haplotype denote the number of individuals with that haplotype, followed by the number of populations containing at least one individual with that haplotype. Circle size is representative of the number of haplotypes represented in the sample. 80

Figure 28. ND2 region cladogram based on parsimony analysis.

Cladogram based on ND2 region analysis using Phylip. Haplotypes are indicated by "h___", capital letters refer to individual clades. 91

Figure 29. Haplotype frequency based on parsimony analysis of the ND2 region.

Frequency of haplotypes within each clade as determined by parsimony analysis. Clade A = blue; clade B = red. This figure corresponds to Table 21. 92

Figure 30. ND2 region cladogram based on Bayesian analysis.

Cladogram based on ND2 region analysis using Mr.Bayes. Haplotypes are indicated by "h___", capital letters refer to individual clades. 95

Figure 31. Haplotype frequency based on Bayesian analysis of the control region.

Frequency of haplotypes within each clade as determined by Bayesian maximum likelihood analysis. Clade A = blue; clade B = red. This figure corresponds to Table 22. 96

Figure 32. ND2 region minimum spanning network.

Black circles represent haplotypes that were not sampled or no longer exist. Haplotypes are noted as four digit numbers, the numbers below each haplotype denote the number of individuals with that haplotype, followed by the number of populations containing at least one individual with that haplotype. Circle size is representative of the number of haplotypes represented in the sample. 98

Figure 33. Correlation between latitude and genetic diversity.

There is a trend towards decreased genetic diversity in previously glaciated areas, but exceptions clearly exist. Interior circles indicate ND2 nucleotide diversity, exterior circles indicate control region diversity. Green indicates diversity is below average, yellow indicates diversity is average (within 0.2 of the mean), and red indicates diversity is above average. The dashed line illustrates the extent of the last glacial maximum. Numbers identify sample sites: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB. 102

Figure 34(a). Distribution of ND2 region haplotypes, h1 (solid line) and h12 (dashed line).

Illustrated is the distribution of ND2 haplotypes h1 and h12 in sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB. 103

Figure 34(b). Distribution of ND2 region haplotype h15.

Illustrated is the distribution of ND2 haplotype h15 in the sampled populations (it is likely Lake Winnipeg contains this haplotype but it was not identified in the sample: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie;(8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB. 104

Figure 34(c). Distribution of ND2 region haplotypes h82 (solid line) and h26 (dashed line).

Illustrated is the distribution of ND2 haplotypes h82 and h26 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB. 105

Figure 34(d). Distribution of ND2 region haplotypes h110 (solid line) and h7 (dashed line).

Illustrated is the distribution of ND2 haplotypes h110 and h7 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB;(10)Missouri River, MO; (11) Lake Maloney, NB. 106

Figure 35(a). Distribution of control region haplotypes h0005 (solid line) and h0001 (dashed line).

Illustrated is the distribution of control region haplotypes h0005 and h0001 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB. 107

Figure 35(b). Distribution of control region haplotypes h0011 (solid line) and h0638 (dashed line).

Illustrated is the distribution of control region haplotypes h0011 and h0638 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB. 108

Figure 35(c). Distribution of control region haplotypes h0612 (solid line), h0012 (dashed line) and h0633 (yellow line).

Illustrated is the distribution of control region haplotypes h0612, h0012 and h0633 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB. 109

Figure 35(d). Distribution of control region haplotypes h0655 (solid line), h0654 (dashed line) and h0607 (yellow line).

Illustrated is the distribution of control region haplotypes h0655, h0654 and h0607 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB. 110

Figure 35(e). Distribution of control region haplotypes h6133 (solid line), h6137 (dashed line) and h0668 (yellow line).

Illustrated is the distribution of control region haplotypes h6133, h6137 and h0668 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB. 111

Figure 35(f). Distribution of control region haplotypes h0003 (solid line) and h0014 (dashed line).

Illustrated is the distribution of control region haplotypes h0003 and h0014 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB. 112

Figure 36. Suture lines as indicated by control region shared haplotypes.

Longitudinal suture lines as evidenced by the presence (or absence) of shared control region haplotypes among sampled freshwater drum populations. Two approximations were identified (solid line and dotted line. Sample sites are indicated by number. (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB. 114

Figure 37. Haplotypes in the Ohio River freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b). 122

Figure 38. Haplotypes in the Missouri River freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b). 124

Figure 39. Haplotypes in the Lake Maloney freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b). 126

Figure 40. Haplotypes in the Lake Sakakawea freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b). 129

Figure 41. Haplotypes in the Nelson River freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b). 131

Figure 42. Haplotypes in the Lake Winnipeg freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b). 132

Figure 43. Haplotypes in the Lake Pepin freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b). 134

Figure 44. Haplotypes in the Green Bay freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b). 136

Figure 45. Haplotypes in the Lake Erie freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b). 138

Figure 46. Haplotypes in the Lake Ontario freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b). 141

Figure 47. Haplotypes in the Pickwick Reservoir freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b). 144

List of Appendices

Appendix 1. Sample locations.

Numbers were arbitrarily assigned to each sample. 179

Appendix 2. Concentration of DNA eluent for freshwater drum samples.

Concentration of DNA eluent after extraction in ng/µl. Each sample was measured five times using the Qubit fluorometer, the high and low values discarded (marked with an asterisk*), and the average concentration calculated. See Appendix 1 for sample location associated with each sample number. 180

Appendix 3. Design of primers specific to the mitochondrial control region of freshwater drum.

Primer design was dependent on whether the genetic relationship between freshwater drum and red drum (*Sciaenops ocellatus*) was sufficient to amplify the freshwater drum mitochondrial control region using red drum primers. Certain principles had to be applied to ensure specific and efficient amplification. 188

Appendix 4. Control region haplotype frequencies of freshwater drum in sampled populations.

Frequency of haplotypes in sampled populations of freshwater drum. Column 1 refers to the name arbitrarily assigned to a particular haplotype. All other columns refer to the frequency of each haplotype in a population. 190

Appendix 5. ND2 region haplotype frequencies of freshwater drum in sampled populations.

Frequency of haplotypes in sampled populations of freshwater drum. Column 1 refers to the name arbitrarily assigned to a particular haplotype. All other columns refer to the frequency of each haplotype in a population. 193

INTRODUCTION

Population genetics is the study of the genetic composition of populations and how allelic frequencies change over time whereas phylogeography examines genetically related populations to determine the origins and relationships of taxa with respect to geographic distribution (Avise, 2000; Zink, 2002; Althukov, 2006; Brito and Edwards, 2009). The genetic composition of populations can be used to determine the phylogeographic makeup of populations. Intraspecies relationships are those most commonly studied and can be used to interpret the geographic structure of a population, both historically and in present day (Richards et al., 2007). All genes, haplotypes, and taxa share a common ancestor. Relationships among taxa are determined by the events which transpired in the process of diversification (Templeton, 1998). Data can be utilized to develop and test hypotheses of the evolutionary processes or events which produced taxon structure (Zink, 2002; Brito and Edwards, 2009). Diversification involves both spatial and temporal factors, genetic variation increases over both distance and time (Zink, 2002).

Taxa occupy habitats across the range in which they are distributed. The total population of a taxon is composed of subpopulations that adapt to different local conditions and acquire genetic change. Phylogeographic studies analyze genetic variation and correlations with geographic distribution (Brito and Edwards, 2009). Studies of freshwater fishes are limited to waterways in which the species exist, in my study of freshwater drum I analyzed freshwater regions of central and eastern North America.

Three major drainage basins exist in central North America, the Mississippi River, Great Lakes/St. Lawrence and Nelson River/Hudson Bay drainage systems. Glaciation has affected landscape and hydrography, altering drainage systems while destroying and creating waterways

(Hewitt, 1996). The distribution of species was altered by glaciation, many species were forced south of the ice sheet and into refugia in order to survive. Post glacial effects created new habitats, allowing for range expansion of many species. Fishes with great mobility and wide environmental tolerances were better able to colonize new habitats and increase territory (Bernatchez and Wilson, 1998).

Freshwater drum is ideal for this study as it is the only North American freshwater species of the family Sciaenidae and has a wide latitudinal distribution range (Stewart and Watkinson, 2004). Mitochondrial DNA (mtDNA) is the most common type of data used for animal phylogeographic animal studies, providing information about present and historical demography and processes (Avise, 2000). Glaciation affected taxa by destroying habitats, decreasing population size and segregating population members, this resulted in genetic variation (Hewitt, 1996).

Population Genetics

The genetic makeup of a diploid population is derived from its gene pool, where zygotes will become the breeding population of the next generation. The allelic frequency within the gene pool governs the genetic composition of future generations (Altukhov, 2006; Freeman and Herron, 2007).

The study of population genetics emerged in the 1920's and 1930's based upon the ideas of R. Fisher, J.B.S. Haldane and S. Wright. Their studies united the theories of Darwinism, which supported the idea of gradual evolution, and Mendelian genetics, which supported the idea of major phenotypic change based upon a single mutation. Supporters of Neo-Darwinism and Mendelian inheritance challenged traditional Darwinism. Darwin could not account for the

source of inheritance, for selection to occur gradually over time a continual supply of variation would be necessary. Neo-Darwinists believed that over time genetic variation would disappear, making a population homogeneous and therefore impossible for natural selection to occur. In 1918, Fisher showed that if a continuous trait such as height was under the influence of many Mendelian "factors" (which are now known to be genes), there would be an approximately normal distribution. Models were developed which demonstrated that the genetic composition of a Mendelian population could be altered over time. This enhanced knowledge by providing a mathematical framework that allowed for quantitative analysis of evolutionary hypotheses. This was important in establishing the field of population genetics (Okasha, 2008).

Populations are dynamic; the frequencies of alleles contributing to the gene pool can be altered by factors including mutation, gene flow, natural selection and genetic drift (Klug and Cummings, 2005). Mutation or genetic recombination can influence a population as a whole, yet populations also affect individuals. Reproductive success of individuals with a particular genotype can be dependent on population size, the frequency of the gene within the population, and the genotypes of others within the population (Freeman and Herron, 2007).

Population genetics applies the concepts of Mendelian genetics to a population (Mettler et al., 1988). In my study I analyzed the degree of genetic differentiation to determine relationships between and amongst populations of freshwater drum. Greater genetic variation is observed amongst populations that are not closely related, often a result of isolation over space or time. Understanding the process of genetic recombination, and the forces which drive change (Figure 1), is fundamental to the field of population genetics (Klug and Cummings, 2005).

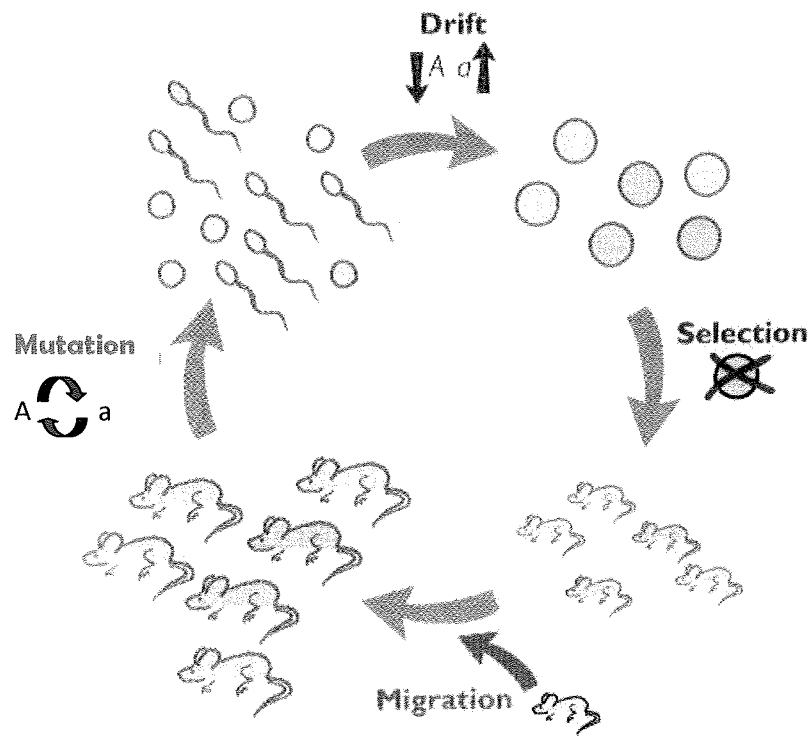


Figure 1: Forces of evolution. Mutation alters an allele, and is the source of all genetic change. Genetic drift affects the distribution of alleles in a population by random chance, it is not adaptive. Selection rids the population of deleterious alleles and preserves beneficial alleles. Migration introduces alleles to a population, altering allelic frequency. These forces, among others, individually or working in concert, affect the evolution of a population.

Source: Freeman, S. & Herron, J.C. (2007). Evolutionary Analysis, Fourth Edition. Pearson Education, Upper Saddle River, NJ. p. 181. Copyright © 2007 by John Wiley & Sons, Inc. Reprinted by permission of John Wiley & Sons, Inc.

Hardy-Weinberg equilibrium illustrates allelic and genotypic proportions in an idealized population. Hardy-Weinberg acts under five assumptions: there is no mutation, no gene flow, no selection, no genetic drift, and mating is random. Under these conditions two properties are observed; allele frequencies are stable over time and genotype frequencies can be predicted after a single generation of random mating. This is the state known as Hardy-Weinberg equilibrium (Klug and Cummings, 2005).

Violation of any of the assumptions of Hardy-Weinberg equilibrium results in evolution. Hardy-Weinberg equilibrium describes a situation in which a population does not evolve, illustrating a state where genetic variability is maintained and dominant traits do not increase over time. Hardy-Weinberg equilibrium is of great importance to population genetics, identifying the forces which cause evolution (Klug and Cummings, 2005; Okasha, 2008). Hardy-Weinberg equilibrium serves as a null model against which evolutionary mechanisms may be identified (Freeman and Herron, 2007). In my study I have been able to infer the cause of altered genotypic frequencies. Statistical analysis allowed me to identify if genetic change resulted from mutation, gene flow, selection or genetic drift.

Mutation is the fundamental source of genetic variation; it is the only manner in which new alleles arise in a species. Mutation is random, varying both in rate and persistence. Mutation rate and stability is influenced by the effect a mutation may have upon an organism. Deleterious mutations may be eliminated by selection, neutral mutations are fixed or lost by genetic drift, and advantageous mutations are swept to fixation by selection in the absence of environmental stochasticity (Freeman and Herron, 2007). Genomes are composed of coding and non-coding regions. Coding regions translate into protein sequences and display lower rates of mutation as function is more likely to be impaired. Non-coding regions have higher

rates of mutation as effects to the organism are minimal or non-existent. Mutation coupled with other forces can alter allelic frequencies within a population (Klug and Cummings, 2005).

Migration in the context of population genetics is gene flow; the introduction of new alleles into a population by individuals which originate from a different geographic area. Migration homogenizes allele frequencies across populations, selection maintains variation among subpopulations. Regions that have been able to maintain gene flow over time, as has occurred in many non-glaciated areas, display less genetic division among populations. Migration has a greater effect on small populations as the gene pool is smaller, and may reduce or eliminate adaptive differences resulting from natural selection. Migration hinders divergence of populations and can counter selection by maintaining the presence of an allele that is opposed by selection in a local population. An example of this is seen in a study of Lake Erie water snakes by King and Lawson (1993). Snakes with banded markings are more common on the mainland; unbanded snakes are more common on the islands. Banded snakes are selected against on the island as they are more visible, and more vulnerable to predation. Migration from the mainland maintains the banded allele while selection on the islands opposes it (Klug and Cummings, 2005; Freeman and Herron, 2007).

Natural selection can result in differential survival and reproductive rates in individuals with genotypes of higher fitness. This can be very influential in altering the frequency of alleles. Subpopulations persist in habitats to which they can adapt. A population existing in the Nelson River of northern Manitoba will likely display different adaptations than a population in Alabama, as there will be differences in temperature, food source and other conditions. Genotypes which confer higher fitness increase the likelihood of individual survival or reproduction, individuals with greater fitness contribute more gametes to the gene pool.

Natural selection results in evolution by increasing the mean fitness of the population. Selection occurs in many forms, natural selection can be negative, removing deleterious mutations, or can be positive, acting on beneficial mutations and increasing their frequency (Klug and Cummings, 2005; Freeman and Herron, 2007). Natural selection and genetic drift alter allelic frequencies. Natural selection is adaptive, genetic drift is not.

Genetic drift is a random force which alters allele frequencies, driving them towards fixation or loss regardless of fitness. It is change that is not associated with selection. Genetic drift does not cause adaptation. The influence of drift is greatest in small populations where the distribution of alleles is more likely to be affected by a single mutation (Freeman and Herron, 2007). Drift in a newly colonized population that has been established by a small number of individuals is referred to as founder effect. This same effect can arise by genetic bottleneck when a large population greatly decreases in number temporarily, decreasing genetic diversity (Klug and Cummings, 2005). Episodes of genetic bottlenecks and founder effect occurred throughout periods of glaciation. Advances of ice resulted in habitat loss and reduced population size, retreats of ice allowed colonization of new habitats by small numbers of individuals (Hewitt, 2000). Genetic drift decreases genetic variation, genetic diversity is introduced and maintained by mutation in existing members of a population, or via gene flow.

Population genetics examines each of these evolutionary forces and their interaction with each other and the population(s). The form of genetic differentiation, coupled with biogeography, can help to determine if a population has been subject to the influence of mutation, gene flow, selection or genetic drift (Avise, 2000). As revealed in this and other studies, populations in previously glaciated regions are more likely to have been affected by genetic drift, whereas populations that have persisted over time will show increased genetic variation resulting from

gene flow (Robison, 1986; Hewitt, 1996, 2000). Genetic diversity among populations is a direct result of these factors. These basic elements have formed the field of population genetics which in turn has given rise to other fields of study including phylogeography.

Phylogeography

Phylogeography (Avise, 2000) involves the study of genetically related populations, focusing on the origins and relationships of taxa in the context of geographic distribution, primarily at the intraspecific level (Zink, 2002; Brito and Edwards, 2009). Phylogeography can be used to analyze the origin and dispersal of subpopulations. This allows for the inference of historical demographic processes which contributed to the past and present demography of a population (Avise et al., 1987; Zink, 2002). A multitude of factors affect the geographic distribution of taxa; the processes contributing to evolution and creating population structure vary greatly. Environmental barriers, including glaciation, may prevent dispersal, while ecological barriers such as behaviour and physiology may prevent mating. Taxa affected by these barriers are subject to effects of genetic drift and selection that may result in both genetic and spatial separation (Avise, 1998). This, and previous studies, have shown that populations existing in previously glaciated regions display decreased genetic variation, often resulting from genetic bottlenecks and drift (Mandrak and Crossman, 1992; Brown et al., 1993; Wilson et al., 1996; Borden and Krebs, 2009). Determining temporal and spatial correlations of a population is not as simple as superimposing a phylogeny onto a map, analytical methods must be employed to ensure these inferences are valid.

Mitochondrial DNA data dominate the study of animal phylogeography. Increasingly other forms of molecular data, such as nuclear DNA (nDNA), are employed. Molecular data are

analyzed for polymorphisms, each unique sequence forming a haplotype. Haplotypes are then used to build a phylogenetic network or tree. Various methods, including parsimony, maximum likelihood, coalescence and Bayesian analysis are available to further interpret the tree as it relates to geographic distributions and to infer historical patterns and the processes which shaped them (Emerson and Hewitt, 2005; Morrone, 2009).

Various group structures can be analyzed using genetic data such as mitochondrial DNA. An example of group structures is a clade, a population which shares the same, or similar, haplotypes (Templeton, 1998). Populations and clades, as well as the relationships between them, can be represented by a phylogenetic tree. The length of a branch on a phylogenetic tree is correlated with the number of genetic changes between lineages (Figure 2) which can be calibrated to time (Avise et al., 1987). It has been hypothesized that mitochondrial DNA mutates at a fairly constant rate, thus can be used as a molecular clock with which to calibrate time. However, differences in mutation rate may make the molecular clock an unreliable source of information for dating evolutionary events (Gibbons, 1998). Inferences based on the structure of the phylogenetic tree are more likely to be reliable. Large gaps between branches or groups of branches are a consequence of large genetic differences which normally result from long term barriers to gene flow. High levels of gene flow over time are represented by trees, or sections of trees, where branches have minimal gapping (Avise et al., 1987). Gene flow maintains similarity and homogeneity between populations, lack of gene flow is the primary factor in diversification of taxa and the evolution of new species.

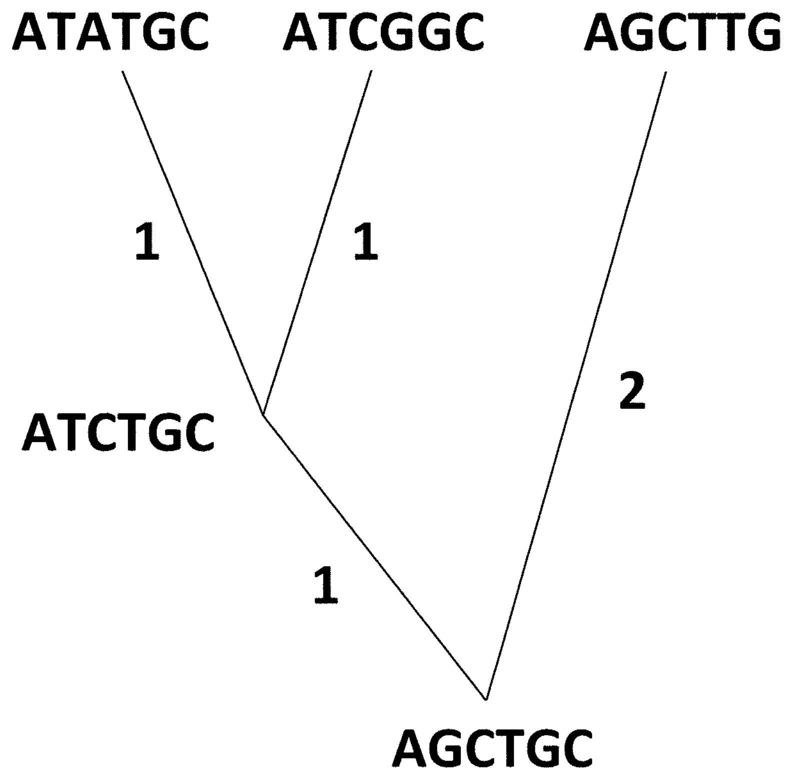


Figure 2. A phylogenetic tree.

A phylogenetic tree, the length of each branch is correlated with the amount of genetic change.

Diversification of taxa occurs when gene flow is minimal or non-existent and directional selection occurs. The absence of gene flow may ultimately result in speciation as distance and time create genetic, behavioural and morphological differences (Zink, 2002). Intraspecific phylogeography examines the evolutionary history of a single species via analysis of genetic variation (Posada and Crandall, 2001; Templeton, 2004, 2009).

Genetic analysis varies depending on the focus of research. The focus of a study determines what types of molecular data are used. Interspecies studies require the use of slowly evolving sequences, such as the 12S or 16S rRNA gene in mitochondrial DNA, as divergence occurred long ago, whereas intraspecies studies examine more recent history (in 10s of thousands of years), thereby requiring analysis of rapidly evolving regions of DNA, including the control and ND2 regions (Lopez et al. 1997, Broughton and Reneau, 2006). The evolutionary history of any organism is based on the sum of multiple genealogies of alleles; each region varies in mutation rate and the longevity of mutations. Thus it is important to analyze data from multiple, unlinked loci to attain a complete and reliable evolutionary history (Emerson and Hewitt, 2005).

Mitochondrial DNA

Mitochondrial DNA is a closed, circular, cytoplasmic genome approximately 16-20 kb in length that encodes 37 genes represented by 22 transfer RNAs, 2 ribosomal RNAs and 13 messenger RNAs (Figure 3) (Avise et al., 1987; Avise, 2006). The mitochondrial genome in other Sciaenid species is approximately 16500 base pairs in length (Cheng et al., 2011). From a phylogenetic perspective the mitochondrial DNA genome represents a single locus as alleles are genetically linked through asexual transmission, thus the entire genome is transmitted as though a single locus. Animal mitochondrial DNA evolves faster than nuclear DNA and exhibits great heterogeneity within and between species. Most variations in mitochondrial DNA involve

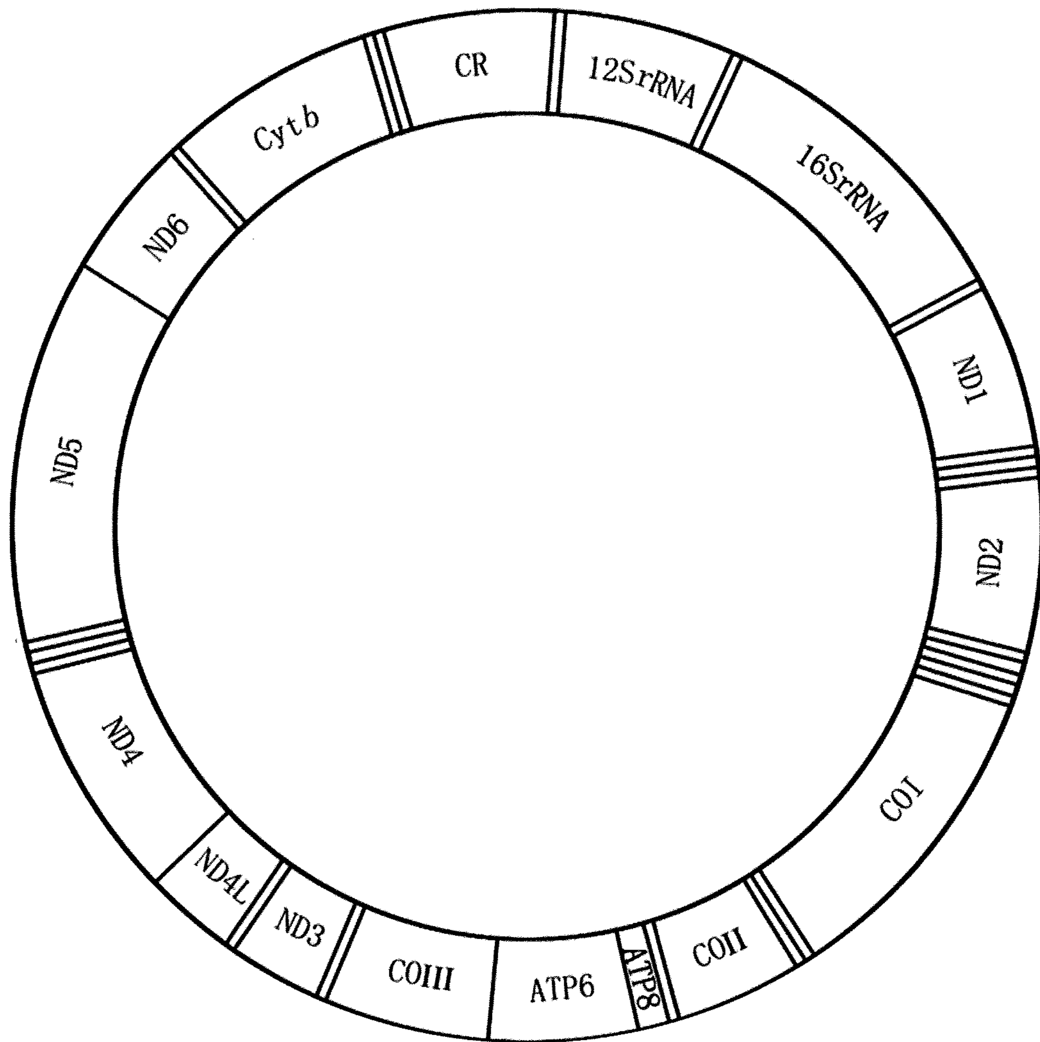


Figure 3. The mitochondrial genome. The mitochondrial genome is a circular genome encoding for 37 genes and is commonly used in phylogeographic studies.

Source: Cui, Z., Y. Liu, C.P. Li, F. You, K.H. Chu. 2009. The complete mitochondrial genome of the large yellow croaker, *Larimichthys crocea* (Perciformes, Sciaenidae): Unusual features of its control region and the phylogenetic position of the Sciaenidae. Gene, **432:** 33-43. © 2008 Canadian Science Publishing or its licensors. Reproduced with permission.

nucleotide substitutions or small changes in length; gene order remains relatively stable (Avise, 2000).

Animal mitochondrial DNA has several features making it appropriate for molecular analysis of intraspecific phylogenetic relationships and is used extensively in evolutionary studies (Villablanca, 1994). These features include its small size, stable organization and above all, high nucleotide substitution rate. In general, mitochondrial DNA is maternally inherited resulting in lack of recombination (Cantatore et al., 1994). Replication of mitochondrial DNA is partially independent of the nucleus. Certain mitochondrial DNA will replicate more than others either by chance or selective advantage, genomes near the nucleus have been shown to exhibit increased replication rates (Barr et al., 2005; White et al., 2008). Sequences from the same taxonomic family can be aligned and polymorphisms identified, making mitochondrial DNA appropriate for phylogenetic analysis. Non-coding regions of mitochondrial DNA lack the structure of coding genes; substitutions and indels (insertions or deletions) are more likely to be tolerated (Lee et al., 1995).

Mitochondrial DNA evolves rapidly and is ideal to differentiate between closely related organisms (Li and Graur, 1991). Mitochondrial DNA mutations are common, often reaching fixation quickly. The high mutation rate of mitochondrial DNA is likely due to inefficiency of DNA repair mechanisms, high exposure to free radicals in the oxidative mitochondrial environment and rapid replicative turnover (Avise, 2000). The rate of mutation in mitochondrial DNA is approximately ten times that of nuclear genes, making mitochondrial DNA ideal for intraspecies studies (Kocher and White, 1989). Animal mitochondrial DNA does not code for proteins directly involved in replication or transcription, thereby evolves rapidly as it tolerates less accuracy in translation (Avise, 2000).

Mitochondrial DNA inheritance is asexual and almost exclusively maternal, normally occurring without recombination (Avise, 2000). The maternal history of mutations is recorded in an organized manner, without paternal influence and/or recombination. Maternal inheritance via mitochondrial DNA is more sensitive in detecting population subdivisions than nuclear genes (Lee et al., 1995). Transmission of paternal mitochondrial DNA is unlikely to persist into later developmental stages (Barr et al., 2005). When paternal leakage occurs it is normally low level and transient, in part because of mechanisms which degrade paternal mitochondrial DNA, with the exception of bivalve species including *Mytilus* which regularly inherit mitochondrial DNA from both parents (Zouros et al., 1994; Avise, 2000).

Two regions of mitochondrial DNA were targeted in this study, the ND2 and the control regions. The ND2 region in Sciaenids is approximately 1050 base pairs in length, and the control region is approximately 820 base pairs (Cheng et al., 2011). The targeted area of the ND2 region in my study is 500 base pairs, and 400 base pairs for the control region, comprising approximately 5.5% of the mitochondrial genome, which should be similar in size to that of yellow drum (Cheng et al., 2011). These are both highly variable regions of the mitochondrial DNA molecule and suited to uncovering intraspecies nucleotide differences. The intragenic region of the ND2 gene contains non-coding sections and has a high rate of mutation also making it suitable for intraspecific phylogenetic studies, as evidenced by the study by Koblmüller et al. 2010 of the cichlid tribe Tropheini. The control region is a non-coding region, and has the highest rate of evolution of the mitochondrial genome due to reduced functional constraints (Villablanca, 1994). Hypervariable regions 1 and 2 have high rates of mutation and substitutions accumulate rapidly, making these regions ideal for phylogeographic studies requiring fine scale resolution which identifies recent change (i.e. those which have occurred since the last glacial maximum). The lower mutation rate of the ND2 region makes it ideal for identifying change which occurred

change over a longer length of time, in my study it is used to identify change occurring prior to the last glacial maximum. The high nucleotide substitution rate of these regions is well suited to act as a genetic marker in population genetic studies (Brown et al., 1993).

Several forms of mutation exist including transitions, transversions and indels. Mutations are passed onto offspring. Substitution of a purine (adenine or guanine) for another purine or substitution of a pyrimidine (cytosine or thymine) for another pyrimidine are transitions. Substitution of a purine for a pyrimidine or vice versa is a transversion. Transitions occur more frequently as they cause less disruption and are better tolerated. Base substitutions are the most common mutation, resulting from mistakes in DNA synthesis where an incorrect base is incorporated into DNA. Mutations may be silent causing no change, or be missense mutations which lead to an altered phenotype (Nester et al., 2007). Silent mutations occur at a much faster rate, and are generally neutral or mildly deleterious (Villablanca, 1994). Mutations affecting the first or second position of a codon almost always change the amino acid specified by mRNA, redundancy in the genetic code causes mutation in the third position to often produce no change at all. Indels are another form of mutation where bases are inserted or deleted into the genome (Cantatore et al., 1994). Most mutations in the rapidly evolving unconstrained regions of mitochondrial DNA are neutral and will be preserved (Villablanca, 1994).

Mitochondrial DNA exists in higher copy numbers than nuclear genes, facilitating amplification of mitochondrial DNA sequences. This results in a consensus sequence being derived from a greater template number, avoiding errors resulting from template degradation (Villablanca, 1994). The small size and stability of the mitochondrial genome, along with elevated mutation

rate, lack of recombination, and high copy number make mitochondrial DNA suitable for phylogeographic studies.

Aplodinotus grunniens Rafinesque (freshwater drum)

Aplodinotus grunniens (Rafinesque, 1819) (freshwater drum) is the sole North American freshwater member of the family Sciaenidae. Freshwater drum has the greatest latitudinal distribution of any North American freshwater fish (Figure 4). Historically this species has been recorded from the mouth of the Nelson River on Hudson Bay in Manitoba to the Rio Usumacinta Basin in Guatemala (Burr and Mayden, 1992). However the species has not been seen in the southern portion of its range in several years, its present distribution reaches the southern United States (Berra, 2001). Freshwater drum is found in all Great Lakes with the exception of Lake Superior. It prefers large and shallow waters, and occupies lakes and rivers in both clear and turbid water (Coad, 1995). Specimens inhabiting northern areas exhibit decreased growth rate and increased life span (Stewart and Watkinson, 2004).

Spawning of freshwater drum is not well studied but is thought to occur at water temperatures of 20° to 23°C in July through September (Coad, 1995). Spawning occurs in midwater and eggs float to the surface (Stewart and Watkinson, 2004). This is one of few North American freshwater fish with planktonic eggs, a trait common in marine fishes including all marine members of the family Sciaenidae (Berra, 2001). Sexual dimorphism is apparent, females grow to larger sizes than their male counterparts. Rypel (2004) determined that females of the species were highly mobile, while males were among the most sedentary of those studied. Planktonic eggs, coupled with the high mobility of females may explain the vast species dispersal (Berra, 2001; Rypel, 2007).

Figure 4. Distribution range of freshwater drum.

Present day distribution of freshwater drum. The dispersal range stretches from the mouth of the Nelson River at Hudson Bay to the southern United States. The species exists in all the Great Lakes, with the exception of Lake Superior.

Source: Dr. Joseph Carney, Lakehead University, Thunder Bay, ON.

Freshwater drum feed on immature insects, crayfish, minnows and molluscs. The majority of drums are bottom feeders. Jaws have a section of small villiform teeth, pharyngeal bones at the top and bottom of the throat are large and contain molariform teeth that can crush shelled animals such as clams and snails (Tomelleri and Eberle, 1990; Coad, 1995).

Glacial Geography of North America

The topography of North America has been altered over time. Glaciation has been a predominant force in this change (Hewitt, 1996). The North American landmass was greatly transformed as a result of the last glacial maximum, the ice sheets of which began to recede approximately 18000 years ago (Figure 5).

There have been a number of glacial events, the most recent to affect North America was the Wisconsin glaciation formed by the Laurentide, Cordilleran and Arctic ice sheets (Bernatchez and Wilson, 1998; Trenhaile, 1998). The Laurentide ice sheet covered areas down to 40° north latitude in central and eastern North America (Figure 5) (Hewitt, 1996, 2000). Glacial advances and retreat altered the topography of North America, destroying and forming freshwater habitats. Fishes survived in refugia south of the ice sheet. Cyclical advances and retreat of ice allowed some species of fishes to disperse, while others were extirpated (Hewitt, 1996). Glaciation destroyed drainage systems and restructured habitats (Figure 6) (Bernatchez and Wilson, 1998). The ice age resulted in the extinction of populations while others survived in existing refugia. Populations inhabiting separate refugia diverged to adapt to distinct habitats and various races (or sub-species) resulted from temporal and spatial isolation (Hewitt, 1996, 2000).

Figure 5. North America in the last glacial maximum. Wisconsin glaciation in North America, the greatest extent of the ice sheet dating approximately 18000 years ago. The Great Lakes and Lake Winnipeg, although shown, did not exist.

Source: http://www.isgs.illinois.edu/maps-data-pub/publications/geonotes/geonote3.shtml

Reproduced with permission of ISGS.

Drainage systems of North America were vastly different prior to the last glacial maximum (Figure 7). During the Pliocene, more than 2.5 million years ago, the major river basin of the central United States was the Teays-Mahomet river system which flowed to the Mississippi River into the Gulf of Mexico (Figure 8) (Hocutt et al., 1986; Berendzen et al., 2003). The Teays River originated in the Blue Ridge Mountains of North Carolina, flowing northward to central Ohio, then west to central Indiana and Illinois through the Mahomet valley to the Mississippi River. The Teays River is hypothesized to have connected to the Erigan River in the Lake Erie basin and was a major connection for fauna between Atlantic drainages and the old Ohio River (Burr and Page, 1986; Robison, 1986). Glacial advances in the pre-Illinoian ice age of the Pleistocene diverted flow from the Mahomet valley to the Wabash valley. The exact timeline of these events is uncertain as many of the river valleys were altered by subsequent glacial advances, but likely occurred between 2.5 and 1.8 million years ago (Burr and Page, 1986). A region near Chillicothe, Ohio was dammed, creating Lake Tight which is hypothesized to have existed more than 700,000 years ago (Goldthwait, 1991). Lake Tight eventually breached, spilling water into the Licking River valley (Figure 9). Over time, water was diverted to the old Kentucky River and the old lower Ohio River (Burr and Page, 1986; Hocutt et al., 1986). The old Ohio River and part of the Pittsburgh River, including its tributaries the upper Allegheny and Monongahela Rivers, were also important drainage systems in the central United States which were altered in the Illinoian ice age (between 300,000 and 130,000 years ago) to form the upper Ohio River (Burr and Page, 1986; Hocutt et al., 1986; Robison, 1986).

The primary drainage system of the western central United States in the Pliocene, over 2.5 million years ago, was the old Missouri River. During the pre-Illinoian stage of the Pleistocene,

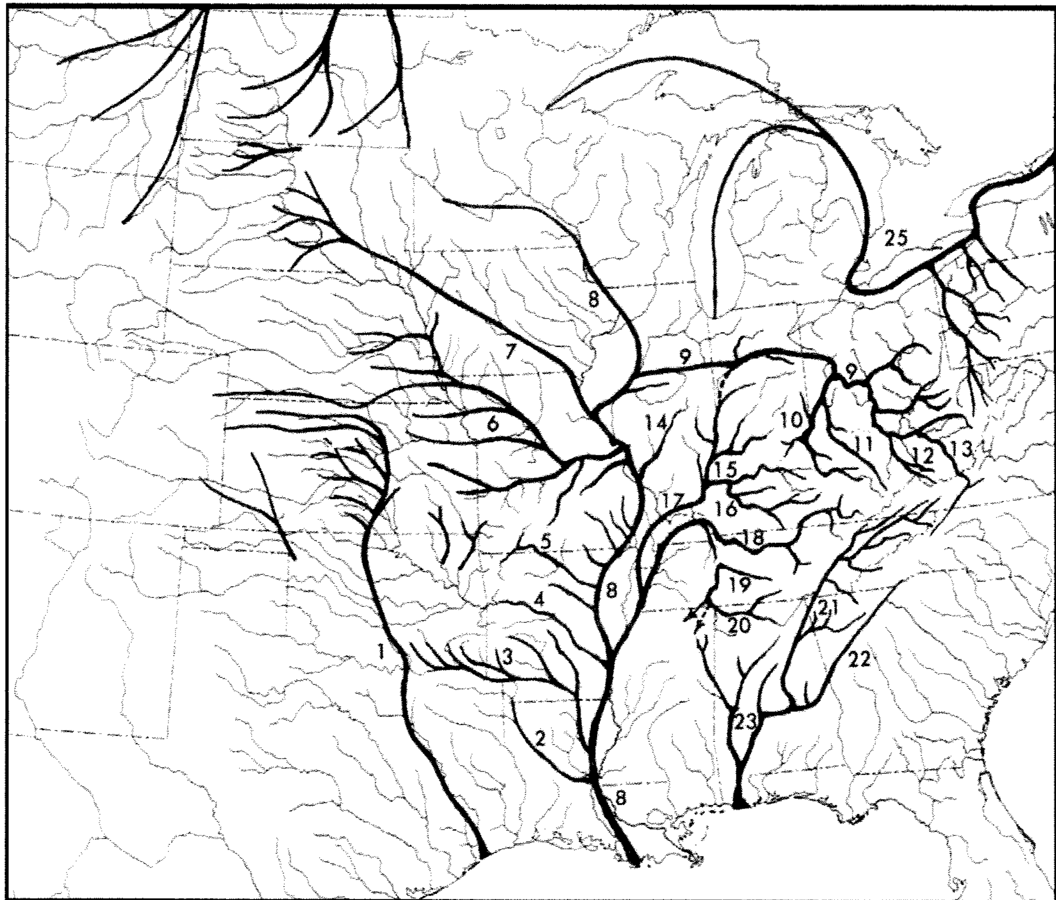
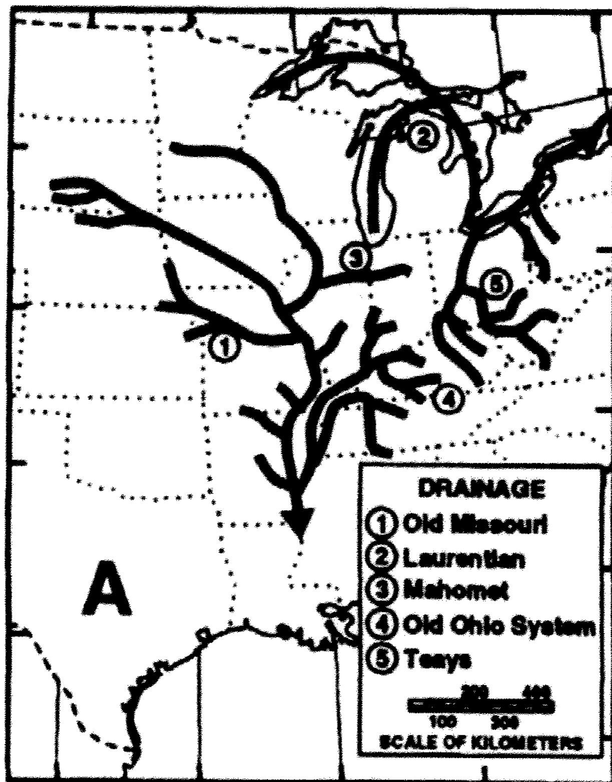
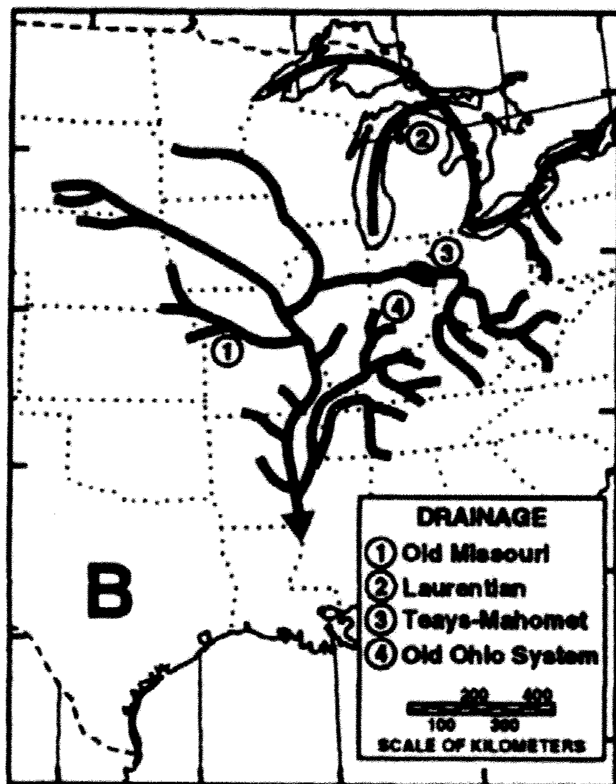

25
8
9
7
9
14
10
6
11
12
13
15
17
16
5
18
8
19
4
21
20
1
3
22
23
2
8

Figure 6. Preglacial drainage systems of North America in the Pliocene.

Pliocene drainage patterns (approximately 2.5 million years ago) are superimposed in dark lines over modern day drainage systems. Drainage systems noted in this study are in **bold**. (1) Plains Stream; (2) Old Red River; (3) Old Ouachita River; (4) Old Arkansas River; (5) White River; **(6) Old Grand-Missouri River**; (7) Ancestral Iowa River; **(8) Old Mississippi River**; **(9) Old Teays-Mahomet River**; (10) Old Kentucky River; **(11) Old Licking River**; (12) Old Big Sandy River; (13) Kanawha River; (14) Kaskasia River; **(15) Wabash River**; (16) Green River; **(17) Old Ohio River**; **(18) Old Cumberland River**; (19) Old Duck River; **(20) Old Tennessee River**; (21) Appalachian River; (22) Old Tallapoosa River; (23) Mobile Basin; **(24) Hudson Bay drainage**; **(25) St. Lawrence River**.

Source: Mayden, R.L. 1988. Vicariance biogeography, parsimony and evolution in North American freshwater fishes. Systematic Zoology, **37:** 329-355. Reproduced with permission.


A
DRAINAGE
1 Old Missouri
2 Laurentian
3 Mahomet
4 Old Ohio System
5 Teays
SCALE OF KILOMETERS


B
DRAINAGE
1 Old Missouri
2 Laurentian
3 Teays-Mahomet
4 Old Ohio System
SCALE OF KILOMETERS

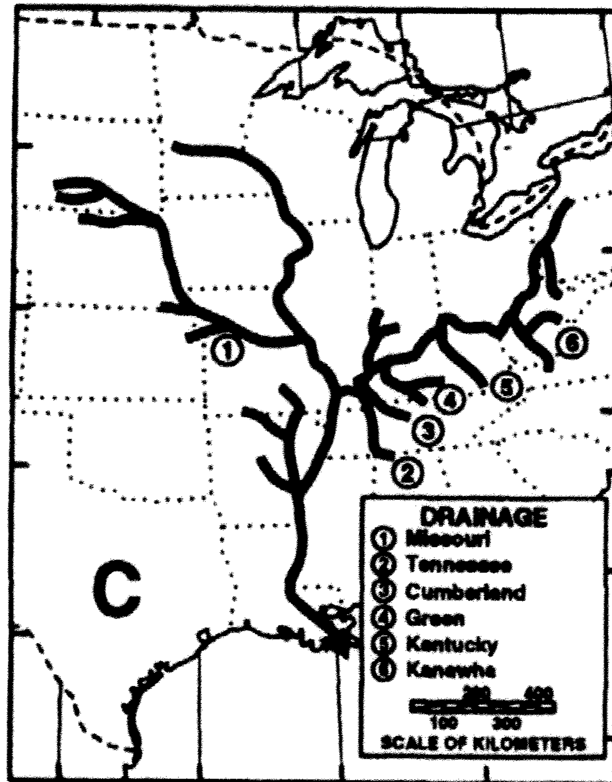


Figure 7. Major North American drainage systems in the Tertiary.

Figure A displays the major drainage systems in the Pliocene. Figure B illustrates major drainage systems in the early Pleistocene. Figure C displays drainage systems in the Holocene. The southern course of the Mississippi River is indicated by the arrow.

Source: Strange, R.M. and B.M. Burr. 1997. Intraspecific phylogeography of North American highland fishes: a test of the Pleistocene vicariance hypothesis. Evolution, **51:** 885-897. © 2008 Canadian Science Publishing or its licensors. Reproduced with permission.

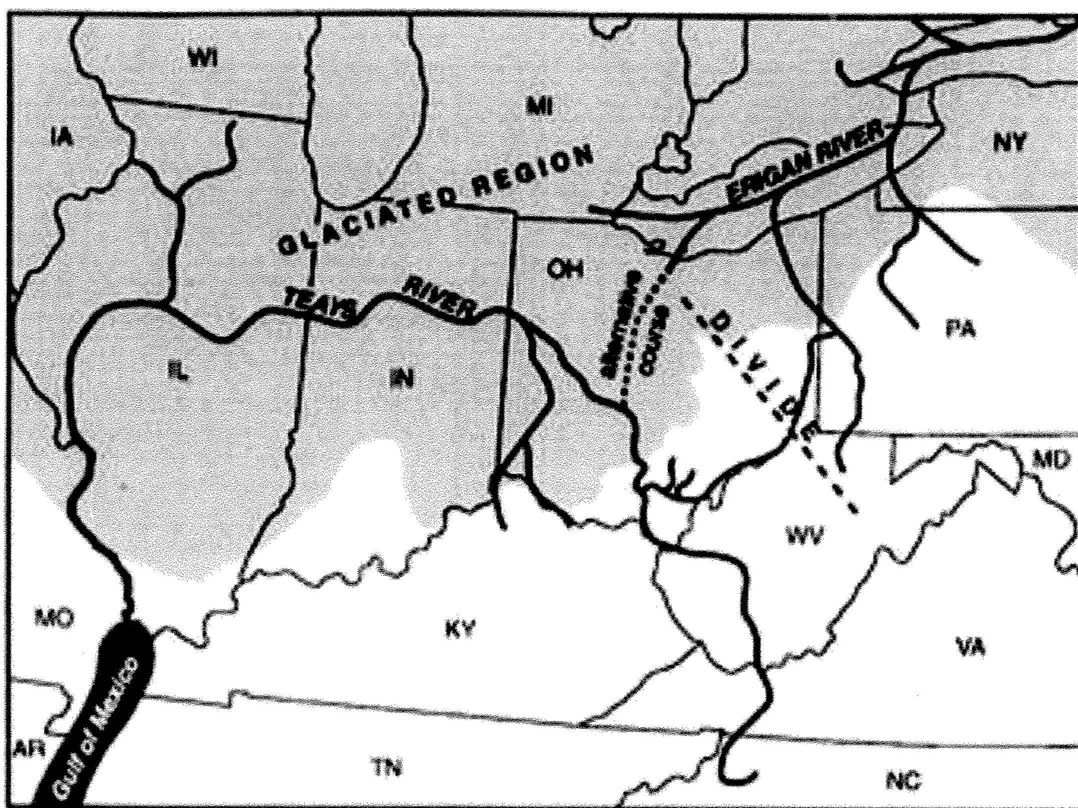


Figure 8. The ancient Teays River.

Drainage systems differed prior to the Wisconsin glaciation. The Teays River drained much of the eastern-central United States into the Gulf of Mexico. The Erigan River, hypothesized to drain into the Teays River, existed in what is now the Lake Erie basin and drained into the St. Lawrence. These systems were altered by the advance and retreat of the ice sheet, the drainage system of the Teays River being altered to become the present day Ohio River.

Source: http://www.mjcpl.org/rivertorail/images/44.jpg

Reproduced with permission of ODNR.

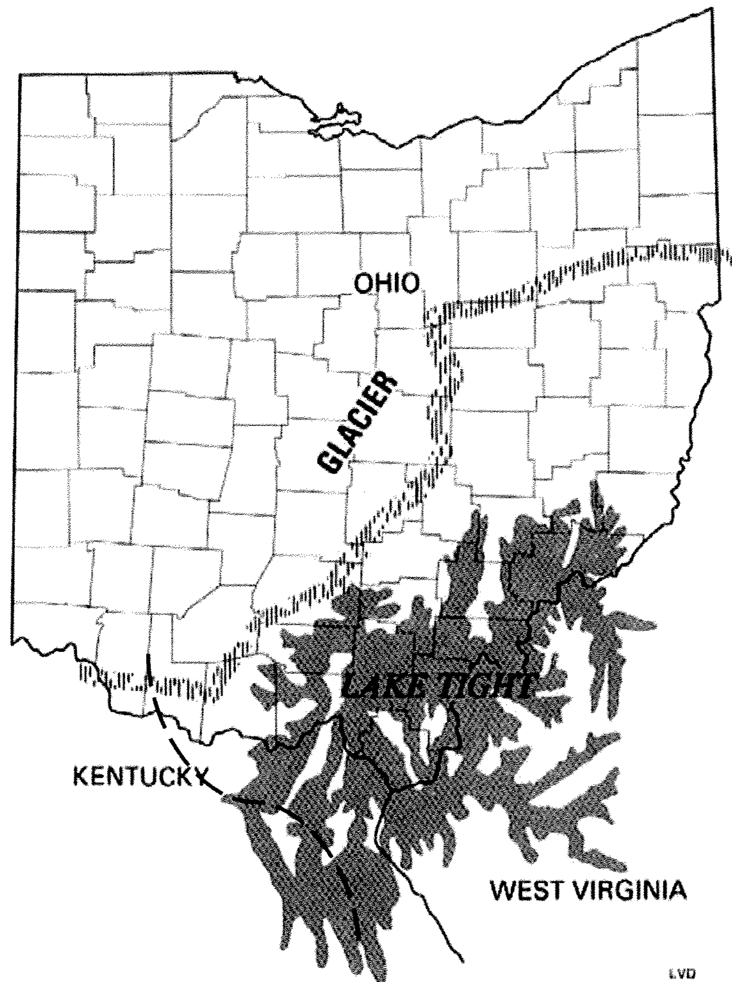


Figure 9. Glacial Lake Tight.

Glacial Lake Tight was formed by damming of the Teays River in Ohio. It is hypothesized that water spilled into the Licking River valley (dashed line) allowing for dispersal of freshwater fishes into the old Kentucky River and old Lower Ohio River.

Source : http://www.ohiodnr.com/Portals/10/pdf/newsletter/Summer87.pdf

Reproduced with permission of ODNR.

at least 800,000 years ago, the edge of the Laurentide ice sheet formed the present day Missouri River. Northern portions of the present day Missouri River were glaciated throughout the Pleistocene (Crossman and McAllister, 1986).

The Mississippi drainage system served as the primary refugium for freshwater fishes during glaciation. The Interior Highlands, including the Ozarks and Ouachita Mountains, also played a significant role as glacial refugia (Borden and Krebs, 2009). Missourian refugia, including the Niobrara River valley, existed in the western central United States (Figure 10) (Crossman and McAllister, 1986; Kaul, 1988).

Periglacial freshwater lakes also provided refugia for species and eventually connected to refugia south of the ice sheet (Wilson et al., 1996). G.K. Warren (1874) hypothesized that Wisconsin harboured such a refugium during the Pleistocene (Figure 11). The Green Bay lobe of the Laurentide ice sheet met the Baraboo Hills to create an ice dam which resulted in the formation of glacial Lake Wisconsin approximately 1.8 million years ago. This dam likely broke and reformed several times throughout the Pleistocene, connecting glacial Lake Wisconsin to the Wisconsin River, a tributary of the Mississippi River (Clayton and Knox, 2008).

The southern border of the Laurentide ice sheet is characterized by lobes created by glacial advances. These lobes dammed rivers and formed large proglacial lakes (Hewitt, 1996, 2000; Siegert, 2001). Major deglaciation occurred between 15000 and 8000 years ago (Bernatchez and Wilson, 1998; Trenhaile, 1998). As ice retreated, glacial lakes in the Great Lakes basins shifted (Figure 12) (Ehlers and Gibbard, 2004). The Great Lakes basin was exposed at the southern edge of the ice margin approximately 15000 years ago. Between 14000 and 10600 years ago glacial Lake Chicago began to form in the basin of Lake Michigan, draining into the

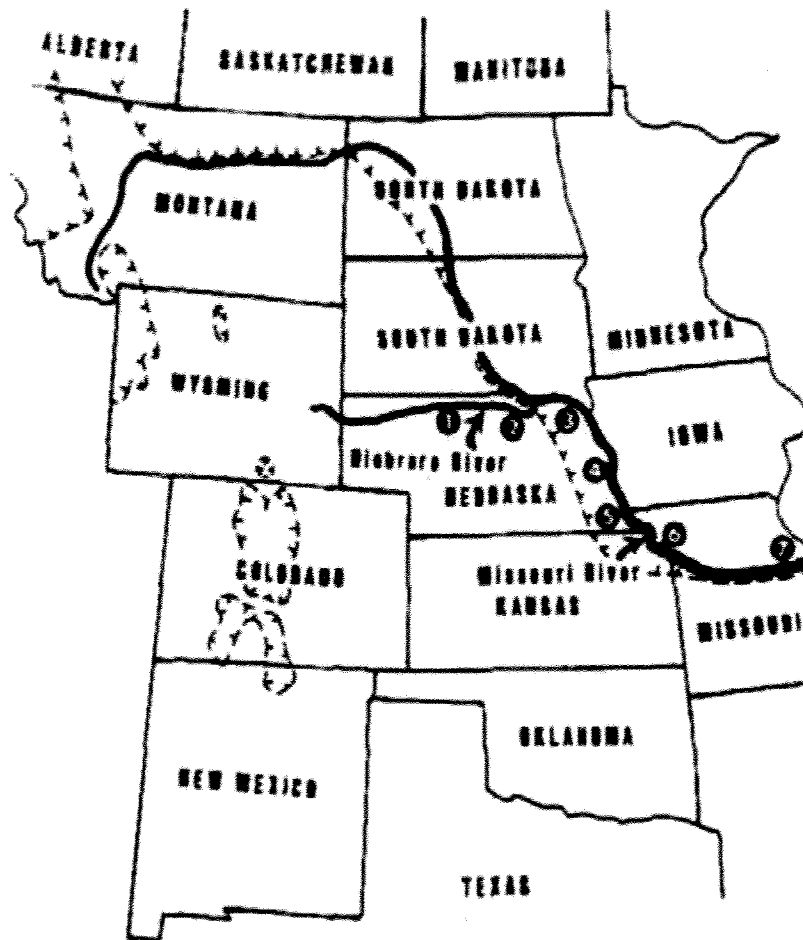


Figure 10. The Niobrara River Valley.

The Niobrara River was a glacial refugium in northern Nebraska which displays great diversity of plant and animal species.

Source: Kaul, R.B., G.E. Kantak and S.P. Churchill. 1988. The Niobrara River valley, a postglacial migration corridor and refugium of forest plants and animals in the grasslands of central North America. The Botanical Review, **54:** 44-81. Reproduced with permission.

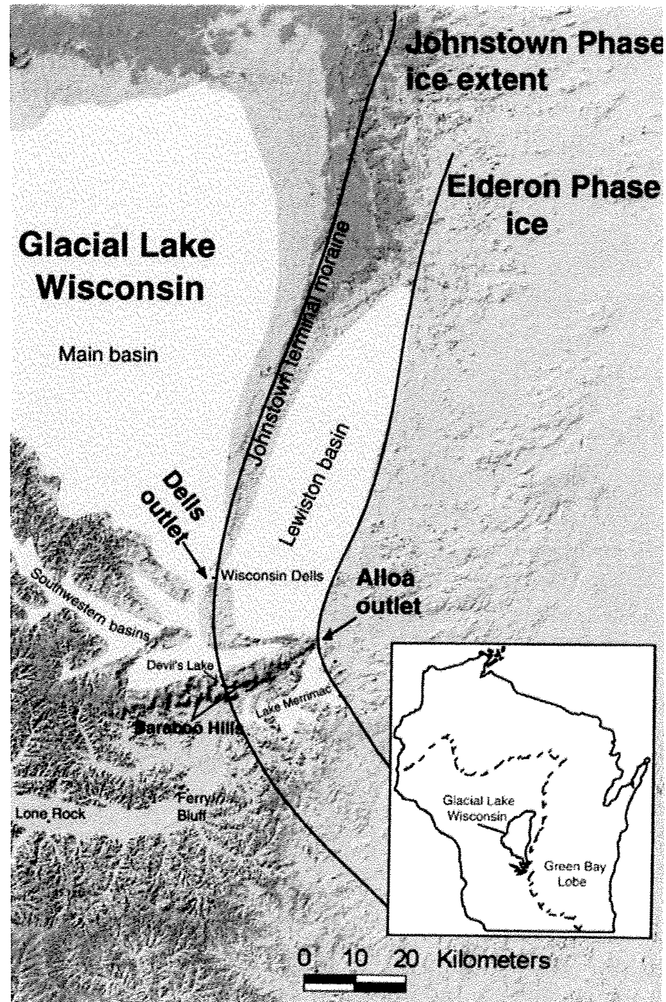


Figure 11. Glacial Lake Wisconsin.

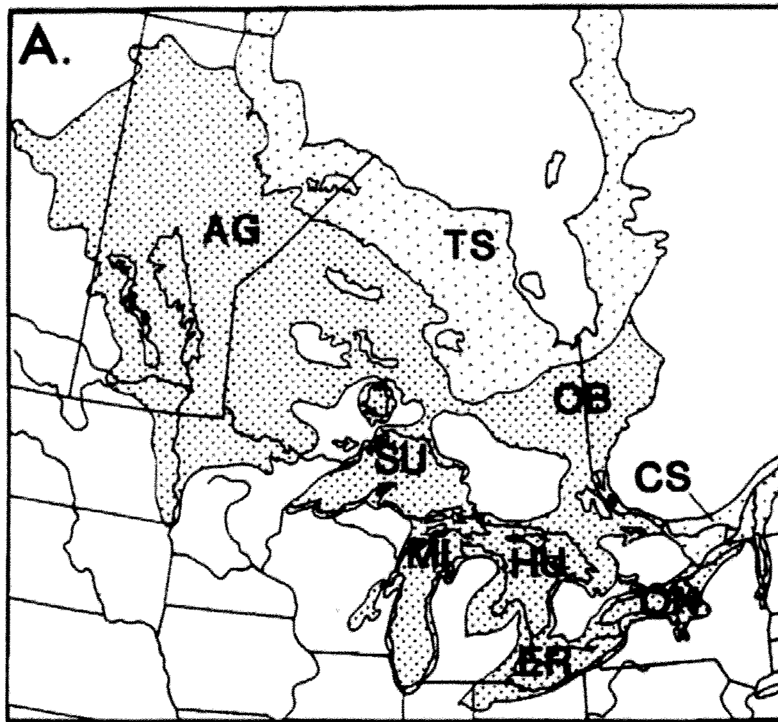
Glacial Lake Wisconsin came into existence approximately 1.8 million years ago. Black lines in the main figure, and dashed lines in the inset, indicate the leading edge of the ice.

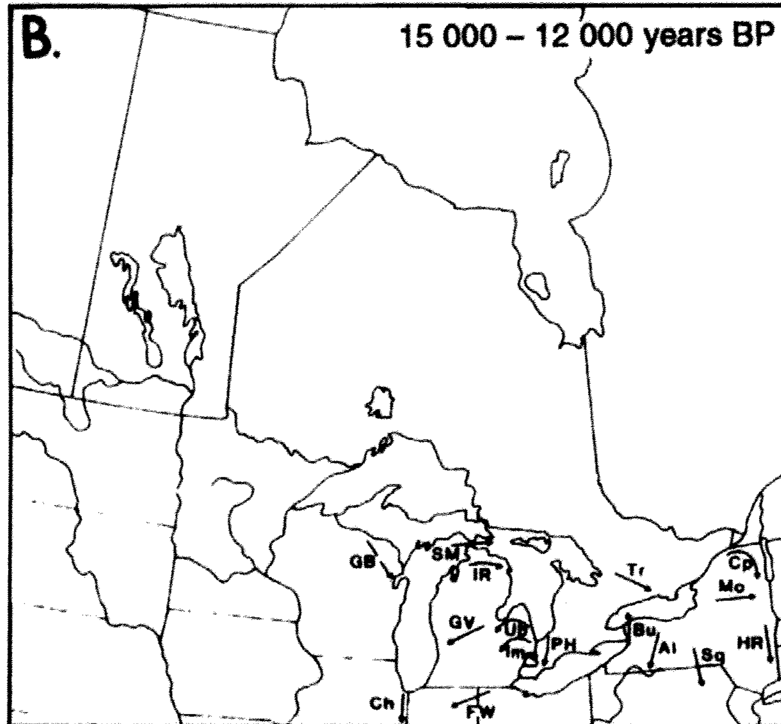
Source: Clayton, J.A and J.C. Knox. 2008. Catastrophic flooding from glacial Lake Wisconsin. Geomorphology, **93:** 384-397. © 2008 Canadian Science Publishing or its licensors. Reproduced with permission.

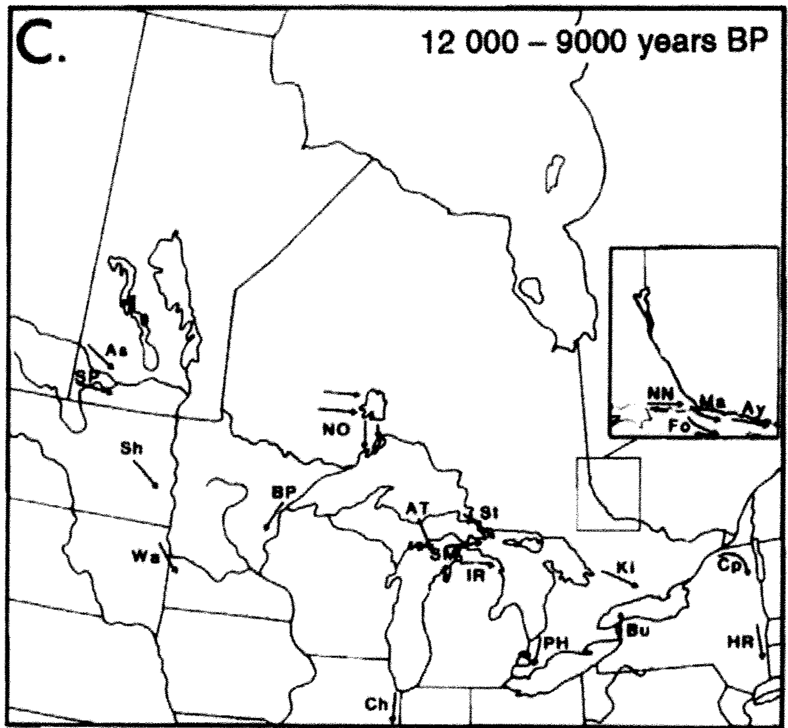
Mississippi drainage system (Underhill, 1986). Lake Maumee began to form in the basins of Lake Erie 14000 years ago, also draining into the Mississippi. Approximately 12700 years ago Lake Iroquois began to form in the Lake Ontario basin, eventually connecting to Lake Maumee. The Upper Allegheny and Monongahela Rivers may have connected the Ohio River basin to Lake Maumee and Lake Iroquois (Bailey and Smith, 1981). A number of drainage outlets connecting proglacial lakes to refugia have been theorized as colonization routes for freshwater fishes (Figure 13).

Lake Michigan is hypothesized to have been accessed by freshwater fishes via the Chicago outlet, which may have existed for up to 8000 years, originating up to 14000 years ago (Mandrak and Crossman, 1992). Multiple access points connected Lake Huron and Lake Erie to the Eastern and Central Highlands, including the Fort Wayne outlet which drained the Lake Erie basin via the Wabash and Maumee Rivers, covering regions of Illinois, Indiana and Ohio. This may have been the first outlet to form after glacial retreat and may have lasted up to 2000 years (Underhill, 1986). Fish may have been able to access what is now Lake Erie via this outlet, through the Ohio and Monongahela Rivers or over the Portage escarpment which separated Ohio and the Great Lakes (Borden and Krebs, 2009) (Figure 13). The Great Lakes remain in existence today, yet even larger proglacial lakes existed, providing habitat and colonization routes for freshwater fishes.

Proglacial Lake Agassiz was an enormous lake covering regions of Saskatchewan, Manitoba, Ontario, Minnesota and North Dakota. Lake Winnipeg and the Nelson River are modern day remnants of Lake Agassiz. Lake Agassiz was formed by the Red River lobe of the Laurentide ice sheet and allowed freshwater fish to disperse to novel habitats (Trenhaile, 1998). Glacial retreat from the Great Plains was rapid, providing early access to Lake Agassiz via the Missouri River


A.
AG
TS
OB
CS
SU
MI
HU
ER
ON


B.
15 000 – 12 000 years BP
GB
SM
IR
Tr
Cp
Mo
GV
Ub
Im
PH
Bu
Al
Sq
HR
Ch
FW


C.
12 000 – 9000 years BP
As
SP
Sh
NO
BP
Wa
AT
St
SM
IR
Ki
Cp
Bu
PH
HR
Ch
NN
Ma
Ay
Fo

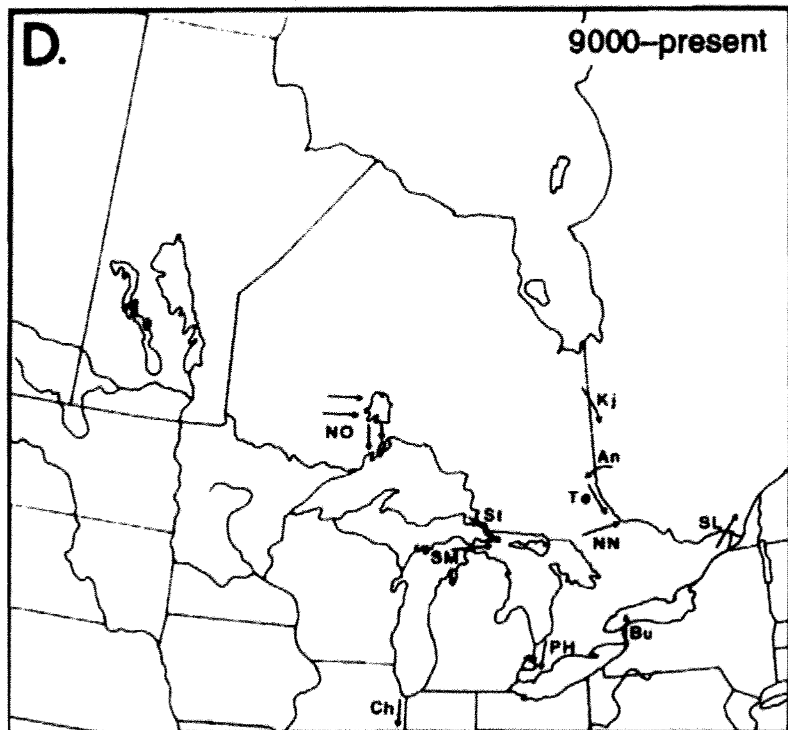

D.
9000–present
NO
Kj
An
Te
St
SM
NN
SL
Bu
PH
Ch

Figure 12. Glacial Great Lakes and Lake Agassiz.

Proglacial lakes which provided dispersal routes for freshwater fishes. A. The maximum extent of glacial lakes in northeastern North America. Light stippling indicates salt water, dark stippling indicates the maximum extent of freshwater lakes. AG, Lake Agassiz; CS, Champlain Sea; ER, Lake Erie; HU, Lake Huron; MI, Lake Michigan; OB, Barlow-Ojibway; ON, Lake Ontario; SU, Lake Superior; TS, Tyrell Sea. Figures B-D illustrate glacial outlets at various periods throughout the Pleistocene. Al, Allegheny; AT, Au Train; Bu, Buffalo; Ch, Chicago; FW, Fort Wayne; GB, Green Bay; PH, Port Huron; Sh, Sheyenne; SL, St. Lawrence; SM, Straits of Mackinac; Sq, Susquehanna; Wa, Warren. Arrows indicate the direction of flow.

Source: Mandrak, N.E. and E.J. Crossman. 1992. Postglacial dispersal of freshwater fishes into Ontario. Canadian Journal of Zoology, **70:** 2247-2259. © 2008 Canadian Science Publishing or its licensors. Reproduced with permission.

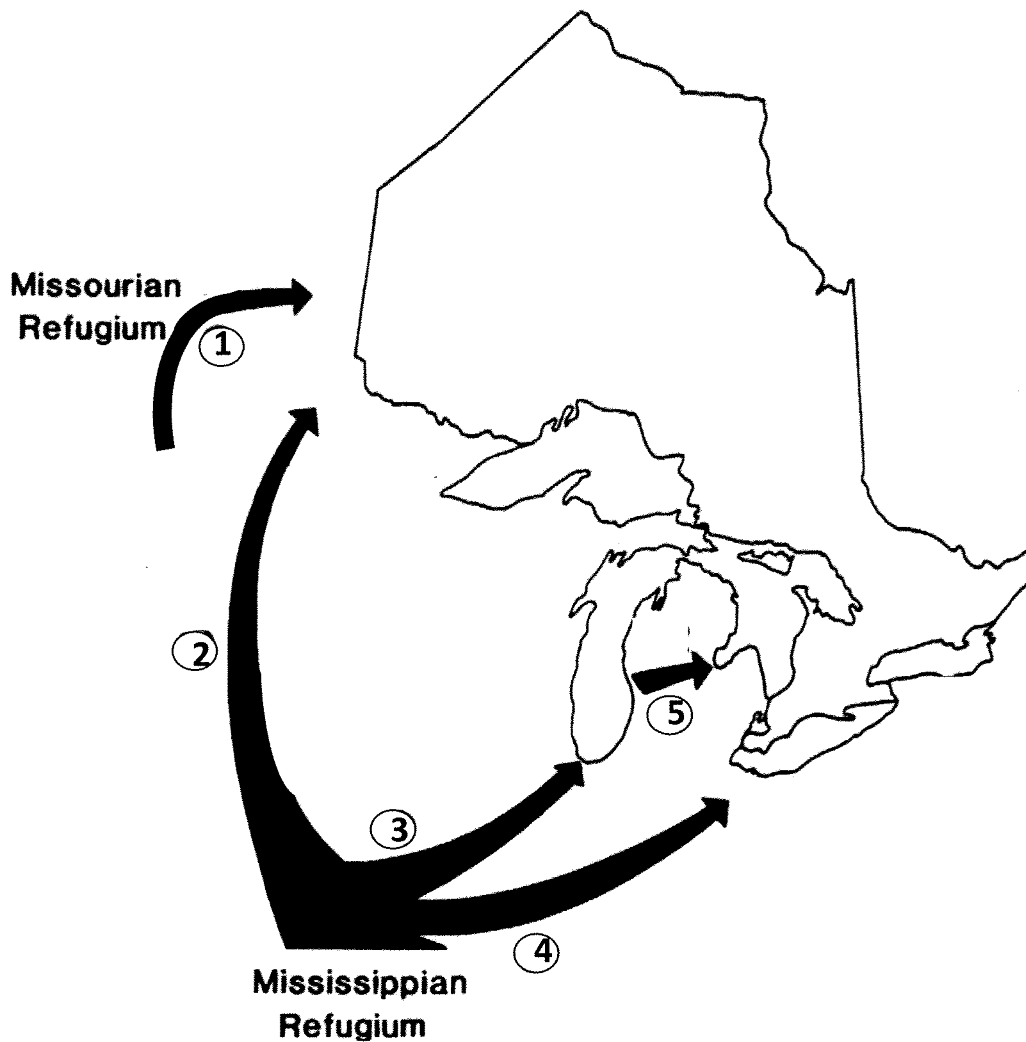


Figure 13. Potential dispersal routes to northeastern North America.

Outlets are numbered as follows: (1) Sheyenne; (2) Warren; (3) Chicago; (4) Fort Wayne; (5) Lower peninsula of Michigan.

Source: Mandrak, N.E. and E.J. Crossman. 1992. Postglacial dispersal of freshwater fishes into Ontario. Canadian Journal of Zoology, **70:** 2247-2259. © 2008 Canadian Science Publishing or its licensors. Reproduced with permission.

(Crossman and McAllister, 1986). Between 12800 and 10000 years ago Lake Agassiz was intermittently connected to the Mississippi drainage system by the Warren outlet (Bailey and Smith, 1981; Borden and Krebs, 2009). Retraction of the Laurentide ice sheet diverted flow of Lake Agassiz from the Mississippi eastward to the St. Lawrence; outflow ultimately discharged northward to the Hudson Bay (Ehlers and Gibbard, 2004).

Retreating ice margins and freshwater runoff greatly influenced postglacial dispersal. Many species of fish existed in waters near the leading ice edge during glacial retreats (Bailey and Smith, 1981). The emergence of large proglacial lakes and river connections allowed great opportunity for dispersal by these surviving populations (Hewitt, 1996). Populations which survived near the ice edge or in periglacial lakes did not acquire the level of genomic variation seen in species which inhabited areas south of the ice sheet. Larger populations existing in non-glaciated areas were more stable than those to the north (Hewitt, 1996, 2000).

North American freshwater fish now inhabiting regions to the north of the last glacial maximum exhibit reduced genetic diversity (Hewitt, 1996). Genomic diversity in glaciated areas remains limited as species have not existed in these regions for sufficient lengths of time, and likely were founded by groups with restricted genetic diversity (Hewitt, 1996, 2000). Leading edge colonization by small groups resulted in bottlenecks and founder effect, reducing genetic diversity in newly colonized populations (Hewitt, 1996).

Suture lines are correlated with a division between homozygous and heterozygous populations among virtually all species. A suture line runs westward from New England to the Dakotas, coinciding with the margin of the last glacial maximum. Suture lines affecting freshwater fishes are frequently correlated with the maximum extent of glaciation, changes in topography resulting from isostatic rebound or differences in water quality (including salinity) (Starnes and

Etnier, 1986; Hewitt, 1996; Berendzen et al., 2003). Suture lines are often indicative of regions where species survived the last glacial maximum (Hewitt, 1996).

Modern day major drainage basins of North America colonized by freshwater drum include the Mississippi, Great Lakes and Hudson Bay/Nelson River drainage systems (Figure 14). The Mississippi river originates in the glacial Lake Itasca in Minnesota and is joined by the Missouri and Ohio rivers, among others, as it flows south to the Gulf of Mexico. In total the basin covers 3,220,000 km^2 of area from 32 of the United States and 2 provinces in Canada. The St. Lawrence River drains the Great Lakes into the Atlantic Ocean draining an area of 1,344,200 km^2. This basin covers parts of Ontario and Quebec, and the northeastern United States. The Hudson Bay/Nelson River covers an area of 892,300 km^2, a portion of this watershed drains Lake Winnipeg into the Hudson Bay (Burr and Mayden, 1992; Benke and Cushing, 2005). These three drainage basins contribute to modern day and historical demography of freshwater drum.

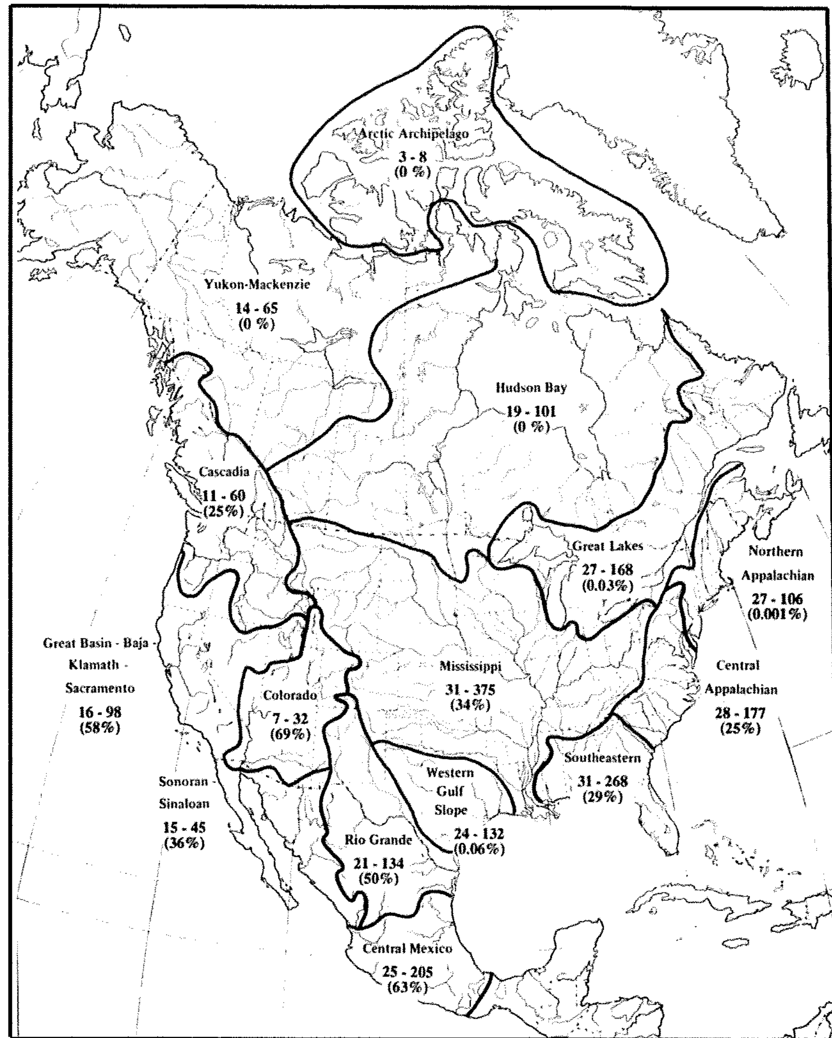


Figure 14. Drainage basins of North America.

Three major river drainage systems are applicable to present day Freshwater Drum habitat (identified by stars): the Mississippi, St. Lawrence and Nelson-Saskatchewan River basins.

Source: Burr, B.M and R.L. Mayden. 1992. Phylogenetics and North American freshwater fishes. Systematics, historical ecology, and North American freshwater fishes, edited by R.L. Mayden. Stanford University Press, Stanford, CA. p. 18-75.

PURPOSE AND HYPOTHESES

One goal of this study was to identify the contemporary population structure of the freshwater drum population. The origin of populations across the range of freshwater drum and the colonization routes used to access various habitats were also identified.

This study attempts to answer three questions (Q) addressing specific hypotheses (H_o) and testing the predictions (P) associated with those hypotheses in order to confirm or reject the associated hypotheses.

Many phylogeographic studies of freshwater fishes have been performed, each study showing evidence of genetic structure within populations (Bernatchez and Wilson, 1988; Wilson et al., 1996; Borden and Krebs, 2009; Stepien et al. 2009). I expect to observe genetic variation among and within sampled populations of freshwater drum, the majority of variation existing within the populations. Sampled populations in geographic proximity, or within the same drainage basin, should exhibit greater genetic similarity. Populations further north, in previously glaciated regions should exhibit a lesser degree of genetic variation (Bernatchez and Wilson, 1988; Hewitt, 1996, 2000).

Q1: Is there greater genetic similarity between populations within geographic proximity and shared drainage basins?

H_1: Genetic similarity will exist between freshwater drum populations within geographic proximity or shared drainage basins.

> **P1-1:** Populations in geographic proximity will display greater genetic similarity.
>
> **P1-2:** Populations within the same drainage basin will display greater genetic similarity.

P1-3: There will be more variation between populations than within populations.

Populations of freshwater fishes primarily survived the Wisconsin glaciation south of the ice sheet. The regions in which they survived can be identified via genetic analysis. Increased genetic diversity is observed in older populations, whereas those recently colonized exhibit less diversity (Hewitt 1996, 2000). The Mississippi has been a major colonization source for many freshwater fishes, as were the Teays and Missouri Rivers (Berendzen et al. 2003, 2008; Stepien et al., 2009).

Q2: In what regions did freshwater drum survive the last glacial maximum?

H_2: Genetic variation will be increased in populations which persisted throughout the last glacial maximum and decreased in recently colonized populations.

P2-1: Populations south of the last glacial maximum will exhibit increased genetic variation.

P2-2: Populations displaying increased genetic variation indicate regions which survived throughout the last glacial maximum and served as refugia.

Q3: What are the source populations of freshwater drum and what colonization routes were used to access new habitats?

$H_{3\text{-}1}$: Ohio River and Missouri/Mississippi River are source populations for other sampled populations.

P3-1(1): Genetic diversity will be increased in these populations.

P3-1(2): Haplotypes will be shared between the Ohio River and Missouri /Mississippi River populations and populations they colonized.

P3-1(3): Populations colonized by the Ohio River and Missouri River will be geographically proximal to the source population of the Ohio River or Missouri/Mississippi River.

$H_{3\text{-}2(1)}$: Populations of the Pickwick Reservoir were colonized by members of the Mississippi/Teays (Ohio) River drainage.

$H_{3\text{-}2(2)}$: Populations of the Great Lakes were colonized by members of the Mississippi/Teays (Ohio) River drainage.

P3-2(1): Genetic correlations in the form of shared haplotypes will exist between the Pickwick Reservoir and Great Lakes populations and the Mississippi/Teays (Ohio) River population.

$H_{3\text{-}3(1)}$: Populations of the Hudson Bay drainage were colonized by members of the Mississippi/Missouri River drainage.

$H_{3\text{-}3(2)}$: Populations of the Lake Sakakawea drainage were colonized by members of the Mississippi/Missouri River drainage.

$H_{3\text{-}3(3)}$: Populations of the Lake Pepin drainage were colonized by members of the Mississippi/Missouri River drainage.

$H_{3\text{-}3}$: Populations of the Lake Maloney drainage were colonized by members of the Mississippi/Missouri River drainage.

P3-3(1): Genetic correlations in the form of shared haplotypes will exist between the Hudson Bay, Lake Sakakawea, Lake Pepin and Lake Maloney populations and the Mississippi/Missouri River population.

METHODS

Materials and Methods

One hundred and seventy-two *Aplodinotus grunniens* specimens were collected using nets from eleven locations across the United States and Canada, representing the Hudson Bay (Nelson River), Great Lakes (St. Lawrence River) and Mississippi River drainage systems (Figure 15, Appendix 1). Fish were collected by the generous assistance of state and provincial agencies. Fish were frozen until further processing.

Muscle tissue was collected and stored in 95% ethanol. DNA extraction was performed under the protocol outlined in the QIAGEN DNeasy Extraction Kit (Qiagen, 2004). A small sample of muscle tissue, approximately 3mm x 5mm, was sliced into small pieces and placed in a 1.5mL tube. 180 µL of buffer ATL and 20 µL of proteinase K were added to the tube and vortexed. Tissue underwent lysis at 55°C and 400r.p.m. overnight for 8 to 12 hours on an Eppendorf thermomixer. After determining by eye that tissue was dissolved, 200 µL buffer AL was added to the lysed mixture and vortexed. 200 µL of 100% ethanol was added to the mixture and vortexed; the solution was then pipetted into a mini-spin column in a 2mL collection tube and centrifuged at 8000 r.p.m. for 1 minute. The spin column was transferred to a new tube and 500 µL of buffer AW1 added and centrifuged at 8000 r.p.m. for 1 minute. The spin column was again transferred to a new collection tube, 500 µL of buffer AW2 added and centrifuged at 14000 r.p.m. for 3 minutes. The column was placed in a new collection tube and 200 µL of buffer AE slowly pipetted onto the membrane of the spin column. The tube sat at room temperature for 1 minute and was then centrifuged at 8000r.p.m. for 1 minute. Extracted DNA eluent was stored at 4°C.

Figure 15. Freshwater drum sample collection sites.

Eleven sites were sampled across the distribution range. (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NE.

All samples were quantified using the Qubit fluorometer and Qubit dsDNA HS assay kit, performed under the following protocol. 199 μL of buffer was added to 1 μL of reagent; the mixture was vortexed and 195 μL pipetted to a 0.5mL assay tube. 5 μL of DNA eluent was added to the assay tube and vortexed. The mixture sat at room temperature for 2 minutes. The concentration of each sample was measured five times, the lowest and highest values discarded, and the average calculated (Appendix 2).

Control region

Initial amplification

Initial amplification utilized published primers for red drum (*Sciaenops ocellatus*), a marine species of family Sciaenidae (Gold et al., 1993). The forward primer used was L15943 (5'- GTA AAC CGG ATG TCG GGG GTT AG-3'), the reverse primer used was H16484 (5'-GGA ACC AGA TAC CAG GAA TAG TTC A-3'). These primers target a partial sequence of the control region of mitochondrial DNA.

Modifications to the polymerase chain reaction (PCR) were necessary to successfully amplify samples using red drum primers and are described as follows. The initial reaction was performed with 2.5 mM Thermopol buffer (New England Biolabs), 250 μM DNTPs (Fermentas), 0.25 μM forward primer (Operon), 0.25 μM reverse primer (Operon), 2 -U *Taq* DNA polymerase, 5 μL sample, and 15.8 μL sterile water to yield 25 μL of reaction mixture. PCR was performed under the following cycling parameters; 95°C initialization, followed by 50 cycles consisting of 95°C denaturation for 30 seconds, annealing at twelve graded temperatures incrementally

ranging from 45°C to 65.0°C for 30 seconds, and 72°C extension for one minute. Cycles were completed with a final extension of 72°C for five minutes. Annealing temperature was determined to be 59˚C using these primers.

Following amplification, 5 µL of PCR product was added to 3 µL 6X loading dye (New England Biolabs) and pipetted into lanes of a 6% polyacrylamide gel (Figure 16). 2 µL of 50 base pair ladder (New England Biolabs) was used to size amplified products. Gels ran at 118 volts for 45 minutes and were then stained with ethidium bromide solution for 20 minutes. Bands were visualized under ultraviolet light and photographed to be viewed with AlphaEaseFC, allowing for enhanced image and improved examination of bands.

Samples were purified using the QIAquick purification protocol as follows. 100 µL of buffer PB was added to 20 µL of amplified sample, transferred to a 1.5mL collection tube, vortexed and transferred to a QIAquick spin column in a 2mL collection tube. The column was centrifuged at 13000 r.p.m. for 1 minute and the flow-through discarded. 700µL of buffer PE was added to the column and centrifuged for 13000 r.p.m. for 1 minute and the flow-through was discarded. The column was then centrifuged for 13000 r.p.m. for 1 minute to remove any residual flow-through. The spin column was transferred to a sterile 1.5mL microfuge tube and 30µL of buffer EB slowly added to the center of the QIAquick membrane. The column stood at room temperature for 1 minute, then centrifuged at 13000 r.p.m. for 1 minute. Purified DNA was stored at -15°C until sequenced.

Purified DNA was sequenced using 3µL of purified sample, 25µM buffer and 2µL BDT (Applied Biosystems). The solution was transferred to a 0.2mL PCR collection tube and 0.15 µM of forward or reverse primer added. The sequencing reaction was performed under the following

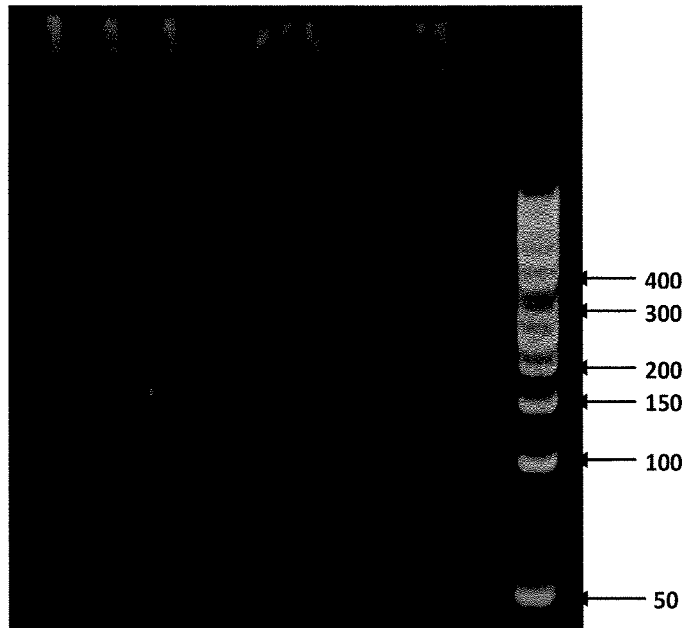


Figure 16. Polyacrylamide gel of freshwater drum control region sequences using red drum primers.

465 base pair control region sequence using red drum primers. Bands are faint and non-specific. Amplified region is circled. Base pairs corresponding to bands are marked on the right of the figure.

cycling parameters; 35 cycles of 95°C denaturation for 30 seconds, annealing temperature of 50°C for 15 seconds and an extension of 60°C for 4 minutes.

The resulting product underwent an ethanol/sodium acetate precipitation. 10 µL of the product of the sequencing reaction was mixed with 10 µL sterile water and added to 3 µL of sodium acetate, 62.5 µL 100% ethanol, and 14.5 µL sterile water in a 1.5 mL collection tube. Tubes sat at room temperature for 20 minutes and were then centrifuged for 20 minutes at 13000 r.p.m with the orientation of the tube noted. The portion of the collection tube oriented upward collects the DNA, aspiration was performed at the bottom of the tube. Supernatants were discarded and 250 µL of 70% ethanol added to the precipitate, and vortexed for 20 seconds; the tube was then centrifuged for 5 minutes at 13000 r.p.m in the same orientation as before. Supernatants were aspirated and discarded and precipitated samples dried at 45°C in a vacuum centrifuge for 15 minutes. Products were stored at -15°C until further processing.

Products were analyzed by the Lakehead University Paleo-DNA laboratory using DNA Sequencing Analysis Software Version 3.7 software on the Genetic Analyzer ABI3100 computer. Upon receipt of results, sequences were aligned using Gap4 software and confirmed manually. The initial sequence was determined to be 465 base pairs, after verification by NCBI BLAST the sequence was confirmed to be similar (90% maximum identity) to that of *Sciaenops ocellatus*.

Amplification with primers specific to freshwater drum

Primers specific to freshwater drum were designed after successful amplification using red drum primers (Appendix 3). The resulting sequence was examined manually to determine sites ideal as candidate primers. Candidate primers were tested using the Operon analysis tool (Figure 17).

The designed forward primer is 20 base pairs, nucleotide position 15591 (5'- CCT ACT GCA TCA AAG AGG AG -3') and has a GC content of 50%, Tm of 60.4°C, and contains one primer-dimer complement 4 base pairs in length. The reverse primer, nucleotide position 16028 (5' – CCA GAT ACC AGG AAT AGT TCA C -3') is 22 base pairs with a GC content of 42.1%, Tm of 55.8°C and no primer-dimer interactions. There are no primer-dimer interactions between the two.

PCR conditions were optimized for use with the designed primers. A 25 µL reaction was performed with 2.5 mM Thermopol buffer, 250 µM DNTPs, 0.125 µM forward primer, 0.125 µM reverse primer, 1 -U *Taq* DNA polymerase, 5 µL sample, and 16.4 µL sterile water. The PCR reaction was performed under the following cycling parameters; 95°C initialization, followed by 35 cycles consisting of 95°C denaturation for 30 seconds, annealing at twelve graded temperatures ranging from 45°C to 65.6°C for 30 seconds, and 72°C extension for one minute. Cycles were completed with a final extension of 72°C for five minutes. The optimal annealing temperature was determined to be 55°C.

Amplified DNA sequences were visualized (Figure 18), purified, sequenced, precipitated and analyzed as for red drum primers.

CONTROL-REGION Analysis

MW: 6135.049

GC Content: 50%

TM: 60.4 °C

Reverse Complement: 5' - CTCCTCTTTGATGCAGTAGG - 3'

	C	C	T	A	C	T	G	C	A	T	C	A	A	A	G	A	G	G	A	G
C							X								X		X	X		X
A			X			X				X										
C							X								X		X	X		X
T				X					X			X	X	X		X			X	
T				X					X			X	X	X		X			X	
G	X	X			X			X			X									
A			X			X				X										
T				X					X			X	X	X		X			X	
A			X			X				X										
A			X			X				X										
G	X	X			X			X			X									
G	X	X			X			X			X									
A			X			X				X										
C							X								X		X	X		X
C							X								X		X	X		X
A			X			X				X										
T				X					X			X	X	X		X			X	
A			X			X				X										
G	X	X			X			X			X									
A			X			X				X										
C							X								X		X	X		X
C							X								X		X	X		X

eurofins mwg operon

Longest primer-dimer complement: 0

Figure 17. Analysis of control region primers.

Primers were tested using the Operon analysis tool to verify suitability. MW, molecular weight; GC, guanine/cytosine; TM, melting temperature.

Source: http://www.operon.com/tools/oligo-analysis-tool.aspx?

ND2 region

Primers were designed based on the *Aplodinotus grunniens* ND2 gene sequence at GenBank accession number AY225720 using the same design parameters as for designing control region primers and tested using the Operon oligo tool (Figure 19). The forward primer, nucleotide position 421 (5' - CTC CTC CTC CAA ATC TAC CCC GC – 3'), is 23 base pairs, with a GC content of 60.9%, Tm of 68.1°C, and no primer dimer complement. The reverse primer, nucleotide position 937 (5' – CCT TGA CGC CTC CAA TCA TCC CC - 3'), 22 base pairs in length, with a GC content of 59.1%, Tm of 66.4°C, and no primer-dimer interactions. There is also no primer-dimer complement between the two primers. These primers were designed to have a high melting temperature so as to be highly specific to freshwater drum.

Polymerase chain reaction was optimized using a 10 µL reaction of 0.4 µM $MgCl_2$, 1.0 µM buffer, 100 µM DNTP, 0.1 µM each of forward and reverse primer, 1 -U platinum *Taq* DNA polymerase, 2.0 µM sample and 5.9 µM H_2O. The reaction was run with initialization of 94°C for 2 minutes, followed by 35 cycles of denaturation at 94°C for 30 seconds, annealing at 60°C for 30 seconds, and extension at 72° for 1 minute.

Samples were visualized using 2 µL of amplified product in 2 µL 6X loading dye, which was pipetted into lanes of a 1% agarose gel stained with 1 µL ethidium bromide (Figure 20). 2 µL of 1 kilobase pair ladder was used to size amplified products. Gels ran at 118 volts for 20 minutes and bands were visualized under ultraviolet light.

Samples were purified using the QIAquick purification protocol as described for control region purification, with the following modification. As the PCR reaction was performed as a 10ul reaction only 40 µL of buffer PB was added to the remaining 8 µL of amplified sample. The

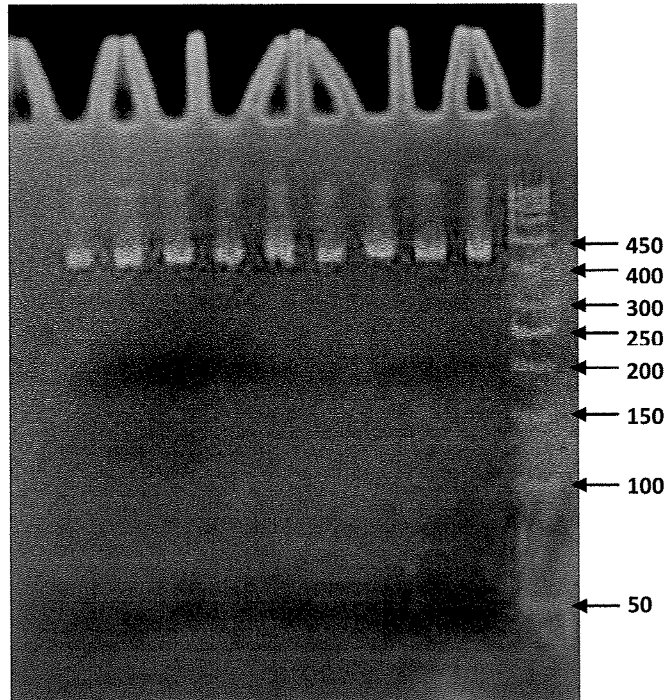


Figure 18. Polyacrylamide gel of freshwater drum control region sequences.

438 base pair control region sequence using primers designed for freshwater drum. The size marker on the right of the gel is a 50bp ladder. Base pairs corresponding to bands are marked on the right of the figure.

ND2-REGION Analysis

MW: 6800.44

GC Content: 60.86956%

TM: 68.11739 °C

Reverse Complement: 5' - GCGGGGTAGATTTGGAGGAGGAG - 3'

	C	T	C	C	T	C	C	T	C	C	A	A	A	T	C	T	A	C	C	C	C	G	C
C																						X	
C																						X	
C																						X	
C																						X	
T											X	X	X				X						
A		X			X			X						X		X							
C																						X	
T											X	X	X				X						
A		X			X			X						X		X							
A		X			X			X						X		X							
C																						X	
C																						X	
T											X	X	X				X						
C																						X	
C																						X	
G	X		X	X		X	X		X	X					X			X	X	X	X		X
C																						X	
A		X			X			X						X		X							
G	X		X	X		X	X		X	X					X			X	X	X	X		X
T											X	X	X				X						
T											X	X	X				X						
C																						X	
C																						X	

Longest primer-dimer complement: 0

eurofins mwg|operon

Figure 19. Analysis of ND2 region primers.

Primers were tested using the Operon analysis tool to verify suitability. MW, molecular weight; GC, guanine/cytosine; TM, melting temperature.

Source: http://www.operon.com/tools/oligo-analysis-tool.aspx?

samples then underwent the sequencing reaction and precipitation and were analyzed as outlined for the control region.

Sequence analysis

Sequences were aligned and analyzed using Gap4 software (Staden et al., 2000) and confirmed manually. Polymorphisms were identified (Figure 21).

The amplified control region sequence was approximately 450 base pairs, the ND2 sequence approximately 500 base pairs. Verification by NCBI BLAST confirmed the control region sequence was similar to that of other Sciaenidae. The amplified *Aplodinotus grunniens* sequence showed 90% maximal identity to a partial sequence of the mitochondrial control region of *Sciaenops ocellatus* at GenBank accession number EU363526.1 (Figure 22). The ND2 region corresponded to the *Aplodinotus grunniens* sequence at GenBank accession number AY225720 (Near et al., 2003).

Sequences were then analyzed for pairwise differences, shared haplotypes, haplotype frequencies, mismatch distribution, linkage disequilibrium, Tajima's D, AMOVA and F_{ST} using Arlequin 3.5.1.2. (Excoffier and Lischer, 2010). Phylogenetic trees were formed using Phylip version 3.6.9. (Felsenstein, 2009) and visualized using Treeview version 1.6.6 (Page, 1996). Bayesian inference was performed using Mr. Bayes version 3.1.2. (Ronquist and Huelsenbeck, 2003). Haplotype frequency and polymorphic data were both included in analysis.

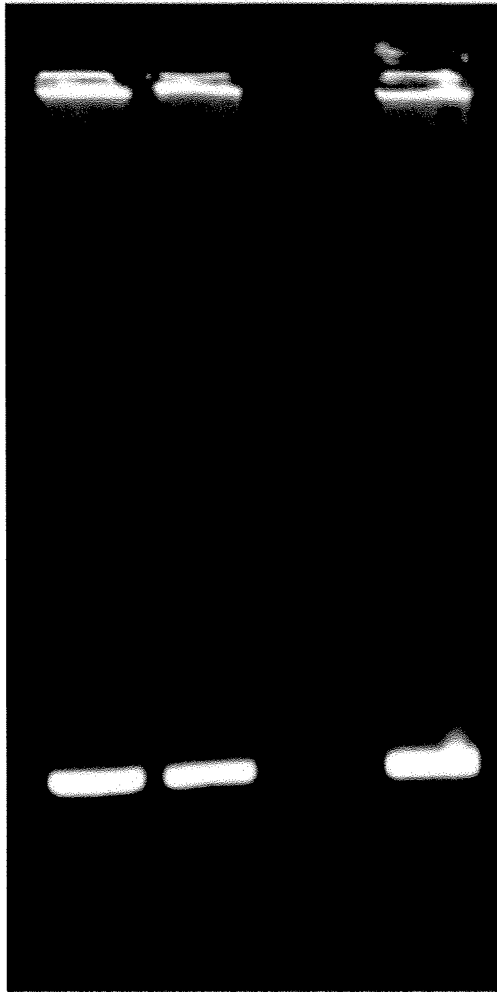


Figure 20. Agarose gel of freshwater drum ND2 sequences.

1 kbp ladder is not shown. Sequences are 500 bp, lane 3 is a negative control.

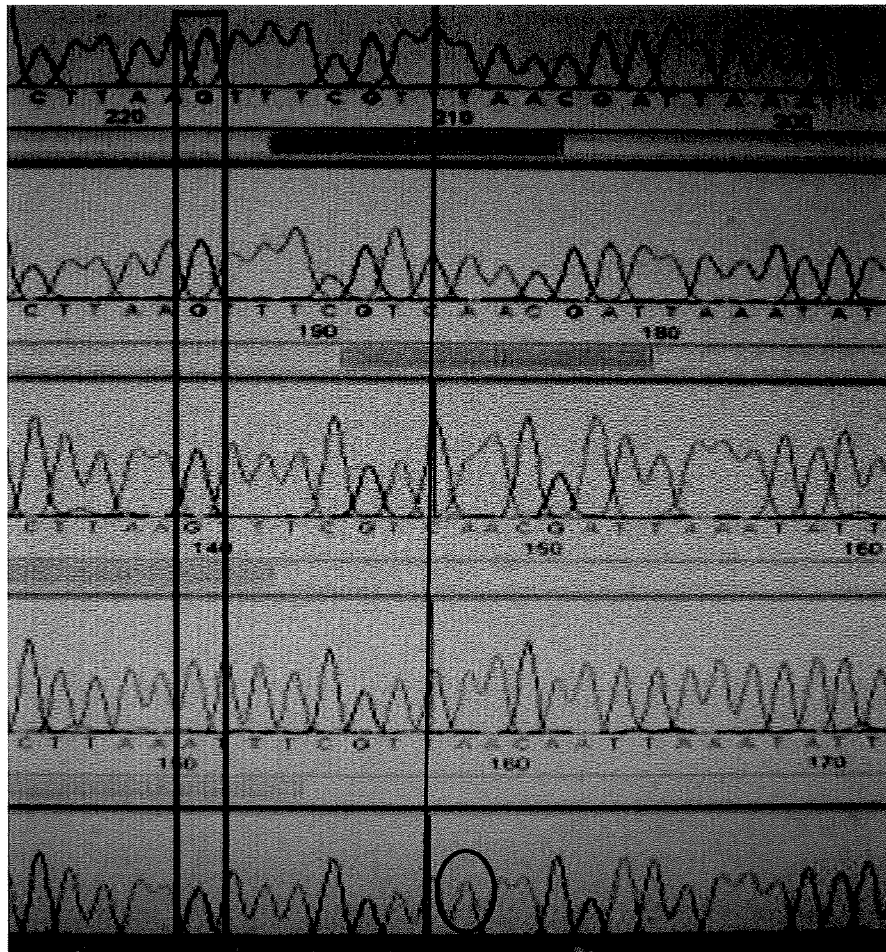


Figure 21. Alignment and analysis of sequences.

Sequences were aligned and analyzed using Gap4 software. Polymorphisms were identified and checked by eye. Illustrated is a partial section of control region analysis. A base substitution is displayed in the black rectangle – the fourth sequence from the top has substituted adenine for guanine. A base substitution of thymine for cytosine is depicted in the lowest 2 sequences, identified by the black line. An insertion of thymine is illustrated inside the black circle.

```
Query  9     CAAAGAGGAGGGAATTTAACCCCCACCACTAATTCCCAAAACTAGTATTCTAAACTAAAC  68
             ||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Sbjct  36    CAAAGAGGAGGGATTTTAACCCCCACCACTAATTCCCAAAACTAGTATTCTAAACTAAAC  95

Query  69    TACTCTTTGTACATATATGTATTTACACCATGAATTTATATTAACCATATCAATAGCATT  128
             |||||||||| |||||||||||||||||||| || ||||||||||||||||||||| |||
Sbjct  96    TACTCTTTGTGCATATATGTATTTACACCATTAAATTATATTAACCATATCAATAGTATT  155

Query  129   CAAGTACATAAATGTTCA-ATTAACATTTCTTGGT-TTAACACATTcaaacagcaaaatt  185
             ||||||||||| |||| | || | || | |||||| |||   |||||| ||| ||  ||
Sbjct  156   CAAGTACATAACTGTTTATAT-AGCAGTACTTGGTATTATT-CATTCACACAGCACCATA  213

Query  186   tagaagaaggatgtacataaagcataaatagtagagcgcaacaaggaaatgatgttagga  241
             || |  ||  || ||||||||||||||||| || | | ||||||  |||| || |||  |
Sbjct  214   TACAG-AATTATATACATAAAGCATAAATAGTATACCCCAACAAG-AAATTATATTACCA  271

Query  242   ttGTTGACGAAACTTAAGATCTAATA-CTTTAATCCAT-GAACAAAGATATACCACGAAC  299
              ||  | ||||| |||||| |||| | ||  || |||| |  ||| ||||||||||| ||
Sbjct  272   CTG--G-CGAAATTTAAGACCTAACAACTG-AACCCATAGGTCAA-GATATACCACGGAC  326

Query  300   CCAACATCCCGTCA-ACTTCACAATTTTAATGTAGTAAGAACCGACCATCAGTTGATTTC  358
              |||||||||| || || ||| ||| ||||||||||||||||||||||||||||||||||
Sbjct  327   TCAACATCCCGCCACACCTCAAAATCTTAATGTAGTAAGAACCGACCATCAGTTGATTTC  386

Query  359   TTAATGCATACGGTTATTGAAGGTGAGGGACAATTATCGTGGGGGTTTCACACAGTGAAC  418
             |||| ||||||||||||||||||||||||||||  || ||||||||||||||||||||||
Sbjct  387   TTAAGGCATACGGTTATTGAAGGTGAGGGACAAAAATAGTGGGGGTTTCACACAGTGAAC  446

Query  419   TATTCCCTGGTATC  432
             |||||| ||| |||
Sbjct  447   TATTCC-TGGCATC  459
```

Figure 22. Comparison of *Aplodinotus grunniens* and *Sciaenops ocellatus* mitochondrial control region sequences.

Overall maximal identity was determined to be 90% between sequences. Polymorphisms are highlighted in yellow. Query is the amplified sequence of *Aplodinotus grunniens* from this study, Sbjct refers the mitochondrial control region of *Sciaenops ocellatus* at GenBank accession number EU363526.1.

Source:http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&BLAST_PROGRAMS=megaBlast&PAGE_TYPE=BlastSearch&SHOW_DEFAULTS=on&LINK_LOC=blasthome

Intra-population analysis

The mean number of pairwise differences among haplotypes was calculated by Tajima's formula:

$$\hat{\pi} = \frac{n}{n-1} \sum_{i=1}^{k} \sum_{j=1}^{k} p_i p_j \hat{d}_{ij}$$

where *i, j, k* is the number of haplotypes, p_i and p_j are the frequency of the haplotype, *n* is the number of samples, and d_{ij} estimates the number of mutations subsequent to divergence of haplotypes *i* and *j* (Tajima, 1993). Nucleotide diversity was measured for each region sampled. Increased diversity is indicative of an older population, low levels of nucleotide diversity indicate a population has been recently colonized (Tajima, 1993).

Sequences that were determined to be identical among populations were recorded as shared haplotypes. Haplotype frequencies among the total population were assessed based on these data.

Mismatch distribution measures sequence differentiation among alleles (Excoffier and Lischer, 2010). An excess of haplotypes occurring at a low frequency indicates purifying selection or recent population expansion. Mismatch distribution analysis determines if regions underwent rapid expansion by analyzing haplotype dispersal (Hartl and Clark, 1997). 1000 bootstrap replicates were used in analysis, adequate to diminish sampling error. Samples were tested for goodness of fit to the demographic expansion model (Schneider and Excoffier, 1999) using the formulas:

$$\theta_0 = \sqrt{v - m}$$

$$\tau = m - \theta_o$$

where Θ is the mutation parameter, v is the variance, m is the mean, and t is the time to expansion. P values were estimated using the formula:

$$P = \text{number of } SSD_{sim} \text{ larger or equal to } SSD_{obs}/B$$

where B is the number of simulated samples (Schneider and Excoffier, 1999).

Populations were also sampled for goodness of fit to the spatial expansion model (Excoffier, 2004), using the formula:

$$F_0(S; M, \theta_0; \theta_1, \tau) = \frac{\theta_1^S}{A^{S+1}} + \sum_{j=0}^{S} \left(\frac{(M e^{-\tau} + C)\theta_0^j \tau^{S-j}}{(M+1)(\theta_0+1)^{j+1} (S-j)!} - \frac{\tau^j \theta^{S-j} C}{j! A^{S-j+1}} \right)$$

where T is time in generations, N is deme size, m is migration rate, $\theta_0 = 2N_{0\mu}$, $\theta_1 = 2N_{1\mu}$, $\tau = 2T\mu$, $A = \theta_1 + M + 1$ and $C = e^{-\tau A/\theta_1}$ (Excoffier, 2004). P values were also measured for this model.

Sequence differentiation is correlated with time since divergence. Tau is an index of the time which has passed since population expansion determined by the formula:

$$t_0 = \frac{N_1}{N_0} E(t_{01}) + \frac{N_2}{N_0} E(t_{02})$$

where N_1 and N_2 are haploid populations of constant size, N_0 is the ancestral population size, t_0 is the average coalescence time of two genes from the same population and t_{01} and t_{02} are the average coalescence time of two genes from different populations (Gaggiotti and Excoffier, 2000).

Linkage disequilibrium is the non-random association of alleles at loci. The level of linkage disequilibrium should be inverse to the level of recombination and decreases with each

generation. Regions with high mutation rates should exhibit low levels of linkage disequilibrium (Hartl and Clark, 1997). The presence of linkage disequilibrium indicates selection on multilocus genotypes, genetic drift or blending of populations, therefore can assess drift in populations. Analysis of linkage disequilibrium entails the use of a Markov chain. Contingency tables are written as two rows with the *i* allele (i_1 and i_2) and two columns with the *j* allele (j_1 and j_2). The Markov chain decreases the counts of cells i_1, j_1 and i_2, j_2 by 1 and increases the counts of cells i_1, j_2 and i_2, j_1 by 1, resulting in the allele count remaining unchanged. The probability of the new table is determined by the ratio:

$$R = \frac{L_1}{L_0}$$

The chain is repeated 10,000 times, the P value is determined to be the proportion of simulated tables that have probability equal to or less than that of the observed table (Raymond and Rousset, 1995).

Tajima's D test of neutrality measures the neutrality of sequence differentiation, meaning whether evolution is occurring randomly or non-randomly. The formula used is:

$$D = \theta\pi - \theta s/\sqrt{var(\theta\pi - \theta s)}'$$

where $\Theta\pi$ = mean number of pairwise differences, Θs = $\theta s = S / \sum_{i=0}^{n-1}$, and S is the number of segregating sites. Positive values of Tajima's D values signify low levels of low and high frequency polymorphisms, indicating a decrease in population size or balancing selection. Negative Tajima's D signifies an excess of low frequency polymorphisms indicative of population size expansion, positive selection, or a recent bottleneck (Tajima, 1993).

Inter-population analysis

Molecular variance within and between populations was analyzed using AMOVA. A large population divided into isolated subpopulations typically exhibits lower levels of heterozygosity than a single undivided population. AMOVA analyzes the sum of squared differences between all pairs of haplotypes to develop an analysis of molecular variance which is segmented into covariance components. The formula used to determine covariance components from the sum of squared deviations is:

$$x_{ijk} = a_k + b_{ik} + c_{ijk}$$

where *i* is the haplotype frequency in *j* population in group *k*, *a* is a group, *b* is a population and *c* is a haplotype. A group is defined as a region of interest, in the case of this study there is a single group consisting of all sample locations. The haplotype frequency of each unit determines the covariance components σ_a^2, σ_b^2, and σ_c^2 which represent the differences among populations, differences among haplotypes in different populations in a group, and differences among haplotypes within a population. The sum of covariance components is σ_T^2 , the total molecular variance (Table 1) :

P is the total number of populations, N is the total number of gene copies. The value of *n* is calculated as:

$$n = (N - \sum_p N_p^2/N)/P - 1$$

Table 1. AMOVA (analysis of molecular variance) table.

The formulas used to determine each partition within the AMOVA table are displayed.

Source of variation	Degrees of freedom	Sum of squared deviations (SSD)	Variance components
Among populations	P-1	SSD (among populations)	$n\sigma_a^2 + \sigma_b^2$
Within Populations	N-P	SSD (within populations)	σ_b^2
Total	N-1	SSD (total)	σ_T^2
FST = $\sigma_a^2 \div \sigma_T^2$			

A minimum spanning tree was constructed based upon the mean number of pairwise differences between all haplotypes using the modified algorithm of Rohlf (1973) (Excoffier et al., 1992).

F_{ST} is used to determine the correlation of alleles in subpopulations to those of the entire population, and can be used to interpret the relationships of all populations, not only that included in the sample, and is calculated as:

$$F_{ST} = (H_T - H_S)/H_T$$

where H_T is the heterozygosity of the total population and H_S is the heterozygosity of the subpopulation. P values < 0.05 were considered to be statistically significant, indicative of populations with a high degree of genetic variation (Slatkin and Voelm, 1991).

Phylogenetic trees were drawn using the programs “dnapars” and “drawgram” in PHYLIP. Dnapars uses parsimony to estimate phylogenetic trees. This method assumes that sites and lineages evolve independently, the probability of mutation at a site is small over the time

represented by a branch, multiple changes in a recurrent branch are not more probable than a single change in a branch which occurs infrequently, and two changes in one site are not more probable than one change in another site. Drawgram uses the output file from dnapars to form phylogenetic trees (Felsenstein, 2009). Phylogenetic trees were viewed with Treeview (Page, 1996).

Bayesian analysis was performed using Mr. Bayes version 3.1.2. Mr. Bayes uses Monte Carlo Markov chains to determine phylogenies. This analytical method employs the principles of maximum likelihood to determine phylogenies and the model of best fit. The relative rate of nucleotide substitution is analyzed, models differ based on the parameters included (Ronquist and Huelsenbeck, 2003).

RESULTS

Control region

Intra-population analysis

The control region sequence is approximately 450 base pairs in length with 39 different haplotypes. Diversity indices for the control region displayed a wide range (Table 2). The greatest values were found in the Ohio River (10.952), Missouri River (9.067) and Lake Sakakawea (9.033). Nelson River exhibited the lowest value of diversity (1.905).

Haplotype distribution and frequency (Appendix 3) was highly variable and haplotypes were shared among many populations (Table 3). Significant commonalities among populations have been noted. Lake Erie and Lake Ontario share the same predominant haplotype (h0005, frequency of 0.8 and 0.429 respectively), the other Great Lakes drainage sample, Green Bay, shares the same haplotype at a frequency of 0.188. The predominant haplotype of Green Bay (h0012, 0.562) is shared with only Lake Erie (0.1). Nelson River and Lake Winnipeg, the only sample sites from the Hudson Bay drainage, share the same predominant haplotype (h0001, 0.8 and 0.625 respectively). Lake Maloney is composed of only 4 haplotypes, the two predominant haplotypes (h0007(0.5), h0647(0.375)) not being shared with any other population.

The demographic expansion model is statistically significant for Lake Maloney (0.011) and Lake Sakakawea (0.014) based on SSD (Table 4). No statistical significance was found under the spatial expansion model for any sampled locations. Tau values indicate that Nelson River, Lake Pepin and Pickwick Reservoir are the most recently colonized populations while Lake Maloney, Green Bay and the Ohio River appear to be oldest.

Table 2. Analysis of pairwise differences of the control region of freshwater drum.

Nucleotide diversity was calculated based on the number of differences between all pairs of haplotypes in each sample. Samples with higher values of π are more diverse. Samples marked with an asterisk* are significantly higher or lower than the mean value.

Sample location	Number of haplotypes	Nucleotide diversity (π)
Lake Ontario	4	6.286
Lake Erie	4	4.263
Green Bay	8	6.417
Nelson River	3	1.905*
Lake Winnipeg	5	4.700
Lake Sakakawea	7	9.033
Lake Pepin	9	5.549
Lake Maloney	6	7.792
Missouri River	11*	9.067
Ohio River	6	10.952*
Pickwick Reservoir	10*	4.338
Mean	**6.6**	**6.391**

Table 3. Shared control region haplotypes among sampled populations.

Number of haplotypes shared among each population sampled.

Sample site	Lake Ontario	Lake Erie	Green Bay	Nelson River	Lake Winnipeg	Lake Sakakawea	Lake Pepin	Lake Maloney	Missouri River	Ohio River	Pickwick Reservoir
Lake Ontario	*	3	1	0	0	0	1	0	1	1	0
Lake Erie	3	*	2	0	0	0	1	0	1	1	0
Green Bay	1	2	*	1	1	1	1	0	1	1	1
Nelson River	0	0	1	*	3	1	1	1	1	0	2
Lake Winnipeg	0	0	1	3	*	1	2	2	2	0	3
Lake Sakakawea	0	0	1	1	1	*	1	0	0	2	2
Lake Pepin	1	1	1	1	2	1	*	2	2	1	5
Lake Maloney	0	0	0	1	2	0	2	*	1	0	2
Missouri River	1	1	1	1	2	0	2	1	*	1	1
Ohio River	1	1	1	0	0	2	1	0	1	*	1
Pickwick Reservoir	0	0	1	2	3	2	5	2	1	1	*

Table 4. Control region mismatch distribution. Column 2 illustrates the distribution of differences between pairs of haplotypes. Columns 3 and 5 illustrate tau values under the demographic or spatial expansion models, and columns 4 and 6 illustrate goodness of fit for these models. P < 0.05 is statistically significant and marked in **bold**.

	Mismatch observed variance	Tau – Demographic expansion	Demographic expansion model, P (sum of squared deviation)	Tau – Spatial expansion	Spatial expansion model, P (sum of squared deviation)
Lake Ontario	32.914	16.326	0.232	4.530	0.108
Lake Erie	38.915	3.0000	0.054	15.511	0.413
Green Bay	43.472	14.625	0.292	14.103	0.505
Nelson River	9.683	3.0000	0.056	3.359	0.436
Lake Winnipeg	25.153	10.455	0.162	9.610	0.561
Lake Sakakawea	34.810	16.025	**0.014**	14.552	0.093
Lake Pepin	17.006	3.984	0.447	3.639	0.659
Lake Maloney	38.738	14.824	**0.011**	14.159	0.088
Missouri River	34.164	1.496	0.149	11.398	0.076
Ohio River	19.048	14.867	0.094	14.804	0.140
Pickwick Reservoir	7.588	4.133	0.785	3.125	0.861

Control region sequences show low levels of linkage disequilibrium in Lake Ontario, Nelson River, Lake Winnipeg, Pickwick Reservoir, Ohio River, and Lake Pepin. Lake Sakakawea, Green Bay and the Missouri River display moderate levels while Lake Maloney and Lake Erie show high levels of linkage disequilibrium.

The P value of Tajima's D is statistically significant (0.043) for Nelson River (Tajima's D = -1.61385) indicating population size expansion, positive selection, or a recent bottleneck (Table 5). No other populations were found to have significant values for Tajima's D.

AMOVA analysis comparing individual populations indicated that most variance is found within populations (72.94%) (Table 6a), based upon drainage basins, most variation (81.82%) is found within drainage basins (Table 6b).

Control region

Inter-population analysis

AMOVA analysis determined the F_{ST} value for the total population was 0.27060 (Table 6), and for drainage basins was 0.18177 (Table 6a, 6b).

Analysis of F_{ST} values revealed that each population was significantly correlated with a minimum of three other populations. The Missouri River was significantly correlated with the greatest number of populations while Lake Erie, Green Bay, Nelson River and Lake Maloney showed significant differentiation from the majority of populations (Table 7, 8). There appears to be some correlation among drainage basins. The two samples from the Hudson Bay drainage system are similar, as are Lake Erie and Lake Ontario in the Great Lakes drainage basin.

Table 5. Control region analysis of Tajima's D test of neutrality. P < 0.05 are statistically significant are marked in **bold**. Negative Tajima's D values indicate population size expansion, positive selection or a recent bottleneck.

Sample location	Tajima's D (control region)	Tajima's D, P=(D simulated < D observed) control region
Lake Ontario	-0.52405	0.31900
Lake Erie	-0.41266	0.34800
Green Bay	-0.02362	0.55000
Nelson River	-1.61385	**0.04300**
Lake Winnipeg	-0.94501	0.18400
Lake Sakakawea	0.96600	0.88400
Lake Pepin	-0.91964	0.20300
Lake Maloney	2.06562	0.99200
Missouri River	0.064436	0.77600
Ohio River	0.50306	0.71200
Pickwick Reservoir	-0.98116	0.16800

Table 6. Control region AMOVA analysis of populations.

(a) Analysis of the variation among and within populations.

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of variation
Among populations	10	190.395	1.12269 Va	27.06
Within Populations	147	444.852	3.02620 Vb	72.94
Total	157	635.247	4.14890	
Fixation Index FST: 0.27060				

(b) Analysis of the variation among and within drainage basins.

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of variation
Among populations	2	60.959	0.80852 Va	18.18
Within Populations	104	378.518	3.63959 Vb	81.82
Total	106	439.477	4.44811	
Fixation Index FST: 0.18177				

Table 7. Population pairwise F_{ST} values based on control region analysis. 0.00000 indicates no differentiation between populations, 1.00000 indicates complete differentiation. Negative values should be considered as 0.00000. Theoretically, all values >0.25000 indicate significant differentiation and are highlighted in **bold**.

	Lake Ontario	Lake Erie	Green Bay	Nelson River	Lake Winnipeg	Lake Sakakawea	Lake Pepin	Lake Maloney	Missouri River	Ohio River	Pickwick Reservoir
Lake Ontario	0.00000										
Lake Erie	-0.03543	0.00000									
Green Bay	**0.38594**	**0.48573**	0.00000								
Nelson River	**0.32047**	**0.42476**	**0.61194**	0.00000							
Lake Winnipeg	0.14041	**0.28300**	**0.47299**	0.00983	0.00000						
Lake Sakakawea	0.18882	**0.30286**	0.23405	**0.43282**	**0.25009**	0.00000					
Lake Pepin	0.01356	0.15309	**0.42676**	0.10085	-0.00800	0.19670	0.00000				
Lake Maloney	**0.28184**	**0.39863**	**0.30585**	**0.45715**	**0.32705**	0.22073	**0.28794**	0.00000			
Missouri River	0.17326	**0.30520**	0.10807	**0.37179**	0.23651	0.09418	0.18475	0.14302	0.00000		
Ohio River	0.13100	**0.27976**	0.10904	**0.44508**	**0.25292**	0.03847	0.18492	-0.01223	-0.00203	0.00000	
Pickwick Reservoir	0.05242	0.17497	**0.50066**	0.15735	0.04436	**0.25401**	-0.03543	**0.31941**	**0.26577**	**0.25011**	0.00000

Table 8. Control region pairwise F_{ST} P values. Mean probability plus or minus standard deviation. P values <0.005 are significantly different and are marked in **bold**.

	Lake Ontario	Lake Erie	Green Bay	Nelson River	Lake Winnipeg	Lake Sakakawea	Lake Pepin	Lake Maloney	Missouri River	Ohio River	Pickwick Reservoir
Lake Ontario	*										
Lake Erie	0.57227+/- 0.0168	*									
Green Bay	**0.01172+/- 0.0030**	**0.00000+/- 0.0000**	*								
Nelson River	**0.00000+/- 0.0000**	**0.00000+/- 0.0000**	**0.00000+/- 0.0000**	*							
Lake Winnipeg	**0.03320+/- 0.0060**	**0.00000+/- 0.0000**	**0.00000+/- 0.0000**	0.30957+/- 0.0129	*						
Lake Sakakawea	**0.01855+/- 0.0044**	**0.00098+/- 0.0010**	**0.00781+/- 0.0024**	**0.00000+/- 0.0000**	**0.00000+/- 0.0000**	*					
Lake Pepin	0.33984+/- 0.0159	**0.01074+/- 0.0033**	**0.00098+/- 0.0010**	**0.02637+/- 0.0046**	0.46777+/- 0.0168	**0.00293+/- 0.0016**	*				
Lake Maloney	**0.00098+/- 0.0010**	**0.00000+/- 0.0000**	**0.00098+/- 0.0010**	**0.00000+/- 0.0000**	**0.00098+/- 0.0010**	**0.00000+/- 0.0000**	**0.00000+/- 0.0000**	*			
Missouri River	**0.03027+/- 0.0058**	**0.00000+/- 0.0000**	**0.04199+/- 0.0070**	**0.00000+/- 0.0000**	**0.00586+/- 0.0026**	**0.04883+/- 0.0071**	**0.01465+/- 0.0037**	**0.01855+/- 0.0041**	*		
Ohio River	0.10449+/- 0.0095	**0.01074+/- 0.0026**	0.06348+/- 0.0089	**0.00000+/- 0.0000**	**0.00195+/- 0.0014**	0.19043+/- 0.0120	**0.00977+/- 0.0029**	0.39355+/- 0.0158	0.37012+/- 0.0154	*	
Pickwick Reservoir	0.12500+/- 0.0103	**0.00098+/- 0.0010**	**0.00000+/- 0.0000**	**0.00195+/- 0.0014**	0.11035+/- 0.0091	**0.00000+/- 0.0000**	0.93359+/- 0.0056	**0.00000+/- 0.0000**	**0.00195+/- 0.0014**	**0.00000+/- 0.0000**	*

However, Green Bay, the remaining population in the Great Lakes basin is not significantly correlated with Lake Ontario or Lake Erie. The Mississippi drainage displays great diversity.

Phylogenetic trees were formed using parsimony and Bayesian analysis. Groupings with greater genetic relations within each phylogeny were observed, these do not represent monophyletic groups and were not used in statistical analysis.

Development of a phylogenetic tree using Phylip revealed six groups (Table 9; Figure 23, 24). Group A is comprised of three haplotypes. This group contains the haplotypes found at the highest frequency in two Great Lakes populations, Lake Ontario (0.572) and Lake Erie (0.8). Group B is composed of seven haplotypes, the population of Lake Sakakawea (0.6424) is predominantly contained within this group. Group C includes nine haplotypes. The haplotypes of Nelson River (0.933), Lake Winnipeg (0.75), Lake Pepin (0.5714) and Pickwick Reservoir (0.5286) are predominantly contained within this group. Group D, composed of only two haplotypes, contains the predominant haplotype of Lake Maloney (0.5). The population of the Missouri River (0.5625) is predominant within group E, which is formed by fourteen haplotypes. Group F is composed of eight haplotypes, the majority of the population of Green Bay (0.7495) is found within this group. No single population is found within all six groups, Ohio River is the most widespread, found in all groups with the exception of group 3. All groups contain haplotypes originating from the ancestral Mississippi- Teays system. Group B and a portion of group E contain haplotypes which likely originated from Missourian refugia.

Bayesian analysis was used to form phylogenetic trees using the method of maximum likelihood (Table 10; Figure 25, 26). The cladogram developed was similar to that formed using the parsimony method, also being comprised of six groups. Group A, like that of parsimony analysis, is formed by three haplotypes. One haplotype differs, resulting in Pickwick Reservoir and Lake

Table 9. Haplotype frequency based on Phylip analysis of the control region

Frequency of haplotypes with each group as determined by parsimony analysis using Phylip.

This table corresponds to Figure 23.

Sample location	Group A (h0005, h0631, h6108)	Group B (h0003, h0011, h0665, h6130, h6134, h6137, h6149)	Group C (h0001, h0004, h0022, h0609, h0633, h0638, h0639, h0655, h6133)	Group D (h0607, h0007)	Group E (h0014, h0021, h0024, h0621, h0636, h0641, h0642, h0644, h0645, h0647, h0654, h0668, h6104, h6105)	Group F (h0013, h6103, h6100, h0696, h0695, h0627, h0012, h0612)
Lake Ontario	0.572		0.286			0.143
Lake Erie	0.8		0.05			0.15
Green Bay	0.1875	0.0625				0.7495
Nelson River		0.0667	0.933			
Lake Winnipeg		0.1875	0.75		0.0625	
Lake Sakakawea		0.6424			0.3572	
Lake Pepin	0.0714	0.286	0.5714		0.0714	
Lake Maloney			0.125	0.5	0.375	
Missouri River	0.1875		0.25		0.5625	
Ohio River	0.143	0.143		0.286	0.286	0.143
Pickwick Reservoir		0.4116	0.5286	0.0588		

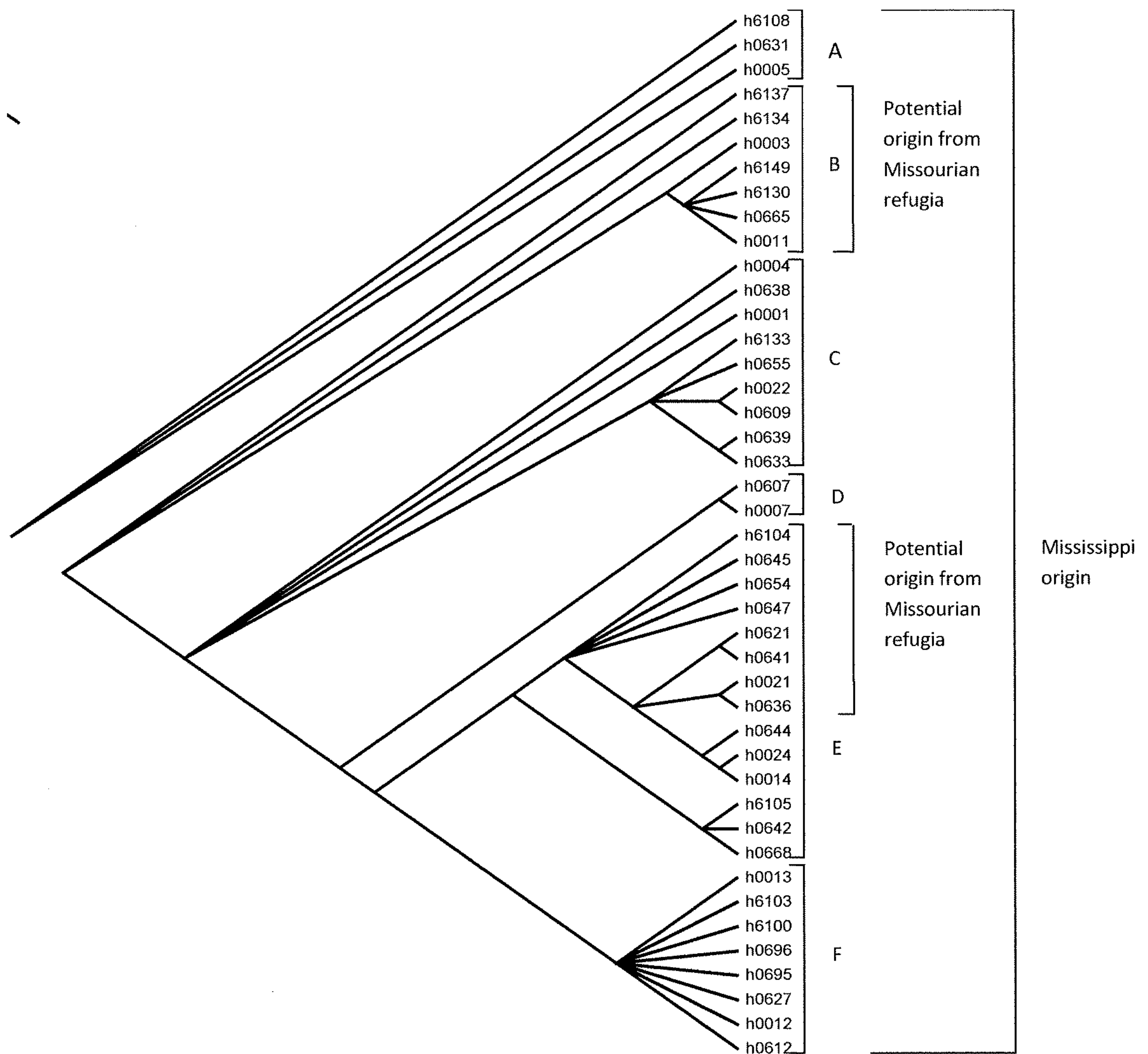


Figure 23. Control region cladogram based on parsimony analysis.

Cladogram based on control region analysis using Phylip. Haplotypes are indicated by "h___", capital letters refer to individual groups.

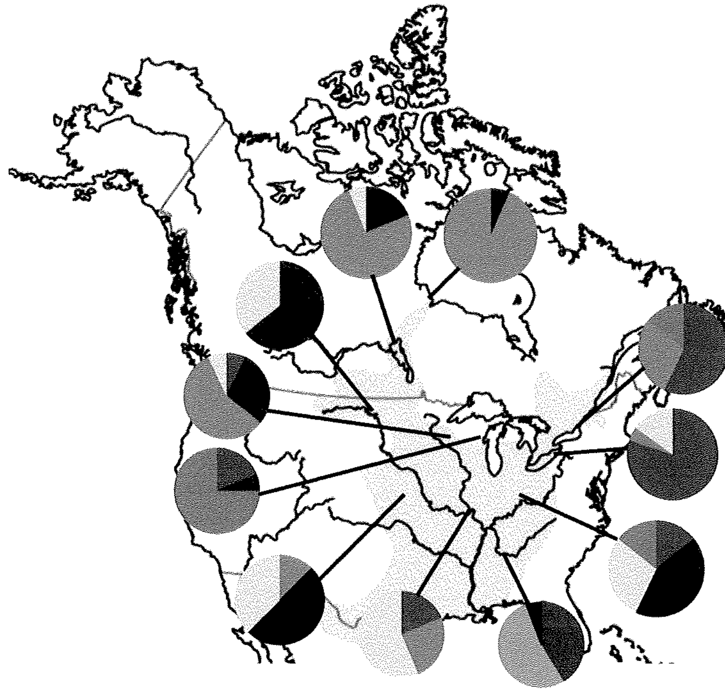


Figure 24. Haplotype frequency based on parsimony analysis of the control region.

Frequency of haplotypes within each group as determined by parsimony analysis. Group 1 = blue; group 2 = red; group 3 = green; group 4 = black; group 5 = yellow; group 6 = orange. This figure corresponds to Table 9.

Table 10. Haplotype frequency based on Bayesian analysis of the control region.

Frequency of haplotypes with each group as determined by Bayesian maximum likelihood analysis. This table corresponds to Figure 25.

	Group A (h0005, h6108, h6137)	Group B (h0001, h0004, h0022)	Group C (h0012, h0013, h0014, h0021, h0024, h0612, h0621, h0627, h0636, h0641, h0642, h0644, h0645, h0647, h0654, h0668, h0695, h0696, h6100, h6103, h6104, h6105)	Group D (h0607, h0007)	Group E (h0609, h0633, h0638, h0639, h0655, h6133)	Group F (h0003, h0011, h0631, h0665, h6130, h6134, h6149)
Lake Ontario	0.429		0.143		0.286	0.143
Lake Erie	0.8		0.15		0.05	
Green Bay	0.188		0.7495			0.0625
Nelson River		0.8			0.133	0.0667
Lake Winnipeg		0.625	0.0625		0.125	0.1875
Lake Sakakawea	0.214		0.3572			0.4284
Lake Pepin	0.143	0.357	0.0714		0.2644	0.2146
Lake Maloney		0.0625	0.375	0.5	0.0625	
Missouri River	0.1875	0.1875	0.5625			
Ohio River	0.143		0.429	0.286		0.143
Pickwick Reservoir	0.176	0.176		0.0588	0.3526	0.2356

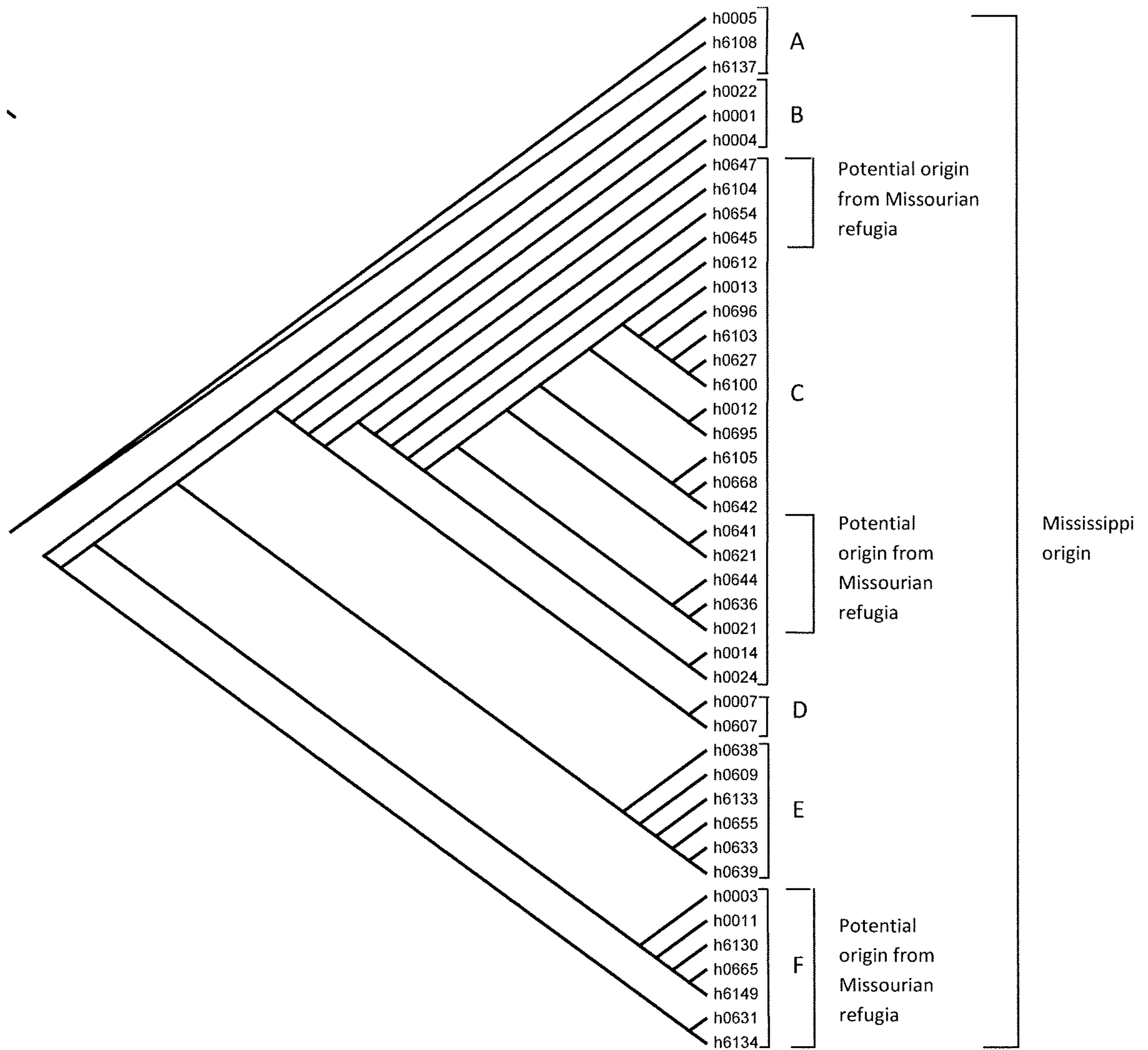


Figure 25. Control region cladogram based on maximum likelihood analysis.

Cladogram based on control region Bayesian analysis. Haplotypes are indicated by "h___", capital letters refer to individual groups.

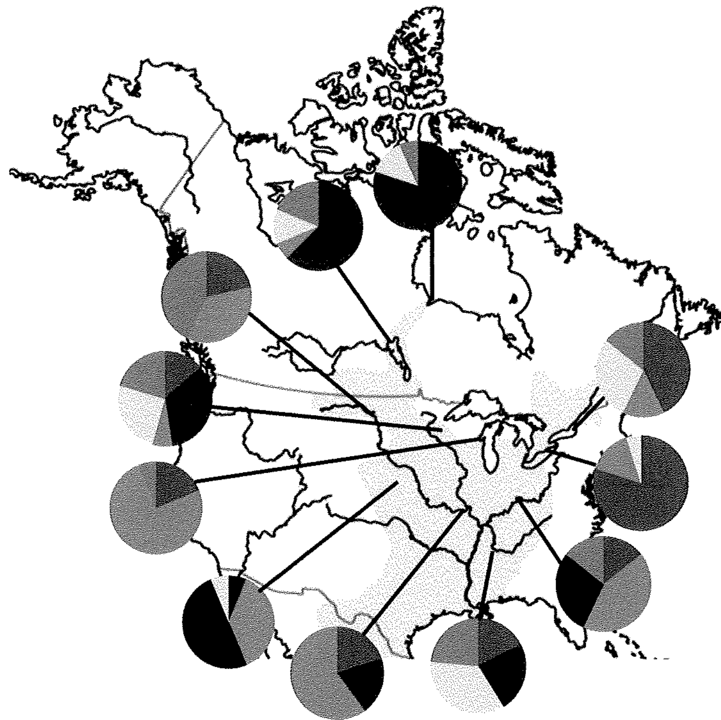


Figure 26. Haplotype frequency based on Bayesian analysis of the control region.

Frequency of haplotypes with each group as determined by Bayesian maximum likelihood analysis. Group 1 = blue; group 2 = red; group 3 = green; group 4 = black; group 5 = yellow; group 6 = orange. This figure corresponds to Table 10.

Sakakawea being included in this group. As in parsimony analysis, Lake Ontario (0.429) and Lake Erie (0.8) are found at the highest frequency in this group. Group B is composed of three haplotypes (h0001, h0004, h0022). Nelson River (0.8), Lake Winnipeg (0.625), and Lake Pepin (0.357) are found at the highest frequency within this group. Group C is a large group, formed by twenty two haplotypes. All populations, with the exception of Nelson River and Pickwick Reservoir, are members of this group. Green Bay (0.7495), Missouri River (0.5625), and Ohio River (0.429) are found at the highest frequency in this group. Group D contains the predominant haplotype of Lake Maloney (0.5). Group E is composed of six haplotypes, Pickwick Reservoir (0.3526) is found at the highest frequency in this group. Group F, comprised of seven haplotypes, contains all populations with the exception of Lake Erie, Lake Maloney and the Missouri River.

Many similarities exist between the phylogenetic trees developed by parsimony and Bayesian analysis. There is little differentiation between the relationships of haplotypes whether analyzed by parsimony or Bayesian analysis. Group A differs by one haplotype in parsimony and Bayesian analysis. Parsimony analysis included the haplotypes of group B (Bayesian analysis) in group C (parsimony analysis) rather than as a separate group. Group D is identical in both forms of analysis. Group E in Bayesian analysis is composed of six haplotypes, which were included in group C in parsimony analysis. Groups E and F in parsimony analysis are the equivalent of group C in Bayesian analysis. Group F in Bayesian analysis is similar to group B in parsimony analysis, with the addition of haplotype h0631.

Bayesian analysis determined that the model of best fit is the HKY model (Table 11) (Hasegawa et al., 1985). This indicates unequal base frequencies and differentiation in the rates of

Table 11. Bayesian model of best fit based on control region analysis.

The Bayesian model of best fit is highlighted in bold.

Model	Harmonic mean	Probability
FTR	-575.53	0.000
HKY	-523.75	**0.942**
GTR	-541.93	0.000
JC	-581.87	0.000
K2P	-531.10	0.000
SYM	-526.55	0.057

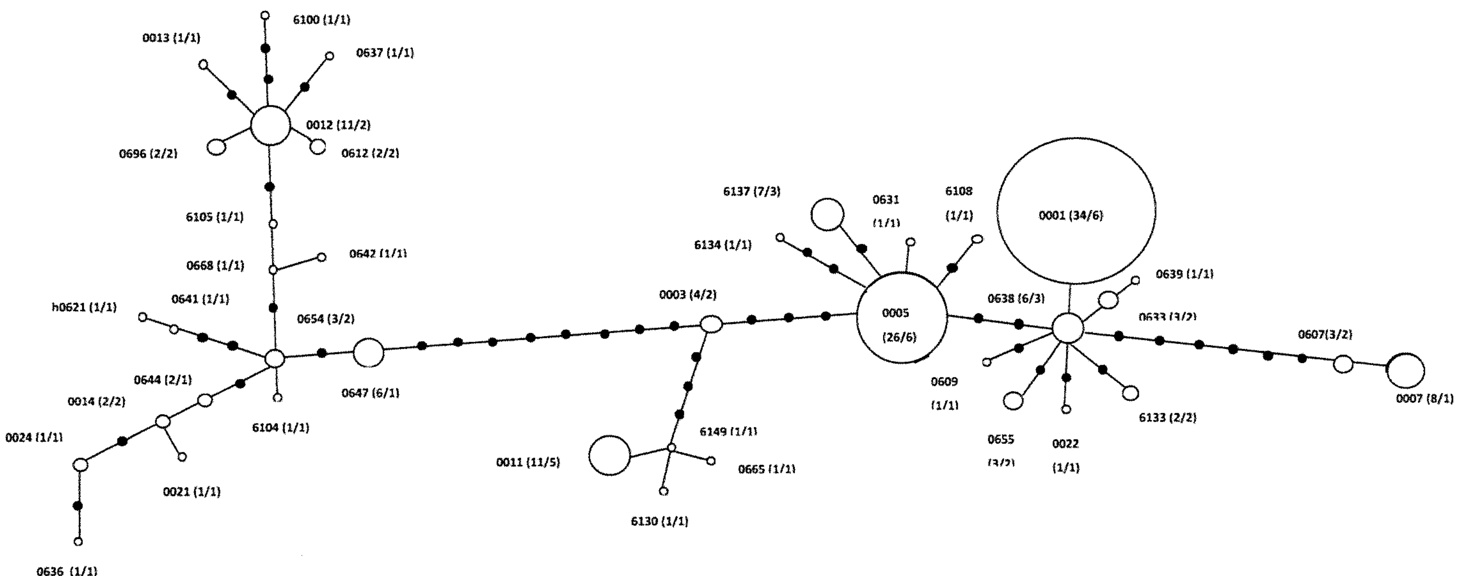


Figure 27. Control region minimum spanning network.

Black circles represent haplotypes that were not sampled or no longer exist. Haplotypes are noted as four digit numbers, the numbers below each haplotype denote the number of individuals with that haplotype, followed by the number of populations containing at least one individual with that haplotype. Circle size is representative of the number of haplotypes represented in the sample.

substitution for transitions and transversions. Each transition rate is equal and each transversion rate is equal.

The minimum spanning network developed indicates periods of isolation in multiple populations followed by dispersal (Figure 27). The form of the network is similar to results of parsimony analysis. There is great genetic distance between the two primary haplotypes of Lake Maloney and the main population.

ND2 region

Intra-population analysis

The sequence obtained from the ND2 region is approximately 500 base pairs. Ten haplotypes were found among the specimens sampled. Diversity indices in the ND2 region were greatest in Lake Maloney (3.513), and lowest in Lake Winnipeg (0.000), Nelson River (0.250) and the Pickwick Reservoir (0.250) (Table 12).

Haplotypes were shared among every population and the frequency of haplotypes varied greatly among populations (Appendix 4). Haplotype h1 is the predominant haplotype of all populations with the exception of Green Bay (0.267) and Lake Maloney (0.125) (Table 13). The h7 haplotype is the predominant haplotype of Lake Maloney with a frequency of 0.5, and is shared only with Pickwick Reservoir (0.0588). The h12 haplotype is the predominant haplotype of Green Bay with a frequency of 0.6 and is shared with specimens from Lake Ontario, Lake Erie, Lake Winnipeg, Lake Maloney and Missouri River.

Mismatch distribution analysis was significant for Lake Ontario under the demographic

Table 12. Nucleotide diversity of the ND2 region of freshwater drum.

Nucleotide diversity was calculated based on the number of differences between all pairs of haplotypes in each sample. Samples with higher values of π are more diverse. Samples marked with an asterisk* are significantly higher or lower than the mean value.

Sample location	Number of haplotypes	Nucleotide diversity (π)
Lake Ontario	2	1.429
Lake Erie	3	0.768
Green Bay	3	2.343*
Nelson River	2	0.248*
Lake Winnipeg	1	0.556
Lake Sakakawea	2	1.513
Lake Pepin	3	0.914
Lake Maloney	4	2.992*
Missouri River	7*	2.477*
Ohio River	2	0.667
Pickwick Reservoir	3	0.235*
Mean	**2.9**	**1.123**

Table 13. Shared ND2 region haplotypes among sampled populations.

Number of haplotypes shared among each population sampled.

Sample site	Lake Ontario	Lake Erie	Green Bay	Nelson River	Lake Winnipeg	Lake Sakakawea	Lake Pepin	Lake Maloney	Missouri River	Ohio River	Pickwick Reservoir
Lake Ontario	*	2	2	1	2	1	1	2	2	1	1
Lake Erie	2	*	2	2	2	2	2	2	3	2	2
Green Bay	2	2	*	1	2	1	1	2	3	1	1
Nelson River	1	2	1	*	1	2	2	1	2	2	2
Lake Winnipeg	2	2	2	1	*	1	1	2	2	1	1
Lake Sakakawea	1	2	1	2	1	*	2	1	3	2	2
Lake Pepin	1	2	1	2	1	2	*	1	2	2	2
Lake Maloney	2	2	2	1	2	1	1	*	3	1	2
Missouri River	2	3	3	2	2	3	2	3	*	2	2
Ohio River	1	2	1	2	1	2	2	1	2	*	2
Pickwick Reservoir	1	2	1	2	1	2	2	2	2	2	*

Table 14. ND2 region mismatch distribution. Column 2 illustrates the distribution of differences between pairs of haplotypes. Columns 3 and 5 illustrate tau values under the demographic or spatial expansion models, and columns 4 and 6 illustrate goodness of fit for these models. P < 0.05 is statistically significant and marked in **bold**.

	Mismatch observed variance	Tau – Demographic expansion	Demographic expansion model, P (sum of squared deviation)	Tau – Spatial expansion	Spatial expansion model, P (sum of squared deviation)
Lake Ontario	5.357	3.00	**0.023**	5.628	0.332
Lake Erie	2.348	3.00	0.389	0.396	0.066
Green Bay	6.054	6.398	0.098	5.313	0.154
Nelson River	0.188	2.930	0.111	0.305	0.291
Lake Winnipeg	2.485	3.000	0.066	5.627	0.197
Lake Sakakawea	4.721	3.250	0.109	5.347	0.315
Lake Pepin	4.021	3.000	0.195	6.521	0.371
Lake Maloney	7.302	6.986	0.092	5.761	0.135
Missouri River	4.330	5.461	0.091	4.617	0.206
Ohio River	0.267	1.053	0.300	1.050	0.210
Pickwick Reservoir	0.196	3.000	0.478	0.265	0.552

expansion model with a P value of 0.024 (Table 14). No other population had a significant value.

Limited or non-existent values of linkage disequilibrium exist in Lake Ontario, Lake Erie, Pickwick Reservoir and Missouri River. Green Bay, Lake Pepin and Lake Maloney show higher levels. Linkage disequilibrium could not be calculated for a single locus in the remaining sites.

Analysis of Tajima's D shows significant P values for Lake Ontario (0.025), Lake Erie (0.017), Lake Winnipeg (0.009) and Lake Pepin (0.026) (Table 15). The value of Tajima's D for each is negative.

AMOVA analysis based on each population individually shows most variation is found within populations (70.97%) (Table 16a). Based upon drainage basins, most variation (92.03%) is also found within populations (Table 16b).

ND2 region

Inter-population analysis

F_{ST} value as determined by AMOVA for the total population is 0.29027, and for drainage basins is 0.07974 (Table 16a, 16b).

F_{ST} values revealed Green Bay to be significantly different from all populations with the exception of Lake Maloney and Missouri River (Table 17, 18). Green Bay is the only population from which Lake Ontario, Lake Sakakawea and Lake Pepin differ significantly.

The phylogenetic tree developed using Phylip revealed two groups (Table 19; Figure 28, 29). Group A, comprised of haplotypes h1, h7 and h15 is the most predominant, specimens from each sample location are found within this group. The highest frequency haplotypes from all populations with the exception of Green Bay are found in this group. Group B contains seven

Table 15. ND2 region analysis of Tajima's D test of neutrality. P < 0.05 are statistically significant are marked in **bold**. Negative Tajima's D values indicate population size expansion, positive selection or a recent bottleneck.

Sample location	Tajima's D	Tajima's D, P=(D simulated < D observed)
Lake Ontario	-1.48614	**0.03500**
Lake Erie	-1.73887	**0.01800**
Green Bay	0.94402	0.81500
Nelson River	-0.39883	0.30400
Lake Winnipeg	-1.95558	**0.00400**
Lake Sakakawea	-0.21776	0.44600
Lake Pepin	-1.76624	**0.02900**
Lake Maloney	2.24034	0.99300
Missouri River	1.38296	0.91600
Ohio River	1.63299	0.96900
Pickwick Reservoir	-1.50358	0.05100

Table 16. ND2 region AMOVA analysis of populations.

(a) Analysis of the variation among and within populations.

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of variation
Among populations	10	46.183	0.27718 Va	29.03
Within Populations	147	99.627	0.67773 Vb	70.97
Total	157	145.810	0.95492	
Fixation Index FST: 0.29027				

(b) Analysis of the variation among and within drainage basins.

Source of variation	Degrees of freedom	Sum of squares	Variance components	Percentage of variation
Among populations	2	8.040	0.06732 Va	7.97
Within Populations	155	120.428	0.77696 Vb	92.03
Total	157	128.468	0.84428	
Fixation Index FST: 0.07974				

Table 17. Population pairwise F_{ST} values based on ND2 region analysis. 0.00000 indicates no differentiation between populations, 1.00000 indicates complete differentiation. Negative values should be considered as 0.00000. Theoretically, all values >0.25000 indicate significant differentiation and are highlighted in **bold**.

	Lake Ontario	Lake Erie	Green Bay	Nelson River	Lake Winnipeg	Lake Sakakawea	Lake Pepin	Lake Maloney	Missouri River	Ohio River	Pickwick Reservoir
Lake Ontario	0.0000										
Lake Erie	-0.02631	0.00000									
Green Bay	**0.42746**	**0.61049**	0.00000								
Nelson River	0.08618	-0.03626	**0.66770**	0.00000							
Lake Winnipeg	-0.05325	-0.01922	**0.62085**	0.01829	0.00000						
Lake Sakakawea	-0.06377	-0.01430	**0.47170**	0.05076	0.03248	0.00000					
Lake Pepin	-0.04508	-0.05075	**0.57589**	-0.02925	-0.02580	-0.03579	0.00000				
Lake Maloney	0.08388	**0.27560**	0.22677	**0.32980**	**0.26879**	0.15671	0.23816	0.00000			
Missouri River	0.06812	0.23099	0.17579	**0.29364**	**0.25226**	0.07658	0.19291	0.06661	0.00000		
Ohio River	0.09361	0.02810	**0.54973**	0.20620	0.23727	-0.04064	0.02116	0.22436	0.14306	0.00000	
Pickwick Reservoir	0.08549	-0.00822	**0.68262**	-0.02602	-0.00135	0.08697	-0.00622	**0.32720**	**0.31940**	**0.33267**	0.00000

Table 18. ND2 region pairwise F_{ST} P values. Mean probability plus or minus standard deviation. P values <0.005 are significantly different and are marked in **bold**.

	Lake Ontario	Lake Erie	Green Bay	Nelson River	Lake Winnipeg	Lake Sakakawea	Lake Pepin	Lake Maloney	Missouri River	Ohio River	Pickwick Reservoir
Lake Ontario	*										
Lake Erie	0.57617+- 0.0139	*									
Green Bay	**0.01758+- 0.0034**	**0.00000+- 0.0000**	*								
Nelson River	0.19434+- 0.0119	0.99902+- 0.0002	**0.00000+- 0.0000**	*							
Lake Winnipeg	0.99902+- 0.0002	0.59668+- 0.0145	**0.00000+- 0.0000**	0.31934+- 0.0155	*						
Lake Sakakawea	0.61035+- 0.0153	0.47168+- 0.0140	**0.00000+- 0.0000**	0.32715+- 0.0152	0.26270+- 0.0150	*					
Lake Pepin	0.78125+- 0.0109	0.99902+- 0.0002	**0.00000+- 0.0000**	0.99902+- 0.0002	0.59766+- 0.0122	0.58984+- 0.0138	*				
Lake Maloney	0.13965+- 0.0100	**0.00195+- 0.0014**	**0.01270+- 0.0042**	**0.00098+- 0.0010**	**0.00391+- 0.0019**	**0.04199+- 0.0057**	**0.00684+- 0.0023**	*			
Missouri River	0.18262+- 0.0131	**0.00586+- 0.0022**	**0.02539+- 0.0047**	**0.00293+- 0.0016**	**0.00195+- 0.0014**	0.11523+- 0.0116	**0.02148+- 0.0049**	0.12402+- 0.0110	*		
Ohio River	0.27930+- 0.0149	0.30859+- 0.0123	**0.00293+- 0.0016**	0.13867+- 0.0115	0.21680+- 0.0121	0.51270+- 0.0171	0.28809+- 0.0159	0.07812+- 0.0073	0.19922+- 0.0133	*	
Pickwick Reservoir	0.29297+- 0.0121	0.53516+- 0.0157	**0.00000+- 0.0000**	0.58496+- 0.0134	0.59473+- 0.0102	0.12598+- 0.0081	0.47754+- 0.0140	**0.00000+- 0.0000**	**0.00098+- 0.0010**	0.10156+- 0.0074	*

Table 19. Haplotype frequency based on Phylip analysis of the ND2 region

Frequency of haplotypes with each group as determined by Phylip parsimony analysis. This table corresponds to Figure 28.

	Group A (h1, h7, h15)	Group B (h12, h26, h41, h46 h82, h113, h131)
Lake Ontario	0.857	0.143
Lake Erie	0.944	0.056
Green Bay	0.308	0.692
Nelson River	1.00	
Lake Winnipeg	1.00	
Lake Sakakawea	1.00	
Lake Pepin	0.9286	0.0714
Lake Maloney	0.616	0.384
Missouri River	0.6665	0.3335
Ohio River	1.00	
Pickwick Reservoir	1.00	

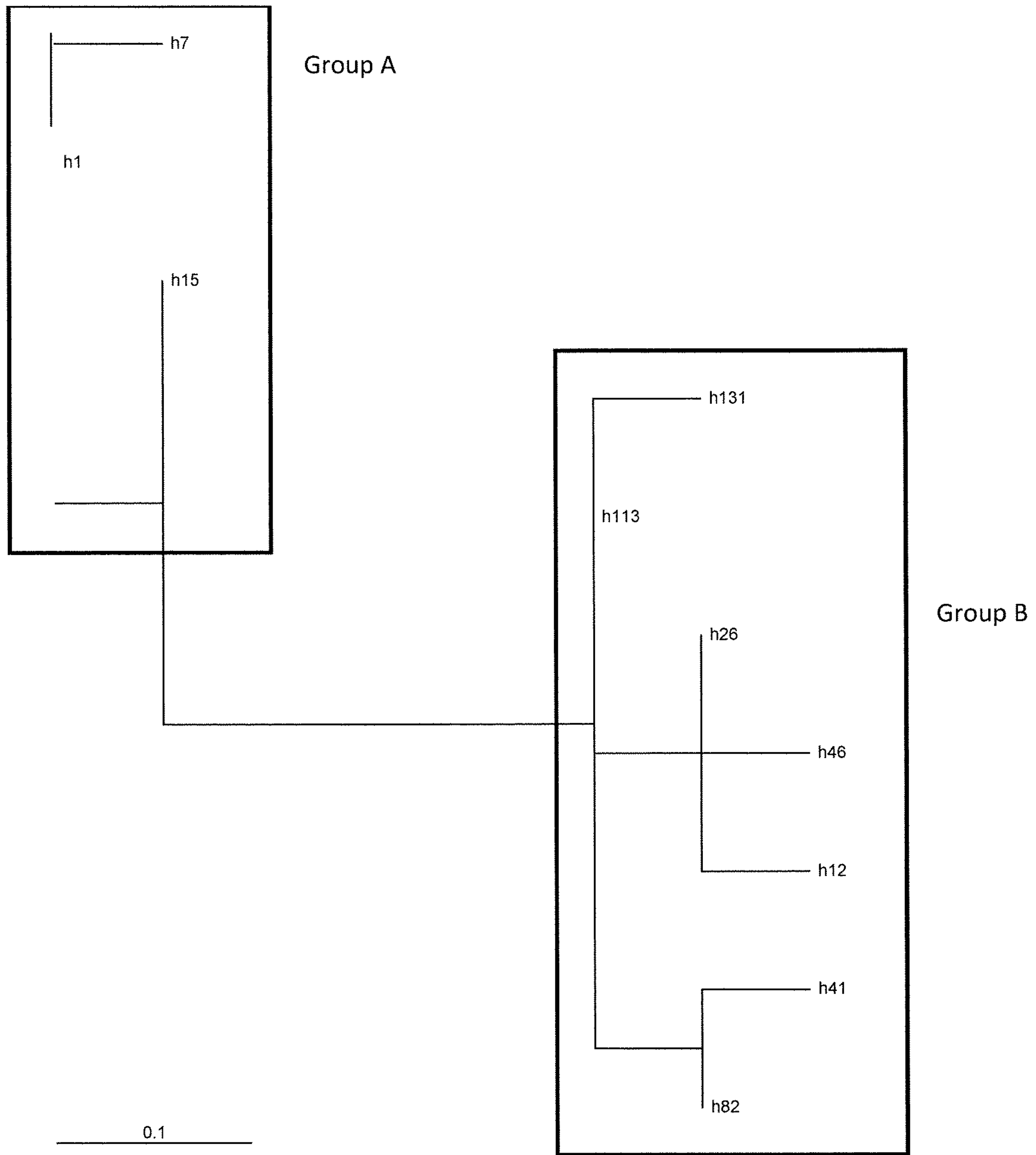


Figure 28. ND2 region phylogram based on parsimony analysis.

Phylogram based on ND2 region analysis using Phylip. Haplotypes are indicated by "h___".

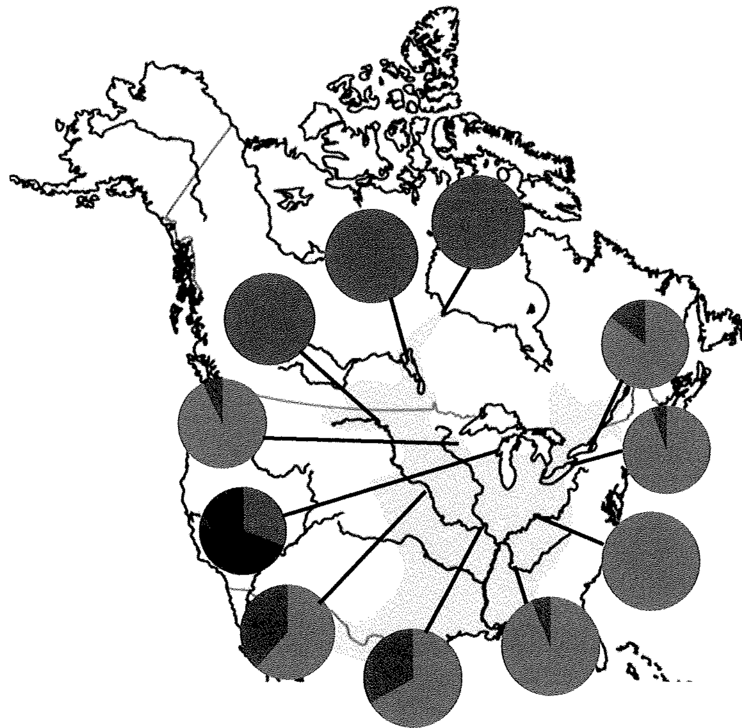


Figure 29. Haplotype frequency based on parsimony analysis of the ND2 region.

Frequency of haplotypes with each group as determined by parsimony analysis. Group A = blue; group B = red. This figure corresponds to Table 19.

different haplotypes, representing all sample locations with the exception of Nelson River, Ohio River and Pickwick Reservoir. This group contains haplotype h12, the most common haplotype in Green Bay. Group A is associated with Mississippian origins, whereas group B appears to be associated with Missourian refugia.

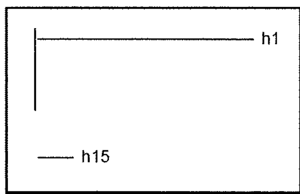
Phylogenetic trees were also formed using Bayesian analysis (Table 20; Figure 30, 31) with great similarity to results derived from parsimony analysis. The primary difference between the results of the two forms of analysis is that haplotype h7 is transferred to group B in Bayesian analysis. Group A appears to be associated with Mississippi refugia and group B is associated with Missourian refugia. Bayesian analysis revealed that the GTR model is the best fit for the ND2 region (Table 21) (Tavaré, 1986). This indicates that base frequencies are unequal as are transition and transversion rates. Individual transition and transversion rates differ from each other.

The minimum spanning network indicates isolation between the primary population and a small population which persisted in the Missourian refugia. Dispersal subsequent to isolation appears to have occurred in both the main and smaller population (Figure 32).

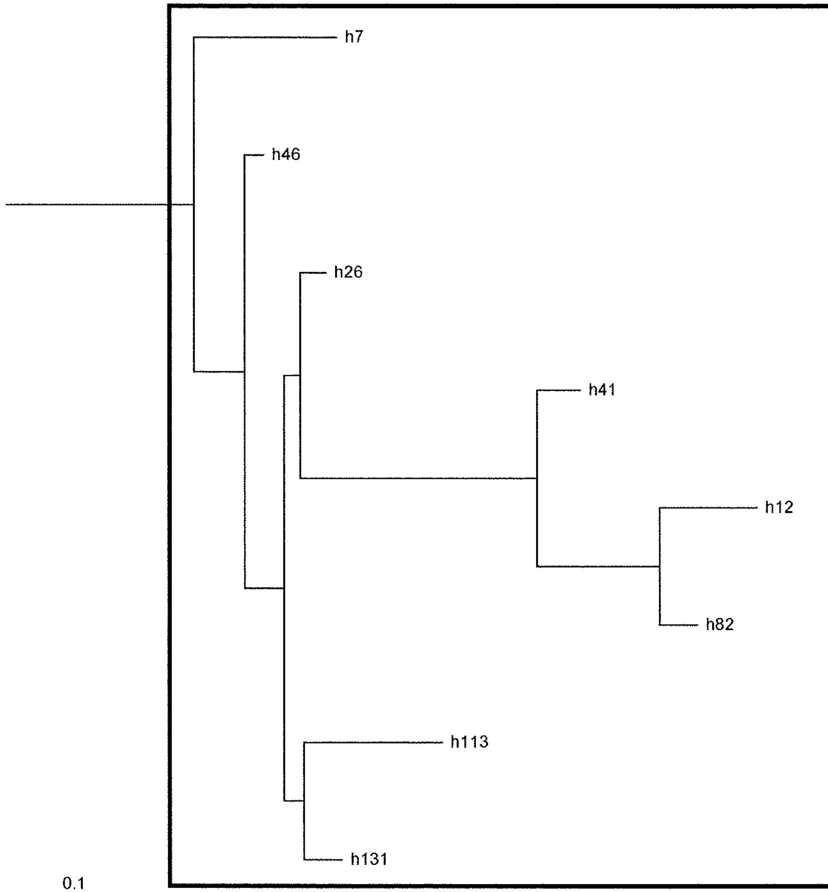
Table 20. Haplotype frequency based on Bayesian analysis of the ND2 region

Frequency of haplotypes with each group as determined by Bayesian maximum likelihood analysis. This table corresponds to Figure 30.

	Group A (h1, h15)	Group B (h7, h12, h26, h41, h46 h82, h113, h131)
Lake Ontario	0.857	0.143
Lake Erie	0.944	0.056
Green Bay	0.308	0.692
Nelson River	1.00	
Lake Winnipeg	1.00	
Lake Sakakawea	1.00	
Lake Pepin	0.9286	0.0714
Lake Maloney	0.154	0.846
Missouri River	0.6665	0.3335
Ohio River	1.00	
Pickwick Reservoir	0.9375	0.0625



Group A



Group B

Figure 30. ND2 region phylogram based on Bayesian analysis.

Phylogram based on ND2 region analysis using Phylip. Haplotypes are indicated by "h___".

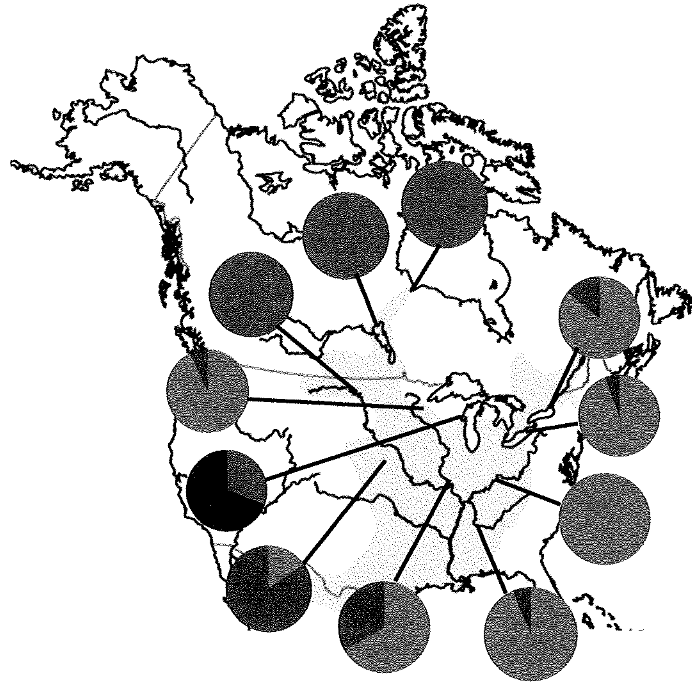


Figure 31. Haplotype frequency based on Bayesian analysis of the ND2 region.

Frequency of haplotypes with each group as determined by Bayesian maximum likelihood analysis. Group A = blue; group B = red. This figure corresponds to Table 19.

Table 21. Bayesian model of best fit based on ND2 region analysis.

The Bayesian model of best fit is highlighted in bold.

Model	Harmonic mean	Probability
FTR	-92.25	0.001
HKY	-88.31	0.057
GTR	-85.76	**0.732**
JC	-94.24	0.000
K2P	-87.23	0.168
SYM	-88.64	0.041

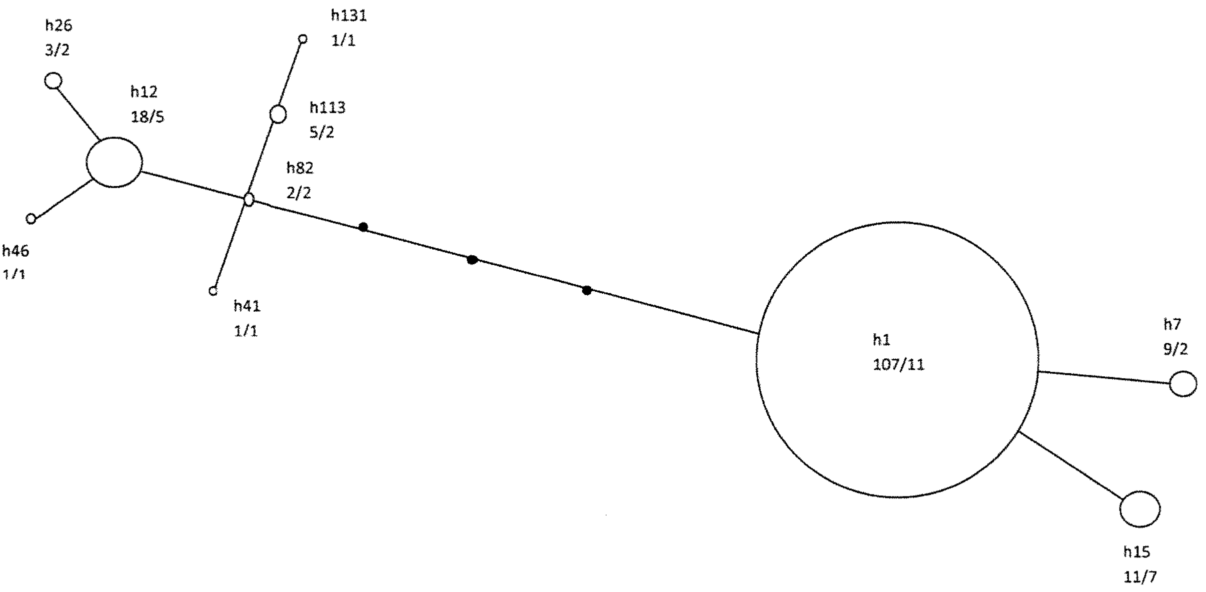


Figure 32. ND2 minimum spanning network.

Black circles represent haplotypes that were not sampled or no longer exist. Haplotype numbers are written as "h_", the numbers below each haplotype denote the number of individuals with that haplotype, followed by the number of populations containing at least one individual with that haplotype. Circle size is representative of the number of haplotypes represented in the sample.

DISCUSSION

The differing mutation rates of the mitochondrial ND2 and control regions results in differences in what they reveal about the history of freshwater drum (Villablanca, 1994). The ND2 region is better suited to identify the historical demography and relationships of lineages. The control region is best suited to analysis of intraspecies demography and relationships over a shorter period of time, such as the last 18000 years since the last glacial maximum (Day et al., 2007). Data from these two regions may appear to be contradictory, but complement each other. In conjunction, these data can be useful in inferring both present day and historical demography (Lopez et al. 2007).

The ND2 region of freshwater drum mitochondrial DNA was successfully amplified using primers developed from the *Aplodinotus grunniens* sequence from GenBank (accession number AY225720). The freshwater drum control region was successfully amplified using red drum primers (Gold et al., 1993). Primers specific to the freshwater drum mitochondrial DNA control region were subsequently developed in order to increase the specificity of amplification. Polymorphisms in both the ND2 and control regions were identified revealing multiple haplotypes, particularly in the control region.

I observed genetic variability in the ND2 region at a much lower level than that of the control region. The difference in mutation rate between the ND2 and control region gene of freshwater drum observed in my study is much greater than that observed in other studies (Lopez et. al, 1997; Koblmüller et al., 2010). Average nucleotide diversity in my study in the control region is 6.391, and 1.123 in the ND2 region (Table 2, 12). This reflects a mutation rate approximately 5.6 times greater in the control region. Studies identify the average divergent rate to be between 1.95 and 2.90 times greater for the control region (Lopez et. al, 1997). Limited information

exists on sequence divergence in fishes. A study of the cichlid tribe Tropheini observed the divergence rate to be 6.97% for the ND2 region and 7.10% for the control region (Koblmüller et al., 2010). The ND2 divergence rate observed in freshwater drum is much less than observed in Tropheini, while the divergence rate of the control region is only slightly less.

Nucleotide diversity can be used to infer the relative age of a population. Populations that are older would be expected to exhibit higher genetic diversity, whereas recently colonized populations would be expected to display lower genetic diversity (Hewitt 1996, 2000). The Ohio River and Missouri River are primary tributaries of the Mississippi River. Portions of these large drainages were not glaciated, thus higher levels of genetic diversity would be predicted in fish which inhabited these regions before and during periods of glaciation. These predictions are met in populations of hogsucker (Berendzen et al. 2008) inhabiting non-glaciated regions, similar results were found in other studies of freshwater fish populations primarily inhabiting central and eastern North America, including gilt darter (Near et al., 2001), rainbow darter (Ray et al., 2006) and walleye (Stepien et al., 2009). Increased nucleotide diversity is observed in the control region for both the Ohio and Missouri River freshwater drum populations, the Missouri River population also displays a high level of diversity in ND2 region analysis. These data suggest that the Ohio and Missouri River populations were founded prior to the Pleistocene and survived glaciation in these regions. The Nelson River displayed the lowest level of genetic diversity (1.905 (CR), 0.248 (ND2)), as expected for a region which was previously glaciated and was likely the last of my sample populations to be colonized due to its northern location (Table 2, 12). Decreased genetic diversity was also detected in Lake Erie, Lake Winnipeg and Lake Pepin accounting for some, but not all, previously glaciated areas (Figure 33). My results are similar to other studies and somewhat conform to phylogeographic predictions of lower genetic diversity in fishes displaced by glaciation (Bernatchez and Wilson, 1998). Green Bay, Lake

Sakakawea and Lake Maloney are regions which exhibit a higher degree of genetic variation than expected (Table 2, 12). These populations are located at or near the leading edge of the ice sheet, thereby may have been able to survive glaciation in these regions. Diversity in the Lake Ontario population is also higher than normal, however this lake could have been accessed by many dispersal routes and this diversity may reflect a population admixture, similar to results reported by Borden and Krebs (2009) in the Lake Ontario smallmouth bass population. Pickwick Reservoir in Alabama displayed levels of genetic diversity lower than expected, it is possible this region was colonized (or recolonized) after glacial retreat as theorized by Berendzen et al. (2008) in their study of bigeye chub. Berendzen et al. (2008) suggest that bigeye chub dispersed from the upper Ohio River both northward and southward (into the Cumberland River) from the upper Ohio River upon glacial retreat. Absence of gene flow between the upper and lower Tennessee Rivers was detected by Berendzen et al. (2008) even though the upper and lower Tennessee Rivers are connected. This genetic separation observed in bigeye chub by Berendzen et al. (2008) exhibits similarity to data from my study of the Pickwick Reservoir population. The reason behind this genetic separation remains unresolved. With the exceptions of Lake Ontario and Pickwick Reservoir my results conform to my prediction 2-1 of hypothesis 2, "populations south of the last glacial maximum will exhibit increased genetic variation".

Shared haplotypes were observed among and between freshwater drum populations (Figure 34, 35). The ND2 region is reflective of the genetic structure of freshwater drum populations prior to the last glacial maximum. ND2 haplotype h1 is shared amongst all populations, indicating common ancestry (Figure 34(a), Appendix 5). The control region is better suited to identify the contemporary population structure of freshwater drum populations because of its higher mutation rate. The distribution of these control region haplotypes is often associated with geographic proximity and shared drainage basins. Control region data reveal a suture line,

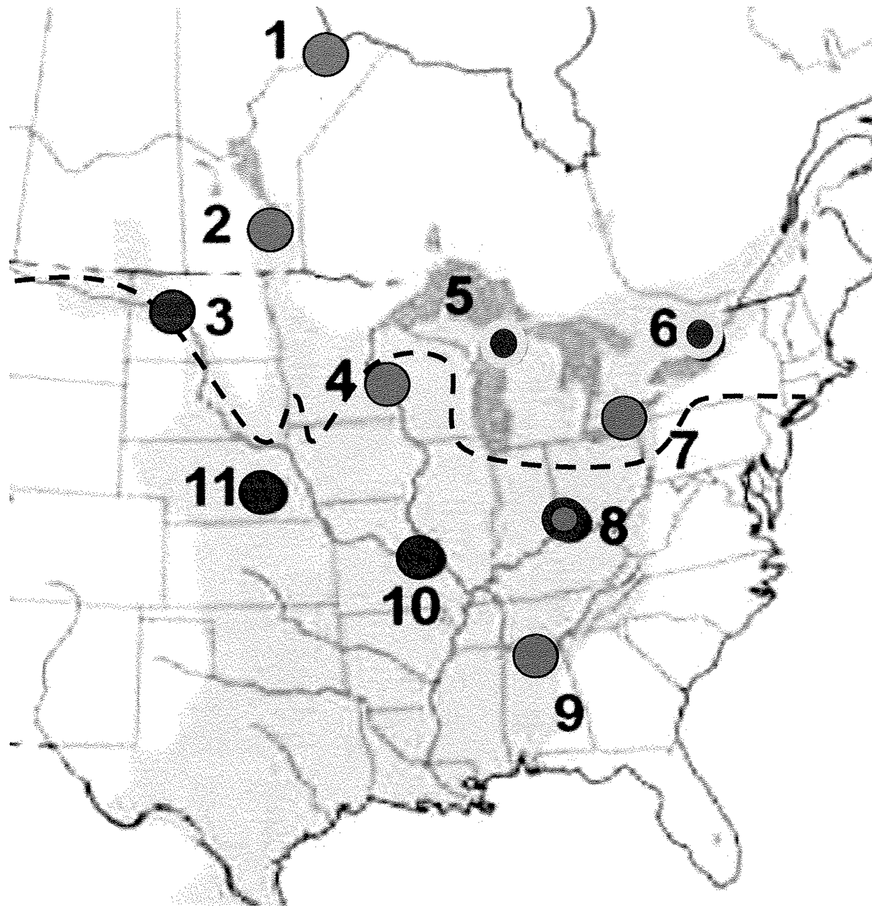


Figure 33. Correlation between latitude and genetic diversity.

There is a trend towards decreased genetic diversity in previously glaciated areas, but exceptions clearly exist. Interior circles indicate ND2 nucleotide diversity, exterior circles indicate control region diversity. Green indicates diversity is below average, yellow indicates diversity is average (within 0.2 of the mean), and red indicates diversity is above average. The dashed line illustrates the extent of the last glacial maximum. Numbers identify sample sites: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB.

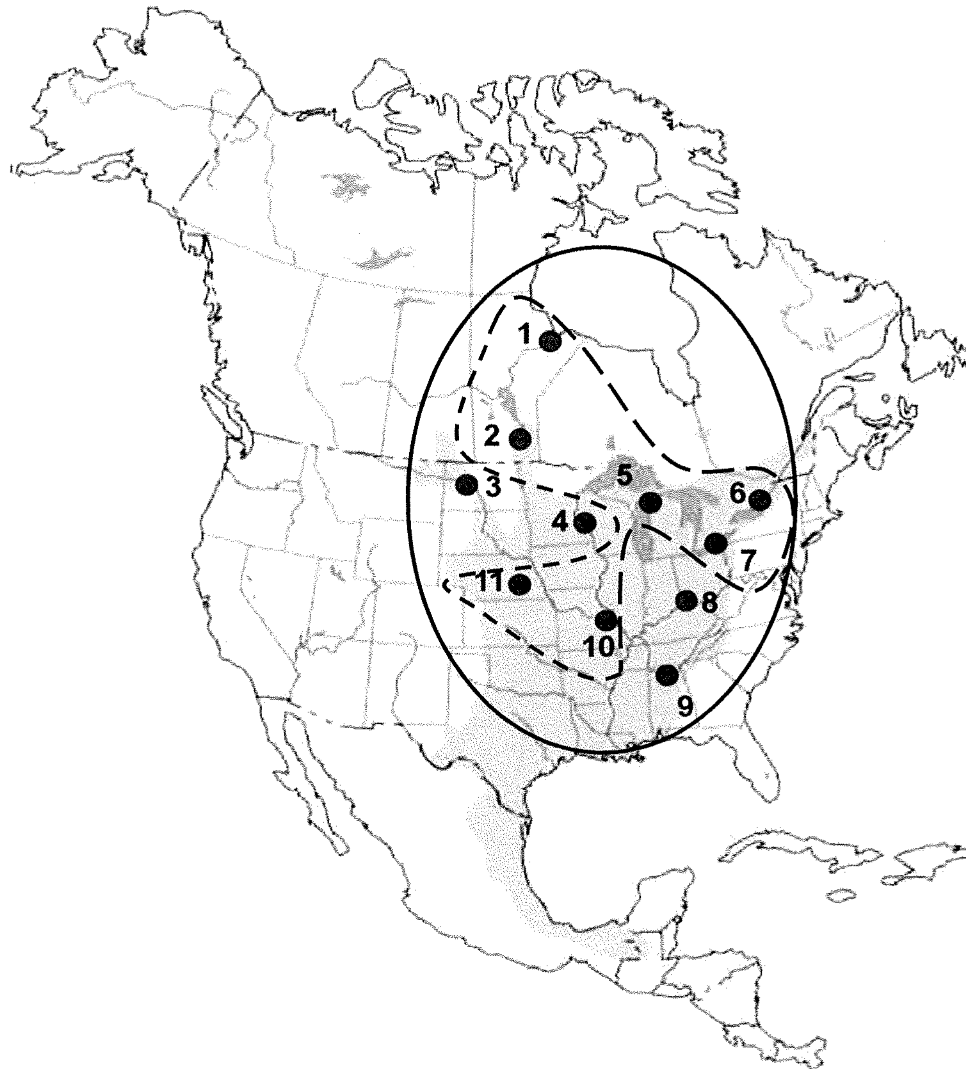


Figure 34(a). Distribution of ND2 region haplotypes, h1 (solid line) and h12 (dashed line).

Illustrated is the distribution of ND2 haplotypes h1 and h12 in sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB.

Figure 34(b). Distribution of ND2 region haplotype h15.

Illustrated is the distribution of ND2 haplotype h15 in the sampled populations (it is likely Lake Winnipeg contains this haplotype but it was not identified in the sample): (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie;(8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB.

Figure 34(c). Distribution of ND2 region haplotypes h82 (solid line) and h26 (dashed line).

Illustrated is the distribution of ND2 haplotypes h82 and h26 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB.

Figure 34(d). Distribution of ND2 region haplotypes h110 (solid line) and h7 (dashed line).

Illustrated is the distribution of ND2 haplotypes h110 and h7 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB;(10)Missouri River, MO; (11) Lake Maloney, NB.

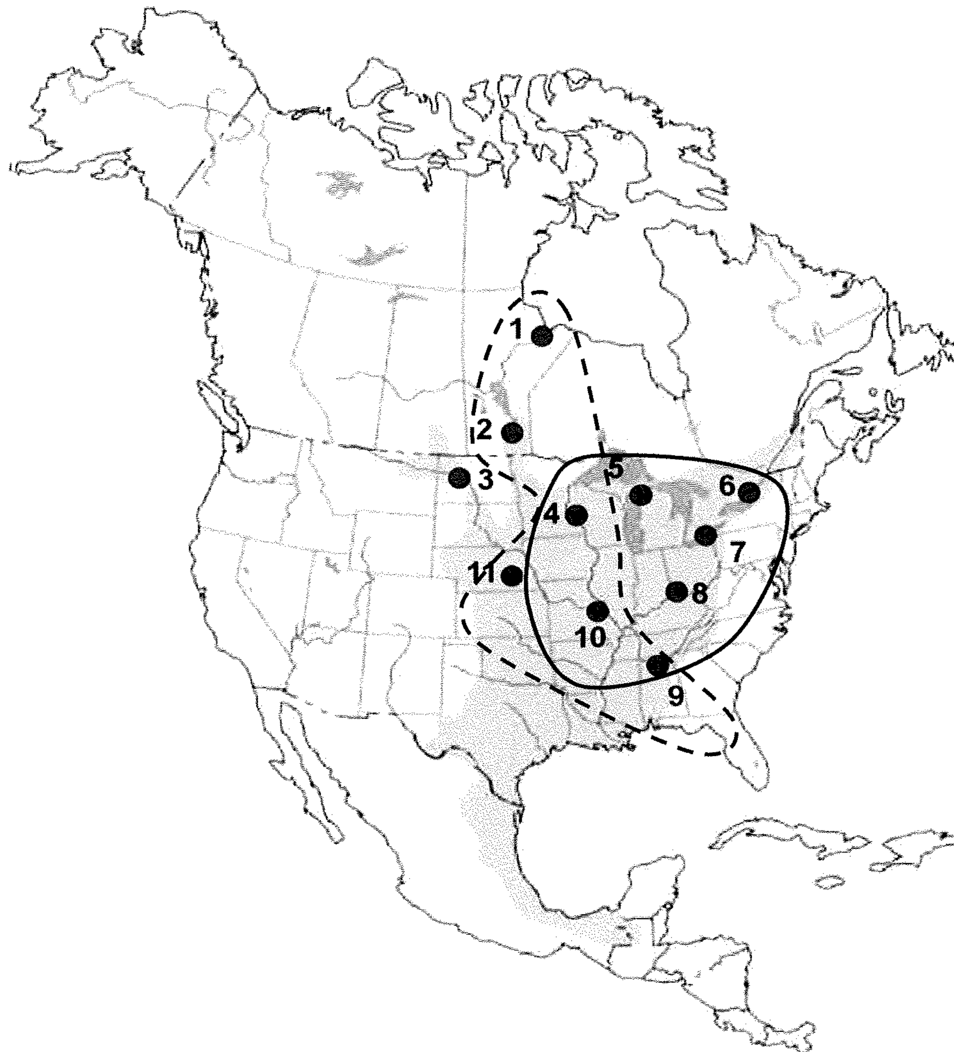


Figure 35(a). Distribution of control region haplotypes h0005 (solid line) and h0001 (dashed line).

Illustrated is the distribution of control region haplotypes h0005 and h0001 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB.

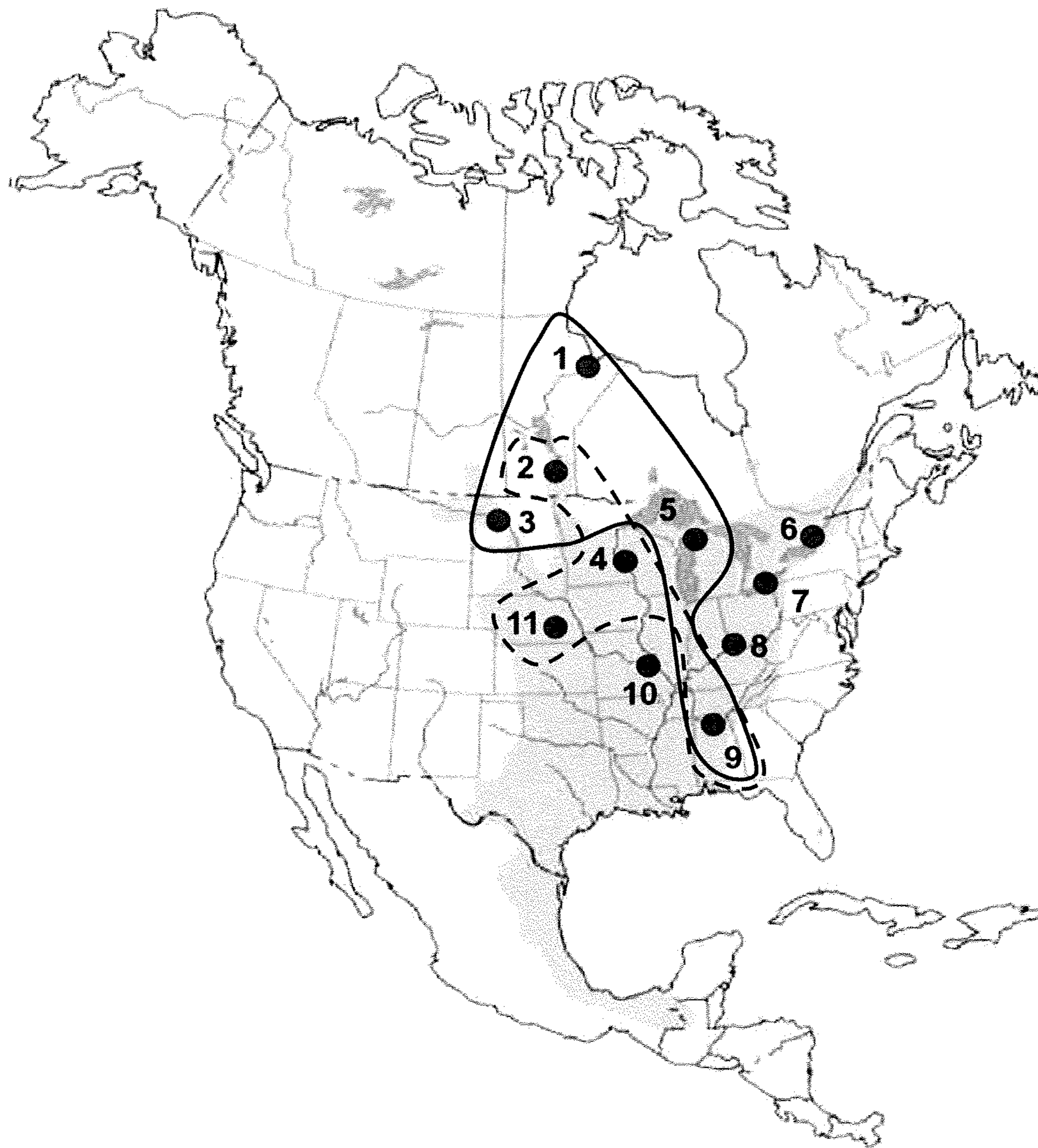


Figure 35(b). Distribution of control region haplotypes h0011 (solid line) and h0638 (dashed line).

Illustrated is the distribution of control region haplotypes h0011 and h0638 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB.

Figure 35(c). Distribution of control region haplotypes h0612 (solid line), h0012 (dashed line) and h0633 (yellow line).

Illustrated is the distribution of control region haplotypes h0612, h0012 and h0633 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB.

Figure 35(d). Distribution of control region haplotypes h0655 (solid line), h0654 (dashed line) and h0607 (yellow line).

Illustrated is the distribution of control region haplotypes h0655, h0654 and h0607 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB.

Figure 35(e). Distribution of control region haplotypes h6133 (solid line), h6137 (dashed line) and h0668 (yellow line).

Illustrated is the distribution of control region haplotypes h6133, h6137 and h0668 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB.

Figure 35(f). Distribution of control region haplotypes h0003 (solid line) and h0014 (dashed line).

Illustrated is the distribution of control region haplotypes h0003 and h0014 in the sampled populations: (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB.

roughly correlated with the course of the Mississippi River, separating the populations into eastern and western groups (Figure 36). This suture line appears to be a common feature of North American freshwater fish species. Both Borden and Krebs (2009) in smallmouth bass populations and Ray et al. (2006) in populations of rainbow darter report a similar feature. A suture line also exists that follows the ice margin of the Wisconsin glaciations in populations west of the Mississippi River, but this is not observed in populations east of the Mississippi River.

Lake Erie and Lake Ontario do not conform to expectations, their population structure may reflect colonization by multiple populations as theorized by Stepien et al. (2009). Pickwick Reservoir shares the most commonalities with populations north of the suture line, further supporting the proposal that this population was recently colonized (Appendix 3).

Mismatch distribution analysis determines if populations have undergone either a demographic or spatial population expansion. Demographic expansion represents an increase in population numbers in the same geographic area resulting in increased population density. By contrast, spatial expansion describes an increase in the size of the habitat populated by the same number of individuals, resulting in decreased density. Tau values measure how long ago the population expansion occurred (Schneider and Excoffier, 1999; Excoffier, 2004). Populations which tested positive for demographic expansion (Lake Sakakawea, Lake Maloney and Lake Ontario) (Table 4, 14) are located near the leading ice edge of the Wisconsin glaciation. We can infer that Lake Sakakawea, Lake Maloney and Lake Ontario were colonized by members of source populations as these areas became accessible. Populations existing in leading edge refugia likely had greater accessibility to newly exposed habitats. However, Lake Pepin, Green Bay and Lake Erie are also located near the leading edge and do not show evidence of demographic expansion. Data from

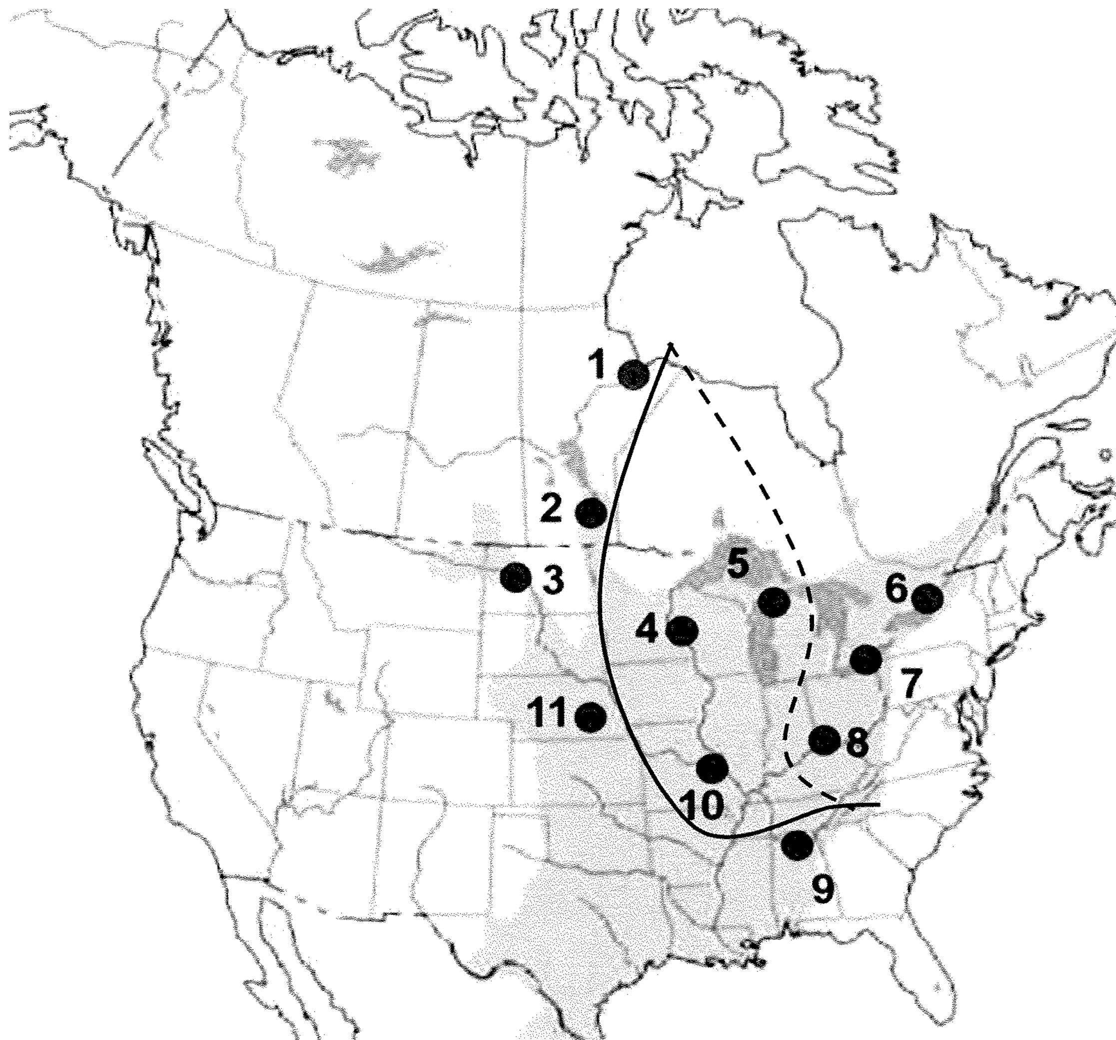


Figure 36. Suture lines as indicated by control region shared haplotypes.

Longitudinal suture lines as evidenced by the presence (or absence) of shared control region haplotypes among sampled freshwater drum populations. Two approximations were identified (solid line and dotted line. Sample sites are indicated by number. (1) Nelson River, MB; (2) Lake Winnipeg, MB; (3) Lake Sakakawea, ND; (4) Lake Pepin, MN; (5) Green Bay, WI; (6) Lake Ontario, ON; (7) Lake Erie, NY; (8) Ohio River, OH; (9) Pickwick Reservoir, AB; (10) Missouri River, MO; (11) Lake Maloney, NB.

my study did not reveal significant values of the sum of squared differences which would suggest spatial or demographic expansion in these populations. Some form of expansion post glaciation, likely spatial expansion followed by demographic expansion, must have occurred in these populations but was not detected. Currat et al. (2006) observed that mismatch distribution can result in false results when using a comparison of haplotype frequencies, severe bottlenecks and sampling error can also skew results. One, or a combination, of these factors has likely affected my data. Tau values for Lake Ontario and Lake Erie are ambiguous. Green Bay and Lake Maloney consistently display high tau value and appear to have existed prior to the last glacial maximum. It is likely that these populations survived in refugia and tau values reflect their older age, as will be discussed in the section on historical relationships and colonization routes. Nelson River, Lake Winnipeg, Lake Pepin and Pickwick Reservoir appear to be the most recently colonized populations. These populations, with the exception of Pickwick Reservoir, are found in previously glaciated regions and would be expected to have been colonized recently. Data from my study suggests Pickwick Reservoir was colonized after the last glacial maximum along with the northern glaciated regions, agreeing with results in a phylogeographic study of bigeye chub (Berendzen et al., 2008). This inference will be further discussed in the following sections.

Linkage disequilibrium decreases with each generation and may be indicative of selection, genetic drift or blending of populations. Low levels of linkage disequilibrium are also correlated with high mutation rates (Hartl and Clark, 1997). In general, linkage disequilibrium results were inconsistent between control region and ND2 analysis. Populations displaying low levels of linkage disequilibrium in control region analysis often displayed high levels of linkage disequilibrium in ND2 region analysis, or vice versa. Populations that are older based on biogeographic data and additional analyses (such as haplotype diversity, nucleotide diversity

and tau values) should display lower levels of linkage disequilibrium. Conversely to what is expected, linkage disequilibrium levels tend to be low in populations which should be younger (colonized after the end of the last glacial maximum), and higher levels are observed in populations that by all accounts should be older (colonized prior to the Wisconsin glaciations). The only exceptions are Lake Ontario and Pickwick Reservoir which display low levels of linkage disequilibrium in both regions of mitochondrial DNA, while Lake Maloney which consistently display high levels of linkage disequilibrium. These results are the exact opposite of what is inferred by biogeographic data and other analyses in my study. Control region analysis detected low levels of linkage disequilibrium in Nelson River, Lake Winnipeg, Ohio River, and Lake Pepin. These populations, with the exception of the Ohio River are located near or within the last glacial maximum and should be young, thus should display high levels of linkage disequilbrium. Moderate linkage disequilibrium is observed in Lake Sakakawea, Green Bay and the Missouri River indicating an intermediate age, suggestive of colonization during (or prior to) the Wisconsin glaciation. Lake Maloney and Lake Erie display high levels of linkage disequilibrium, suggesting they are older populations, a reasonable result for the Lake Maloney population, but contradictory to other statistical data of Lake Erie. Linkage disequilibrium revealed by ND2 analysis was low in Lake Ontario, Lake Erie, Pickwick Reservoir and Missouri River. Green Bay, Lake Pepin and Lake Maloney display higher levels. Many linkage disequilibrium results are contradictory between the control region and ND2 genes, possibly resulting from genetic drift and population blending affecting control region analysis. These data also contradict other analyses within my study including nucleotide diversity, shared haplotypes and mismatch distribution, and is also contradictory to biogeographic data. Linkage disequilibrium is affected by multiple factors including selection, recombination and mutation rate, population structure, genetic drift and non-random mating (Hartl and Clark, 1997). It is possible that a higher

mutation rate exists in some populations making levels lower than they should be, or the effects of genetic drift are prevalent in some populations, maintaining high levels which should have decreased over time. One or a combination of factors appears to be skewing linkage disequilibrium making these data unreliable. The only populations displaying consistent levels of linkage disequilibrium are Lake Ontario, Pickwick Reservoir and Lake Maloney. In Lake Ontario and Pickwick Reservoir linkage disequilibrium is contradictory to known data and to data derived from my study. As linkage disequilibrium data is contradictory to other analyses, and also to geographic data I will not discuss the results of these data further.

Tajima's D measures neutrality of evolution. Positive values indicate a decrease in population size or balancing selection, negative values indicate population size expansion, positive selection, or a recent bottleneck (Tajima, 1993). Significant negative values of Tajima's D were found for Nelson River in the control region, and for Lake Ontario, Lake Erie, Lake Winnipeg and Lake Pepin in ND2 analysis, contradictory to linkage disequilibrium values. These populations are found above the ice edge of the last glacial maximum (with the possible exception of Lake Pepin). It is most likely these data indicate the presence of a recent genetic bottleneck. Results in non-glaciated regions were not significant, although in ND2 analysis all regions identified as older populations display positive values of Tajima's D.

Molecular variance was identified using AMOVA. Based upon each population individually, most variation (72.94% (CR) and 70.97% (ND2)) is found within populations. Based upon drainage basins, most variation (81.82% (CR) and 92.03% (ND2)) is found within populations. This reflects a high degree of variation within populations contrary to other studies, including those of rainbow darter (Ray et al., 2010) and yellow perch (Grzybowski et al., 2010) which indicate higher degrees of variation between populations. This suggests that the freshwater drum

population was widespread prior to the Pleistocene with similar haplotypes among various populations, resulting in minimal variation between populations. This is also indicative of a high level of gene flow between populations of freshwater drum, perhaps reflecting the great mobility of female freshwater drum and planktonic eggs.

F_{ST} determines the correlation between alleles in subpopulations to that of the total population. ND2 analysis indicates significant correlations between many populations (Table 17, 18). This is expected as the mutation rate is lower in the ND2 region than in the control region and is reflective of the relationships between freshwater drum populations prior to the last glacial maximum. There are great variation in control region data (Table 7, 8). Green Bay and Lake Maloney appear to be very distinct from other populations suggesting long term isolation and limited gene flow. The Missouri River, and particularly the Ohio River, are correlated with many populations, and are likely the primary source regions from which dispersal to other regions originated. The correlations between these populations will be discussed further in the sections on contemporary population structure and historical relationships.

Analyses of genetic data derived from my study reveals genetic structure within and among populations. Contrary to my prediction "there will be more variation between populations than within populations" AMOVA analysis of individual populations identified more than 70% of variation existed within populations indicating gene flow between genetic subpopulations (Table 6, 16). AMOVA analysis of drainage basins indicated 81.82% of genetic variation in the control region and 92.03% of genetic variation in the ND2 region is within watersheds (Table 6, 16). Shared haplotypes (Table 3, 14; Figure 34, 35) and F_{ST} analysis (Table 7, 8, 17, 18) indicate partial agreement with the predictions of hypothesis 1 "populations in geographic proximity will display greater genetic similarity" and "populations within the same drainage basin will display greater

genetic similarity". Lake Winnipeg and Nelson River populations are very similar, as are Lake Erie and Lake Ontario of the Great Lakes drainage. In the case of the Hudson Bay drainage, Lake Winnipeg was likely the colonization source for the Nelson River, thus great similarity would be expected between these populations. The pattern observed in these freshwater drum populations is similar to that of lake trout populations within the Hudson Bay drainage system in Manitoba (Wilson and Hebert, 1998) and will be discussed further in the section on historical relationships and postglacial dispersal of freshwater drum. It is likely that the similarity between Lake Ontario and Lake Erie populations is a result of colonization by the same populations and gene flow between them. Green Bay, a member of the Great Lakes drainage, exhibits a degree of similarity to Lake Erie, but minimal correlation with Lake Ontario. Many other phylogeographic studies observed differentiation between fish populations in the upper and lower Great Lakes (Borden and Krebs, 2009; Stepien et al., 2009). The reasons for this differentiation will be discussed in subsequent sections. Similarities exist within the Mississippi River drainage, with greater correlations observed within populations to the east and west of the Mississippi River. Borden and Krebs (2009) detected eastern and western groups in populations of smallmouth bass, as did Ray et al. (2006) in rainbow darter. There is a trend towards increased genetic variation in regions that were not glaciated, in part confirming my prediction "decreased genetic variation will be correlated with regions that were previously glaciated" (Figure 36). This prediction is not completely validated, Lake Erie and Lake Ontario are exceptions to the observed trend. Increased genetic variation in these populations likely reflects populations blending as multiple access points to these regions were available as the Wisconsin glaciation ended. Pickwick Reservoir is another anomaly, decreased genetic variation is observed in this population, yet it is found in a region which remained unglaciated. Bigeye chub presents a similar phylogeographic pattern in this region (Berendzen et al., 2008).

Colonization of these populations will be discussed in the section on historical relationships and post-glacial dispersal. Genetic structure is evident in the freshwater drum population as indicated by nucleotide and haplotype diversity (Table 2, 12), shared haplotypes (Table 3, 13), and differentiation in haplotype frequencies (Appendix 4, 5). This makes it possible to identify subpopulations within the freshwater drum population and to analyze modern day and historical relationships and demography of these populations, which will be discussed in depth in the following sections.

Contemporary population analysis

The current geographic range of freshwater drum encompasses the Mississippi, Great Lakes and Hudson Bay/Nelson River drainage basins, along with separate drainages to the Gulf of Mexico (Berra, 2001; Stewart and Watkinson, 2004). The Mississippi drainage basin is composed of the Mississippi River and several tributaries, including the Missouri, Tennessee and Ohio Rivers (Robison, 1986).

The freshwater drum population is composed of many subpopulations which contribute to the population structure as a whole. Analyses made it possible to determine some of the characteristics of each subpopulation and the relationships between them. The Ohio River supports a large variety of freshwater fishes, many of which survived glaciation in southern refugia in the lower Ohio River and its tributaries (Hocutt et al., 1986. Fishes were able to repopulate the upper Ohio River and its tributaries subsequent to glacial retreat (Hocutt et. al., 1986). The large size of the Ohio River, along with diversity of habitat, is correlated with supporting multiple and large populations of fishes (Burr and Page, 1986). My data indicates a high level of control region nucleotide diversity in the Ohio River population, six different

haplotypes were detected in a sample of eight individuals (Table 2; Figure 37). The Ohio River freshwater drum population is significantly correlated with more populations than any other sampled location in my study despite the small sample size (Table 8, 18). Previous studies of freshwater fishes, including rainbow darter (*Etheostoma caeruleum*) (Ray et al., 2006) and smallmouth bass (*Micropterus dolomieu*) (Borden and Krebs, 2009), also indicate the Ohio River freshwater fish populations are associated with many subpopulations. The Ohio River freshwater drum population shares control region haplotypes with all populations except for Nelson River, Lake Winnipeg and Lake Maloney. Control region haplotypes of the Ohio River population are dispersed among haplotype groupings relatively equally, indicative of this population existing over a long period of time or colonization by multiple populations (Figure 24, 26). Either scenario is likely, as the modern day population of the Ohio River is hypothesized to be largely derived from the ancient Teays River. The freshwater fish populations of the Teays River were fragmented and forced to relocate southward as ice advanced (Hocutt et al., 1986). It is most likely that multiple populations colonized the present day Ohio River, as evidenced by the diversity of control region haplotypes and placement in multiple haplotype groupings (Figure 24, 26, 37). This will be further discussed in the following section of historical relationships and colonization routes. Two ND2 haplotypes were detected in the Ohio River population, including h1, the most common haplotype within the total, continental, freshwater drum population. Both ND2 haplotypes detected in the Ohio River population are confined to a single group (Figure 29, 31, 37). The increased mutation rate of the control region is reflected in these data, differentiation is observed in the rapidly mutating control region, the sample size is likely too small to resolve genetic differentiation in ND2 analysis. The genetic composition of the Ohio River population revealed in my study is indicative of a diverse population that has

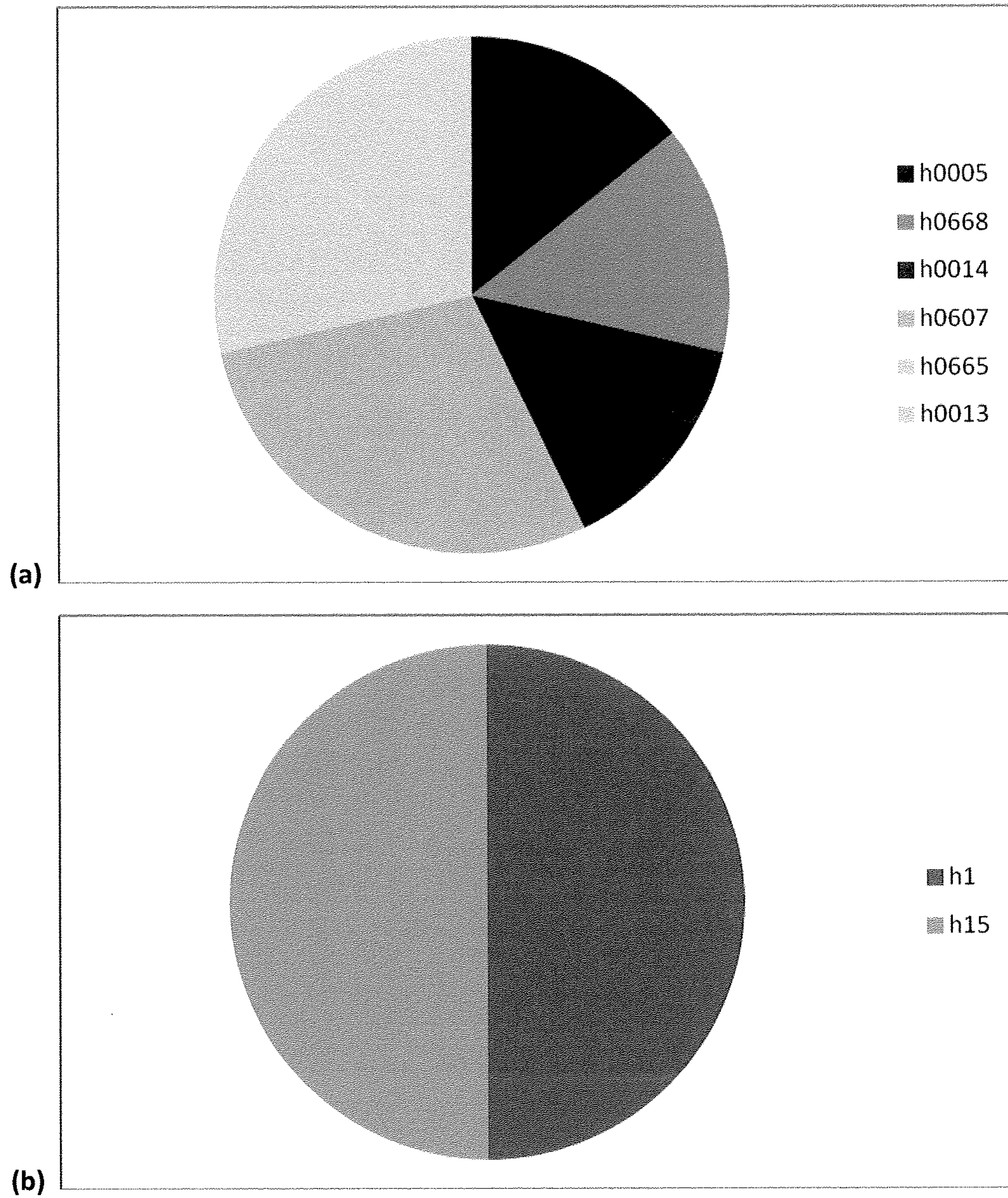


Figure 37. Haplotypes in the Ohio River freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b).

persisted over time, consistent with the genetic composition of Ohio River populations observed in studies of gilt darter (Near et al., 2001) and rainbow darter (Ray et al., 2006).

The Missouri River population sampled is located near the confluence of the Mississippi and Missouri Rivers (Figure 15). This region was not glaciated and historically supported a diversity of freshwater fishes (Cross et al., 1986; Robison, 1986), thus a high degree of genetic variation should be observed in this population. Control region analysis indicates a moderate level of nucleotide diversity within the Missouri River population in comparison to other sampled populations. A high level of haplotype diversity comparative to other sampled freshwater drum populations was observed, with eleven different haplotypes detected (Table 2; Figure 38). Analysis of the more slowly evolving ND2 region also indicates a high level of diversity, seven different haplotypes were detected, much more than were found in any other sample (Table 12; Figure 38). These data suggest the Missouri River population existed prior to the Wisconsin glaciation, possibly dating to the Pliocene. The number of ND2 haplotypes in particular reflects long term evolution within this population. This agrees with data presented by Berendzen et al. (2008), who determined the bigeye chub (*Hybopsis amblops*) populations of the Mississippi and Missouri River regions display a high degree of genetic variation, as have many other studies (Near et al., 2001; Stepien et al., 2009). At least one control region haplotype is shared with each population, with the exception of Lake Sakakawea (which does not share a control region haplotype with the Missouri River) (Table 3; Figure 35). However, three ND2 haplotypes are shared between Lake Sakakawea and the Missouri River (Figure 34), indicating they were historically related but have since diverged. The Missouri River contains eleven different control region haplotypes and seven different control region haplotypes were identified in the Lake Sakakawea population, making it unlikely that contemporary gene flow has been maintained.

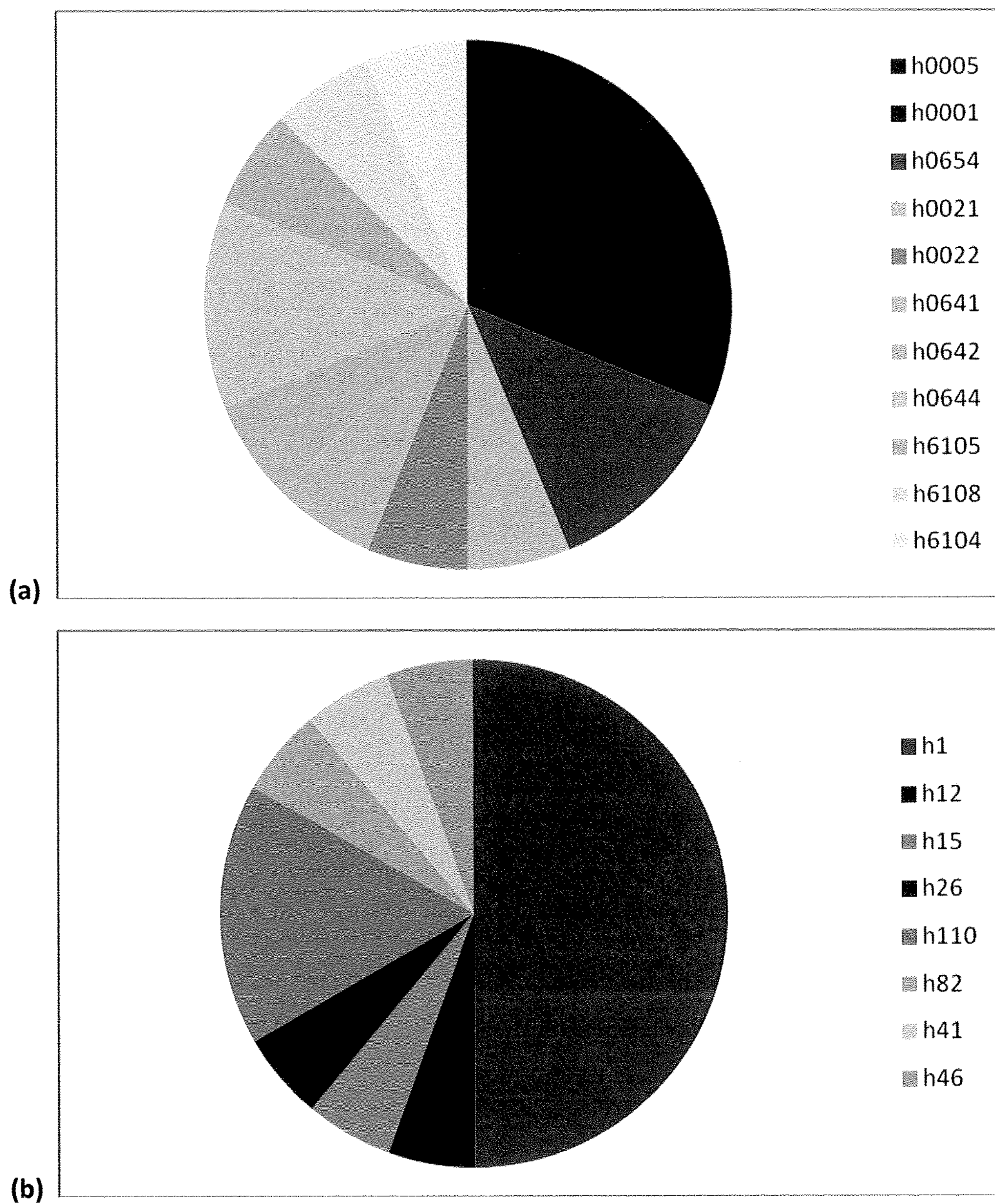


Figure 38. Haplotypes in the Missouri River freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b).

ND2 data also indicate a historical relationship between the Missouri River sample and Lake Erie, Green Bay and Lake Maloney (Table 13). My data indicates a freshwater drum population existed prior to the last glacial maximum and likely prior to the Pleistocene, in the Missouri River. My analyses indicate the Missouri/Mississippi River freshwater drum population has attained great genetic variation over the course of its existence, in accordance with other phylogeographic studies (Berendzen et al., 2003; Borden and Krebs, 2009; Stepien et al., 2009). The large size of the Mississippi and Missouri Rivers makes this area a center of evolution and diversification for fish species (Robison, 1986; Hewitt, 1996, 2000), accounting for the high level of genetic diversity observed in multiple phylogeographic studies. Gene flow between the Missouri and Mississippi River region and other freshwater habitats has had a great effect on the distribution of freshwater fishes (Robison, 1986). The Missouri River freshwater drum population appears to have existing, or historical, relationships with all other sampled populations. Historical relationships will be further discussed in the following section.

Lake Maloney, located in the Platte River in Nebraska as part of the Missouri River drainage, is a unique population. The most common control region haplotypes of Lake Maloney are not found in any other population, and total 87.5% of the Lake Maloney population (Figure 39, Appendix 3). The unique genetic composition observed within the Lake Maloney population suggests this population was colonized prior to the end of the last glacial maximum. F_{ST} analysis of the control and ND2 regions also indicate Lake Maloney is significantly differentiated from the majority of populations (Table 7, 8, 17, 18), parsimony and Bayesian analyses further verify these results (Figures 24, 26, 29, 31). This population also displays a high level of linkage disequilibrium and nucleotide diversity, particularly in regard to ND2 data, where four different haplotypes were observed, more than any population except the Missouri River (Table 2, 12).

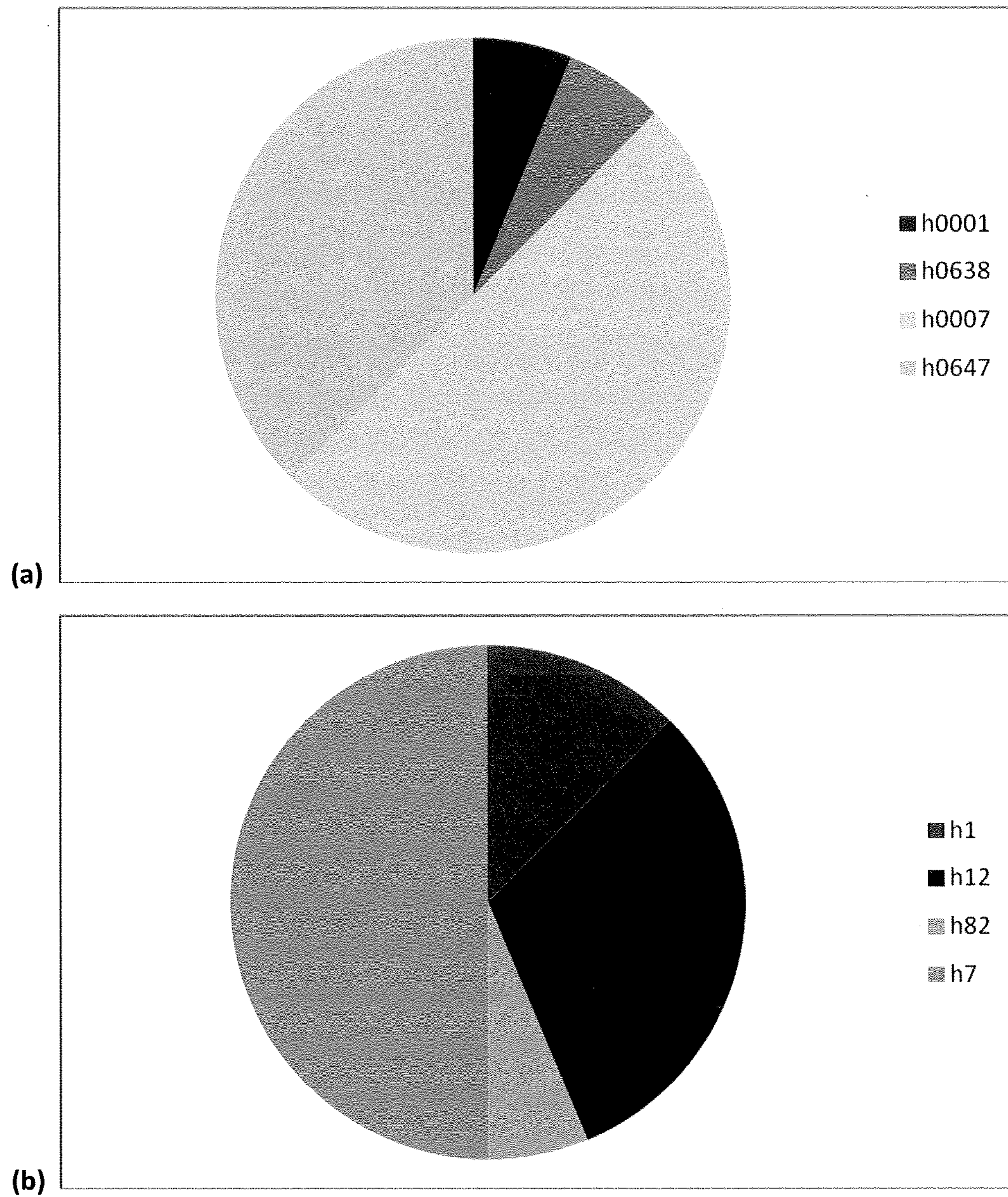


Figure 39. Haplotypes in the Lake Maloney freshwater drum population.

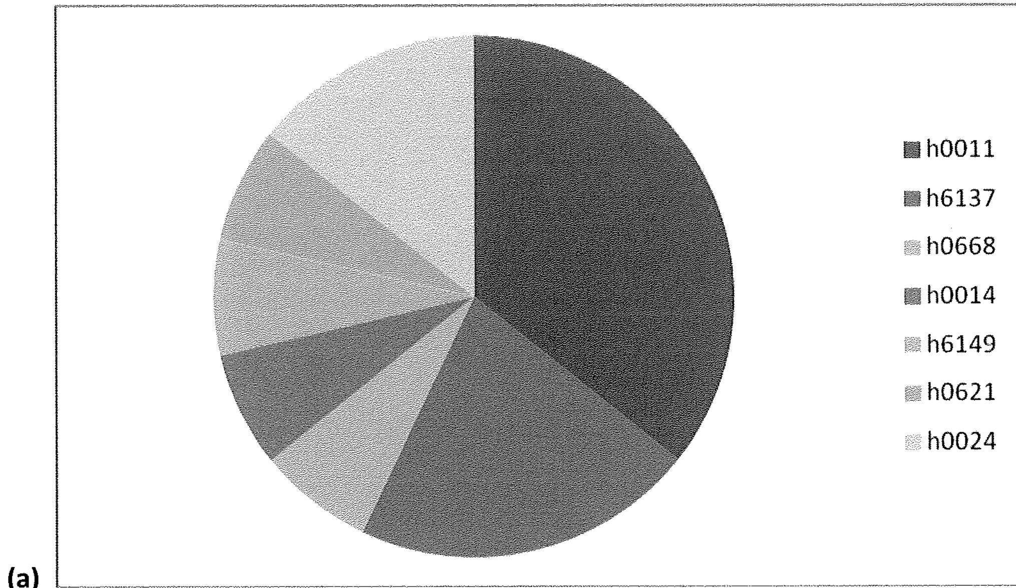
Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b).

These data indicate isolation over a length of time sufficient to result in differentiation. When coupled with the high levels of linkage disequilibrium this is suggestive of genetic drift. Limited haplotype diversity (Figure 39, Appendix 3, 4) also suggests genetic drift is a factor in the genetic composition of this population. Genetic drift within the Lake Maloney population will be further discussed in the section on historical relationships. Mismatch distribution analysis indicates demographic expansion (Table 4), consistent with the Lake Maloney population having undergone a rapid and large expansion in terms of numbers of individuals, resulting in an excess of haplotypes that occur at a low frequency (Appendix 3). Glacial retreat of the ice sheet along the border of present day state of Nebraska likely resulted in improved habitat quality for the Lake Maloney population, possibly by introducing additional food sources, allowing the population to greatly increase in numbers.

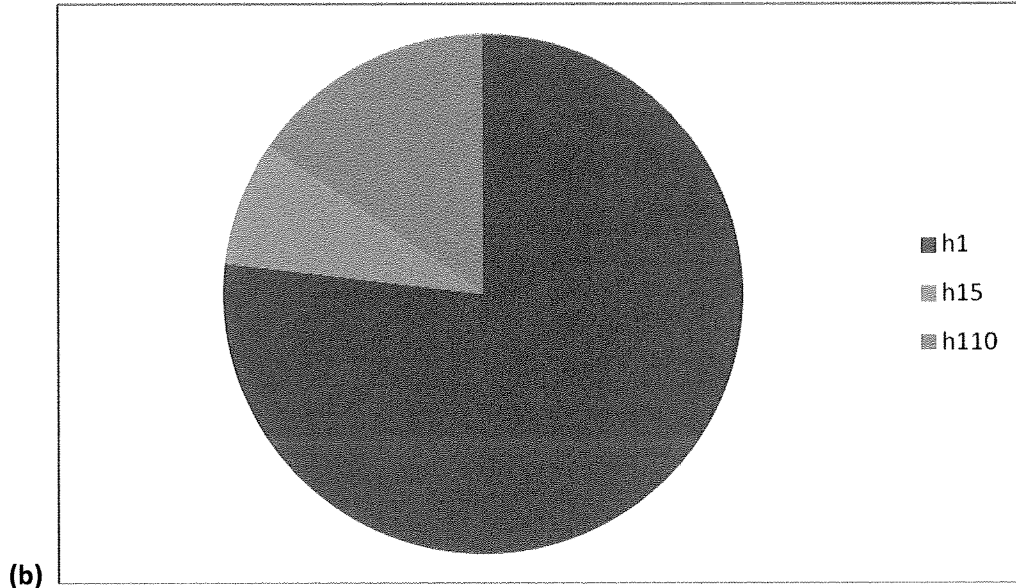
This population of Lake Maloney is highly differentiated from other populations as evidenced by its unique genetic composition, suggesting an absence of gene flow with other populations. There are limited phylogeographic studies with which to compare the genetic composition of the Lake Maloney population. However, a phylogeographic study of killifish (*Fundulus zebrinus*) revealed three clades. Genetic separation was evident in the haplotypes of the Platte River, and the Republican and Smoky Hill drainages of Nebraska, Kansas and Colorado. Together these three drainages formed a "northern" clade, distinct from the other populations sampled (Kreiser et al., 2001), and would contain the Lake Maloney population. This supports my data suggesting that the Lake Maloney population represents a population that has been genetically differentiated from other freshwater drum populations.

Lake Sakakawea exhibits a relatively high level of nucleotide diversity, particularly based on analysis of the control region (Table 2). Data revealed three unique control region haplotypes

(Figure 40; Appendix 4), analysis of control region haplotypes and the populations they are shared with reveals a pattern of shared haplotypes which originate from the area of the Missouri/Mississippi River (Figure 35(b), (e), (f)). Control region haplotypes are not shared with the Great Lakes populations of Lake Erie and Lake Ontario, and only one haplotype is shared with Green Bay (Table 3). This is similar to the pattern observed in yellow perch (*Perca flavescens*) by Grzybowski et al. (2009) which indicated isolation between Midwestern populations (which would also include my sampled populations of Lake Sakakawea, Nelson River, Lake Winnipeg and Lake Pepin) and the Great Lakes populations. Collectively these data suggest separate colonization sources of these populations, which will be discussed in the following section. Control region F_{ST} analysis indicates Lake Sakakawea is significantly differentiated from all populations with the exception of the Ohio River. This genetic relationship between the Ohio River and Lake Sakakawea populations indicates gene flow occurring between these two populations which must have occurred along the Mississippi River (Table 8). Analysis of the ND2 region indicates no significant differentiation of the Lake Sakakawea population from the other sampled populations with the exceptions of Green Bay and Lake Maloney (Table 18). This difference between ND2 and control region analysis suggests Lake Sakakawea has diverged from other populations, and may have been colonized prior to the end of the Wisconsin glaciation. These data suggest the existence of a refugium in this region, which will be discussed in the following section. Mismatch distribution data suggest the population underwent a demographic expansion as ice retreated and conditions improved (Table 4). My data suggest that Lake Sakakawea has existed for a sufficient period of time to attain increased diversity, particularly in the rapidly evolving control region. However, the population has not existed for a sufficient amount of time in which to observe significant



(a)



(b)

Figure 40. Haplotypes in the Lake Sakakawea freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b).

differentiation in the more slowly evolving ND2 region. A complete discussion of historical inferences of the Lake Sakakawea population will follow in the historical relationships section.

Nelson River and Lake Winnipeg are the only populations sampled within the Hudson Bay drainage. As these two populations exhibit a high degree of similarity they will be discussed together. The populations of the Nelson River and Lake Winnipeg are composed of a small number of haplotypes, reflecting the relatively young age of these populations (Figure 41, 42). Control region data indicate Nelson River is significantly correlated only with Lake Winnipeg. Lake Winnipeg also shows a genetic relationship with Lake Pepin and Pickwick Reservoir, shared haplotypes indicate a route roughly correlated with the Mississippi River, extending into the Tennessee River (Figure 35 (a), (b), (d); Table 7, 8). Parsimony and Bayesian analysis place the majority of control region haplotypes within a single haplotype group, with a small percentage contributing to two and three additional haplotype groups respectively (Table 9, 10; Figure 24, 26). This further indicates limited genetic diversity. ND2 data suggest both Nelson River and Lake Winnipeg freshwater drum populations are distinct from Green Bay, Lake Maloney and the Missouri River (Table 17, 18). Analysis places all ND2 haplotypes within a single haplotype group, again reflective of the youth of Nelson River and Lake Winnipeg (Table 19, 20; Figure 29, 31), while also indicating colonization from a single source. Limited nucleotide diversity suggests Nelson River and Lake Winnipeg populations were recently colonized and have not existed for a sufficient amount of time in which to acquire genetic differentiation (Table 2, 12). My results are similar to a study by Stepien et al. (2009) of walleye (*Sander vitreus*) populations where isolation in the Manitoba and upper Mississippi regions are evident. My data also correlate with geological history, as Lake Winnipeg and Nelson River were covered by ice until at least 12800 years ago, when Lake Agassiz began to form (Bailey and Smith, 1981), thereby could

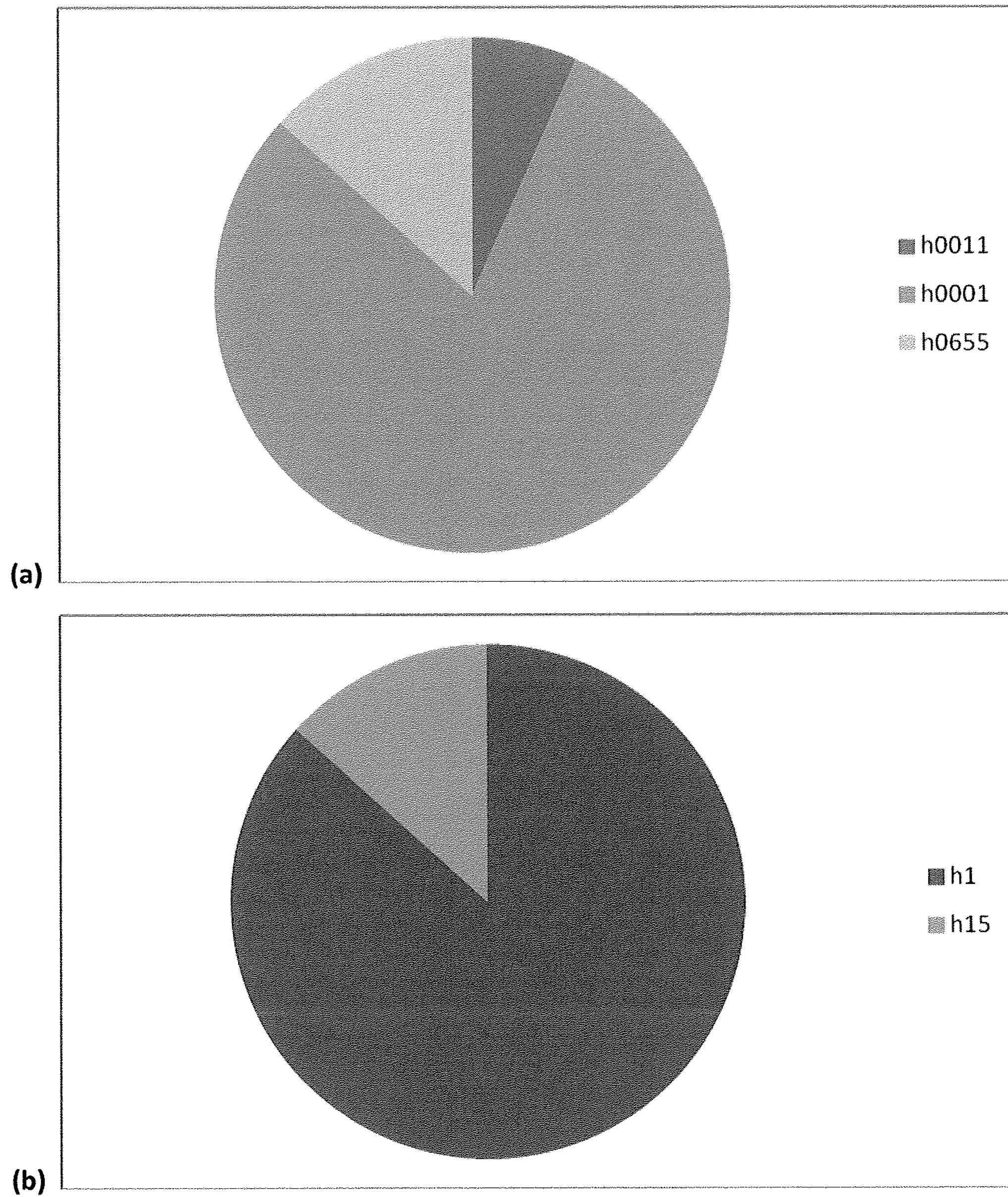


Figure 41. Haplotypes in the Nelson River freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b).

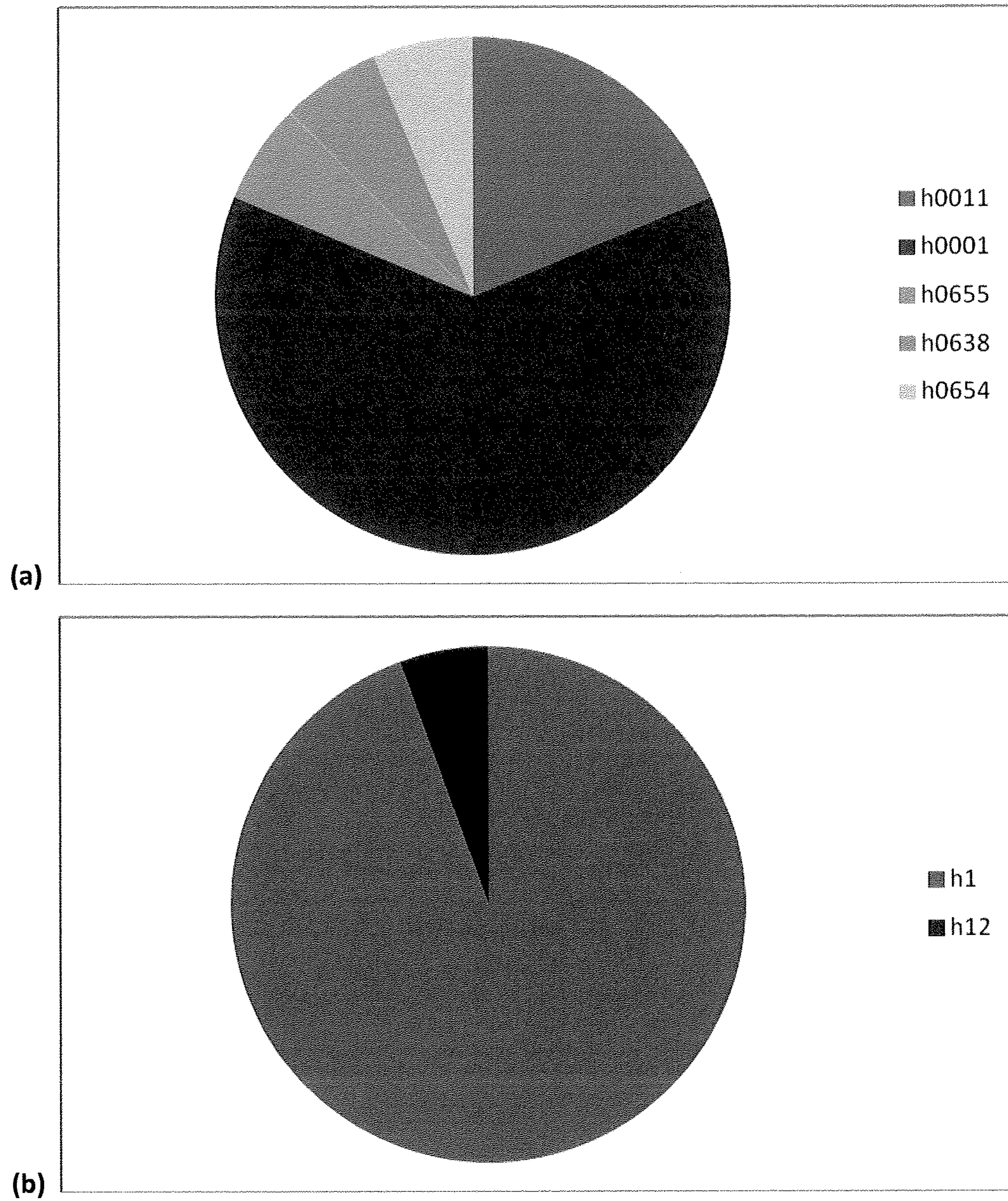


Figure 42. Haplotypes in the Lake Winnipeg freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b).

not have been colonized until that time. Significant values of Tajima's D are consistent with a population bottleneck (Table 5, 15), which correlates with the low levels of diversity observed in these populations. Data suggest these populations were colonized by freshwater drum with limited genetic diversity (founder effect) and have not existed for a great length of time.

Lake Pepin displays a moderate level of nucleotide diversity in both control region and ND2 region analysis in comparison to other freshwater drum populations (Table 2, 12). Nine haplotypes were revealed in the control region more than the average in sampled populations (Figure 43). The haplotypes in this population are distributed amongst a number of groups, four of six based on control region parsimony analysis (Table 9; Figure 24), and five of six based on control region Bayesian analysis (Table 10; Figure 26). This likely indicates the Lake Pepin freshwater drum population was colonized by more than one population. Control region haplotypes are shared with every population, including five with the Pickwick Reservoir, more than any other population, consistent with common ancestry. The pattern of shared control region haplotypes among populations is centered around the central region of the Missouri/Mississippi River (Figure 35). Control region analysis indicates Lake Pepin is significantly differentiated from most populations with the exceptions of Lake Ontario, Lake Winnipeg and Pickwick Reservoir (Table 7, 8). This also suggests colonization by multiple sources as will be discussed in the following section on historical relationships and colonization routes. The haplotypes of the Lake Pepin population are found within both ND2 haplotype groups in parsimony and Bayesian analysis (Figure 29, 31). Based on ND2 analysis, Lake Pepin significantly differs from only the populations of Green Bay and Lake Maloney (Table 16, 18). Data from the ND2 region is reflective of population genetic structure prior to the last glacial maximum, while control region data indicate this population has recently diverged from many other populations. My data differ from studies of smallmouth bass (Borden and Krebs, 2009)

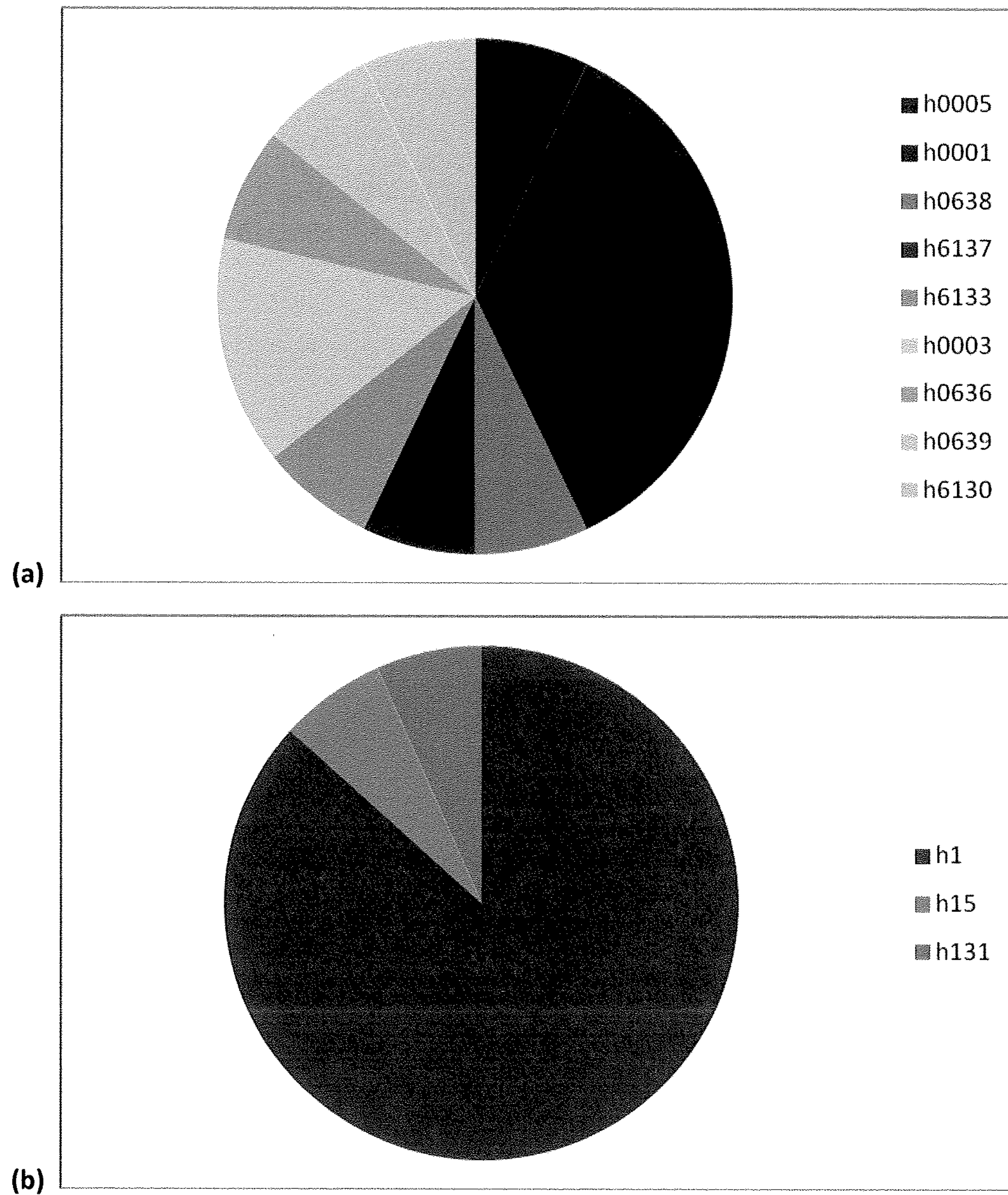


Figure 43. Haplotypes in the Lake Pepin freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b).

and walleye (Stepien et al., 2009) populations which indicate isolation persisting into the post-glacial period in the upper Mississippi regions, whereas significant gene flow appears to have occurred in the freshwater drum population of Lake Pepin.

ND2 and control region nucleotide diversity in the Green Bay population is average, while haplotype diversity is above average in comparison to other freshwater drum populations. Eight different control region haplotypes were detected (Table 2; Figure 44). Five of these control region haplotypes are shared with other populations, no control region haplotypes are shared with the Lake Maloney population (Table 3). Analysis of the control region revealed three unique haplotypes, all at a frequency of 6.25%. The most common control region haplotype (frequency 0.562) is also found only in the Lake Erie population (frequency 0.1) (Figure 35, Appendix 3). The unique genetic structure, as evidenced by rare haplotypes (h0012, frequency 0.562; h0627, h0696, h6100, frequency 0.625 respectively) existing in high frequencies, is indicative of limited gene flow to and from other populations, resulting in diversification of the Green Bay freshwater drum population. Haplotypes h0012, h0627, h0696, h6100 comprise 74.95% of the Green Bay population, only haplotype h0012 is shared with another population. The genetic differentiation detected in this population has been observed in other Lake Michigan fishes, including yellow perch (Grzybowski et al., 2010). Parsimony and Bayesian control region analysis place the majority of this population (0.7495) within a single group, shared only by the populations of Lake Ontario, Lake Erie, and Pickwick Reservoir. These data indicate colonization of Green Bay was largely derived from a single source, which will be further discussed in the following section. Limited gene flow along the Mississippi River and through the Great Lakes exists as evidenced by minimal sharing of haplotypes, and will be discussed in the subsequent section of this paper. F_{ST} control region analysis also indicates Green Bay is

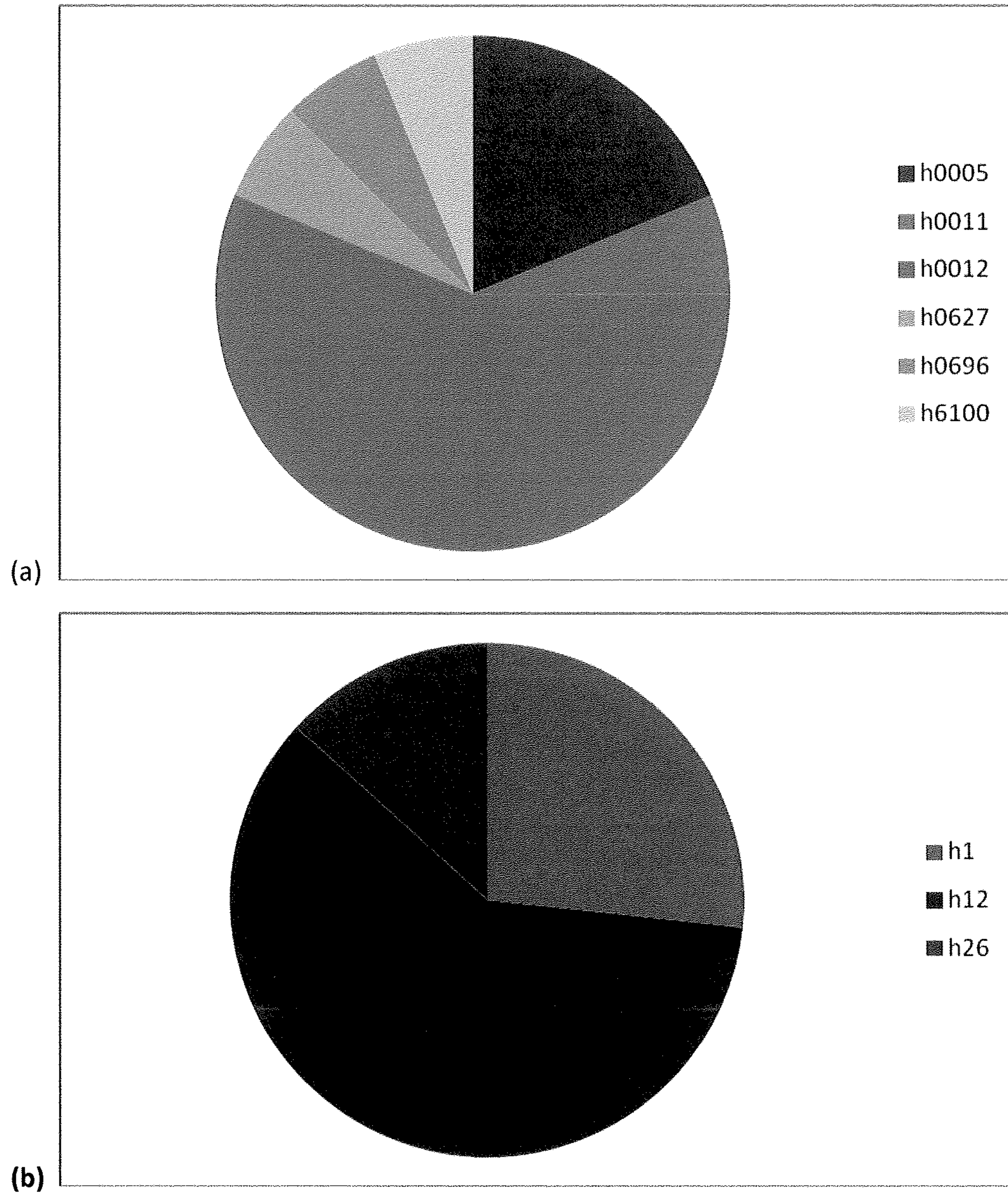


Figure 44. Haplotypes in the Green Bay freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b).

significantly differentiated from all populations with the exception of the Ohio River (Table 8). Gene flow between the Green Bay and Ohio River populations must occur along the Mississippi River. ND2 haplotypes are shared with every other population, including three with the Missouri River (Table 13). A unique ND2 haplotype was not found, however the predominant haplotype of Green Bay (frequency 0.6) is also only shared with the population of Lake Maloney (frequency 0.312) (Appendix 4). ND2 analysis reveals a high level of nucleotide diversity (2.343) in comparison to other populations, indicative of long term evolution (Table 12). Parsimony and Bayesian analysis place the majority of this population within the second ND2 haplotype grouping (Table 19, 20), unlike any other sampled population. ND2 F_{ST} values indicate significant differentiation from all other populations, indicating an absence of gene flow over time (Table 18). Tau values (Table 4, 14) suggest the freshwater drum population of Green Bay existed prior to the Wisconsin glaciation, values are greater than observed in other sampled populations near or to the north of the last glacial maximum, in accordance with other data observed in this population, including the increased haplotype and nucleotide diversity. Gene flow between Green Bay and other populations has been limited, resulting in the unique genetic composition of this population as indicated by the relatively low amount (and frequency) of shared haplotypes in comparison to other freshwater drum populations. Genetic differentiation has been observed in the upper Mississippi and upper Great Lakes in other phylogeographic studies. In some fishes, including smallmouth bass (Borden and Krebs, 2009), populations in these geographic areas contribute to a western clade. Other studies, including that of rainbow darter (Ray et al., 2006) reveal a separate clade containing populations from the upper Mississippi, Missouri River tributaries, Hudson Bay tributaries and western Lake Michigan. The historical factors resulting in this population structure will be discussed in the subsequent section. The freshwater drum population of Lake Erie exhibits a lower level of nucleotide and

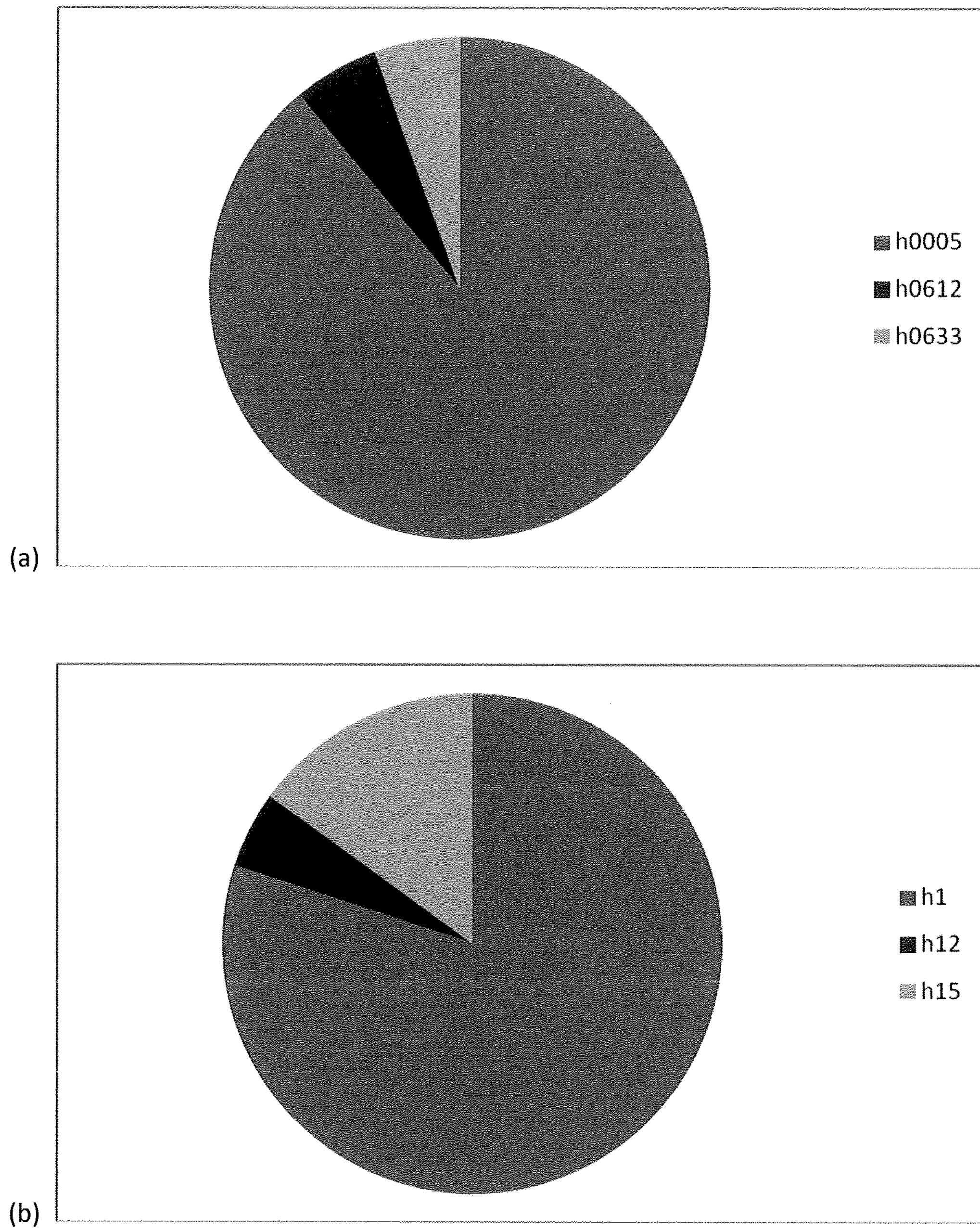


Figure 45. Haplotypes in the Lake Erie freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b).

haplotype diversity than most other populations sampled (Table 2, 12; Figure 45). Similar results were observed in walleye populations (Stepien et al., 2009), revealing a pattern of genetic differentiation in freshwater fishes of Lake Erie. This population does not share control region haplotypes with any population to the west of the Mississippi River or with the Pickwick Reservoir (Table 3). This pattern is also observed in the bigeye chub (Berendzen et al., 2008). However, ND2 analysis shows the population of Lake Erie shares three haplotypes with the Missouri River and two haplotypes with every other population (Table 13, Appendix 4). Control region F_{ST} values indicate Lake Erie is significantly differentiated from every population with the exception of Lake Ontario (Table 8). ND2 F_{ST} values suggest Lake Erie is associated with all populations with the exception of Green Bay, Lake Maloney and the Missouri River (Table 18). ND2 mismatch distribution values approach significance for spatial expansion (Table 14), this inference is corroborated with a statistically significant value of Tajima's D (Table 15). Parsimony analysis places 80% of the Lake Erie population within a single widespread haplotype group, with the remaining haplotypes distributed amongst two separate haplotype groups, one of which is widespread, and the other composed of the three Great Lakes populations and the Ohio River (Table 9). This suggests the Ohio River as being a source population to the upper and lower Great Lakes as will be discussed in the following section. Bayesian control region analysis also places 80% of the Lake Erie population in a single haplotype group, the remaining haplotypes are spread across two groups, all of which are widespread (Table 10). ND2 parsimony and Bayesian analysis place 94.4% of the haplotypes in Lake Erie within a single group containing haplotypes from every sampled population (Table 19, 20). The population of Lake Erie could not have existed until 14000 years ago when Lake Maumee began to form (Bailey and Smith, 1981). Genetic differentiation is evident, and is also observed in the population structure of walleye (Stepien et al., 2009). This has been hypothesized to result from relatively low levels

of gene flow into Lake Erie from other sources, including other Great Lakes. Gene flow is somewhat restricted as Lake Erie is geographically isolated in comparison to other sampled Great Lakes populations. Stepien et al. (2009) hypothesized the observed genetic differentiation in the Lake Erie walleye population resulted from the short and narrow size of the Detroit River restricting access to the west, while Niagara Falls prevented entry from Lake Ontario until the inception of the Welland canal in 1830 (Rapport and Whitford, 1999; Stepien et al., 2009). This resulted in limited gene flow over a period of at least 12000 years from Lake Ontario into Lake Erie, a length of time sufficient in which to observe genetic differentiation, particularly in the control region (Figure 12) (Mandrak and Crossman, 1992).

Haplotype diversity in Lake Ontario is less than that of other freshwater drum populations (Figure 46). Similar to Lake Erie, Lake Ontario does not share control region haplotypes with populations to the west of the Mississippi River, with the exception of the Missouri River population (Table 3). This suggests different colonization sources for freshwater drum populations to the east and west of the Mississippi River, similar to that observed in bigeye chub and smallmouth bass (Berendzen et al., 2008; Borden and Krebs, 2009). These colonization patterns will be discussed in the following section. Control region F_{ST} analysis indicates the Lake Ontario population is significantly different from all populations with the exception of Lake Erie, Lake Pepin, Ohio River and Pickwick Reservoir (Table 8). It is interesting that although Pickwick Reservoir does not share haplotypes with Lake Ontario, F_{ST} analysis indicates these two populations are none the less, very similar (Table 8). Although no identical control region haplotypes are found in Pickwick Reservoir and Lake Ontario, common haplotypes differ by one to two nucleotides, as evidenced in Figure 27. The source populations of Pickwick Reservoir and Lake Ontario are very similar, likely originating from a common historical source, possibly the Teays River. This will be discussed in the following section. As observed in all other sampled

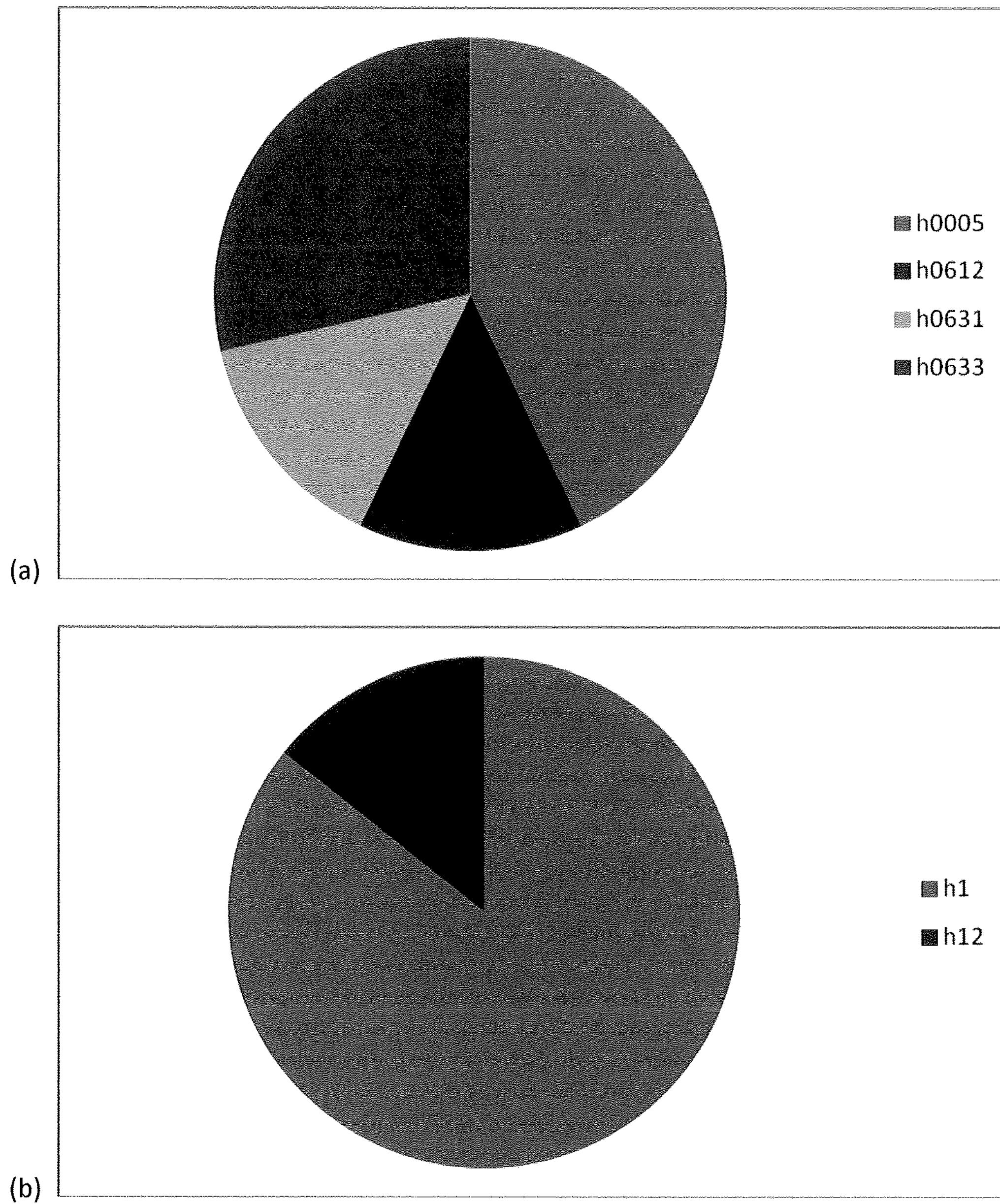


Figure 46. Haplotypes in the Lake Ontario freshwater drum population.

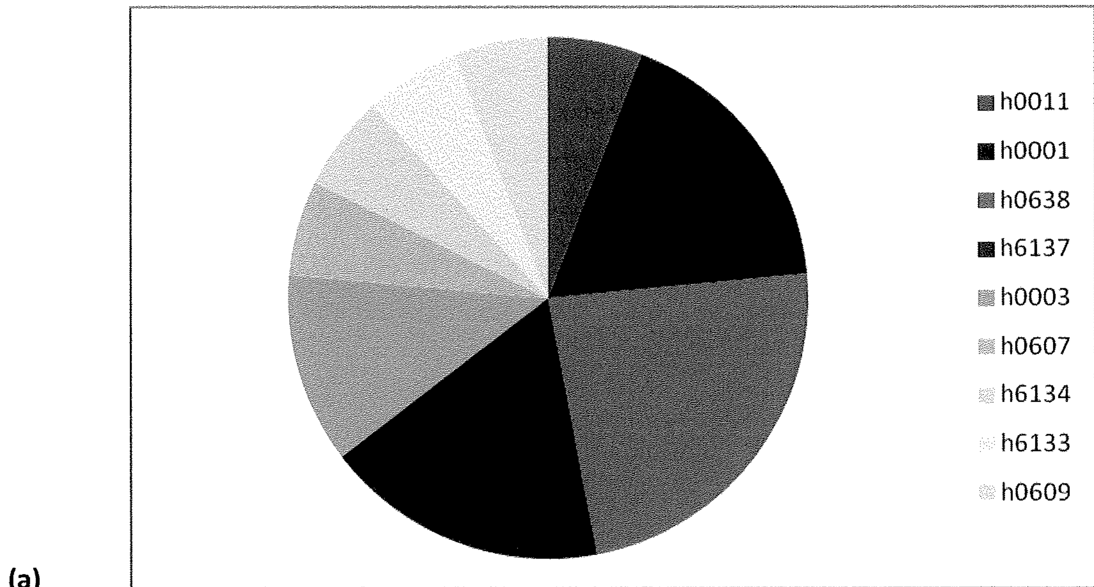
Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b).

locations, the Lake Ontario population shares at least one ND2 haplotype with every other population, consistent with an interpretation of common ancestry (Table 12; Figure 34). ND2 mismatch distribution analysis indicates the population underwent a demographic expansion (Table 14), increasing in population numbers as habitat improved. Tau values indicate this population was colonized recently, biogeographic data tell us the population of Lake Ontario could not have existed prior to 12700 years ago (Table 14) (Bailey and Smith, 1981). The Tajima's D value obtained in ND2 analysis is also significant, further indicating population expansion (Table 15). Bayesian and parsimony analysis of the ND2 region place the majority of haplotypes within a single group (Table 19, 20; Figure 29, 31). Lake Iroquois did not begin to form in the Lake Ontario basin until approximately 12700 years ago, therefore the freshwater drum population of Lake Ontario could not have existed before this time (Bailey and Smith, 1981). Data from my study suggest Lake Ontario is a young population which has begun to diversify, as observed in analysis of the control region.

The three populations sampled within the Great Lakes drainage system (Green Bay, Lake Erie and Lake Ontario) display some commonalities. The genetic composition of Lake Erie and Lake Ontario populations are similar, while Green Bay differs. The most common haplotypes in Lake Erie and Lake Ontario are shared, haplotype h0005 in the control region is found at the frequency of 0.8 and 0.429 respectively, the ND2 haplotype h1 is found at a frequency of 0.8 in Lake Erie and 0.857 in Lake Ontario (Appendix 3, 4). Lake Erie is the more distinct of the two populations, exhibiting significant differentiation, particularly based on ND2 analysis (Table 5). The Lake Erie and Lake Ontario populations are strongly correlated with one another, as expected due to their geographic proximity. Lake Ontario has not diverged to the extent of Lake Erie, likely a result of increased gene flow within the Lake Ontario population. The relative geographic isolation of Lake Erie resulting from the small size of the Detroit River to the west

(although there may be significant gene flow between Lake Erie and Lake Huron which was not sampled in my study) and Niagara Falls to the east may have contributed to the limited genetic diversity observed in the Lake Erie freshwater drum population. The small volume of Lake Erie also increases the chances of inbreeding (Klug and Cummings, 2005). Green Bay is genetically distinct from Lake Erie and Lake Ontario. This is evident in the distribution of haplotypes within groups (Table 9, 10, 19, 20; Figure 24, 26, 29, 31) and the genetic distance between common haplotypes of these populations (Figure 27, 32). Genetic differentiation between the upper Great Lakes (Lake Superior, Lake Michigan and Lake Huron, with the exception of Georgian Bay) and lower Great Lakes (Lake Ontario, Lake Erie and Georgian Bay) has been observed in studies of smallmouth bass and yellow perch (Borden and Krebs, 2009; Grzybowski et al., 2010). The degree of differentiation in smallmouth bass populations of these regions is similar to that shown by freshwater drum in my study. Like my study, analyses included the mitochondrial control region (Borden and Krebs, 2009). The cytochrome b gene was also analyzed in the smallmouth bass study, which has a lower mutation rate than that of the ND2 gene analyzed in my study. Yellow perch phylogeographic structure was studied using microsatellites, and revealed finer scale relationships than those observed in my study. The yellow perch populations exhibited distinct differences in genetic structure between the upper and lower Great Lakes (Grzybowski et al., 2010). Green Bay was likely primarily colonized by freshwater drum from a separate drainage system to that of Lake Erie and Lake Ontario (as will be discussed in the following section), gene flow between these populations is likely limited by the great distance between these lakes.

(a)



(b)

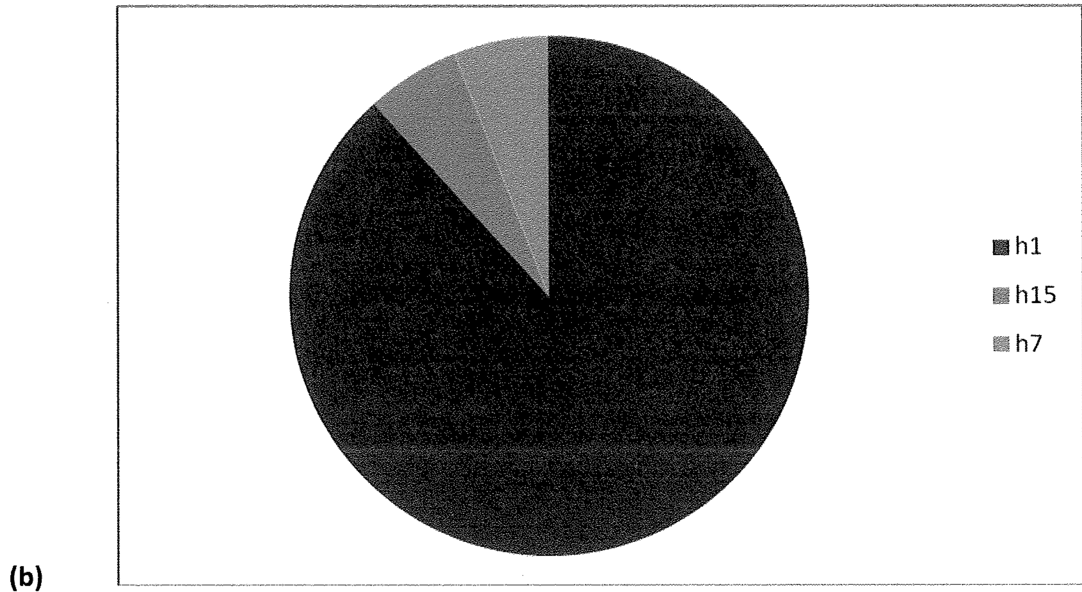


Figure 47. Haplotypes in the Pickwick Reservoir freshwater drum population.

Control region haplotypes are illustrated in (a), and ND2 region haplotypes in (b).

Pickwick Reservoir displays a relatively low level of nucleotide diversity but a high level of haplotype diversity, particularly in the control region (Figure 47). F_{ST} analysis of the control region indicates Pickwick Reservoir is correlated with Lake Ontario, Lake Winnipeg and Lake Pepin (Table 8). A large number of haplotypes are detected within the Pickwick Reservoir, suggesting colonization by multiple sources. It is possible that Pickwick Reservoir shares a central colonization source with Lake Ontario, Lake Winnipeg and Lake Pepin. These results are reflective of the colonization of these populations which will be discussed in the following section. ND2 data also indicate significant relationships between Pickwick Reservoir and Lake Erie, Nelson River and Lake Sakakawea (Table 18). Tau values suggest the population of Pickwick Reservoir is similar in age to the populations in glaciated areas (Table 4, 14). Interestingly, this indicates a correlation between Pickwick Reservoir, the most southerly sampled region, and the majority of populations (with the exception of Green Bay) that exist in previously glaciated areas. Data from my study, and that of Berendzen et al. (2008), suggest that regions of the Cumberland and lower Tennessee (the same geographic region as Pickwick Reservoir) were colonized post-glaciation. Data suggest this population has not been in existence for a period of time that would be expected based on geographic location.

My study revealed a high degree of genetic variation, particularly in the control region. AMOVA analysis indicates greater differentiation within, rather than between, populations (Table 6, 16). This contradicts my prediction "there will be more variation between populations than within populations", and also contradicts results from other studies including that of rainbow darter (Ray et al. 2006). The low level of differentiation between populations may be the result of the high mobility of female freshwater drum, and dispersal of planktonic eggs, maintaining a higher than average level of gene flow. Data produced in this study, including fixation indices,

nucleotide diversity and shared haplotypes, illustrate contemporary relationships between populations.

Evidence suggests some populations, including Lake Erie and Lake Ontario, are closely related. , The freshwater drum populations of Lake Winnipeg and Nelson River are also very similar genetically. Other populations, such as Green Bay and Lake Maloney, show great divergence from other populations. Data indicate limited gene flow in both Green Bay and Lake Maloney, likely reflective of isolation (which will be discussed in the following section "historical relationships of freshwater drum"). It is evident that the populations sampled in the Hudson Bay drainage basin (Lake Winnipeg and Nelson River) are closely genetically related to one another, as expected by their geographic proximity and shared drainage. These populations are also related to Lake Pepin and Lake Sakakawea, which are both close in proximity. Lake Winnipeg and Nelson River populations are not closely genetically related to the lower Great Lakes, Lake Maloney and Missouri River populations. This is not consistent with my prediction "populations within the same drainage basin will display greater genetic similarity", but is consistent with the results of Borden and Krebs (2009) smallmouth bass phylogeographic study. Borden and Krebs (2009) suggested that connectivity among basins is less important in explaining shared alleles than dispersal events are, which will be discussed in detail in the subsequent section of this paper. My prediction "populations in geographic proximity will display greater genetic similarity" can also not be substantiated, data suggest relationships are far more complex.

Analyses of the Great Lakes samples indicate significant correlations between Lake Erie and Lake Ontario, based upon F_{ST} analyses and genotypic composition, including shared haplotypes. Lake Erie contains three ND2 and three control region haplotypes, Lake Ontario contains two ND2

haplotypes and four control region haplotypes. These populations exhibit genetic similarity, while the genetic composition of the Green Bay population in Lake Michigan is distinct. Green Bay contains three ND2 haplotypes and six control region haplotypes. Seven control region haplotypes in total are unique to the Great Lakes system, three of these haplotypes are unique to the Green Bay population.

Genetic diversity is also observed throughout the Mississippi drainage basin. Lake Maloney is unlike any other population, 87.5% of its population was found to have unique haplotypes. Pickwick Reservoir bears more similarity to northern populations than those in closer geographic proximity as evidenced by shared haplotypes and F_{ST} analyses indicating gene flow between Pickwick Reservoir and regions occupying previously glaciated regions, or more likely, colonization from a common central region into the Pickwick Reservoir and previously glaciated populations. Collectively, my data illustrate a complex pattern of relationships among and between freshwater drum populations. As some predictions associated with hypothesis 1 were not met my hypothesis should be refined to state "genetic similarity will exist between freshwater drum populations but is not always reflective of geographic proximity or shared drainage basins". The prediction "there will be more variation between populations than within populations" was rejected, however approximately 30% of variation was found between populations. This indicates there is significant gene flow in the freshwater drum population as a whole, likely indicative of the high mobility of female freshwater drum and the dispersal of planktonic eggs. It was also possible to detect differences between populations based on differences within the populations. The predictions "populations in geographic proximity will display greater genetic similarity", "populations within the same drainage basin will display greater genetic similarity", and "increased genetic variation will be correlated with regions that

were not glaciated" were met in part. Although not always related to geographic proximity, drainage basins and/or glaciation genetic variation exists.

Most freshwater drum populations inhabiting previously glaciated areas display lower genetic diversity than those from unglaciated areas (Table 2, 12), in accordance with my prediction "decreased genetic variation will be correlated with regions that were previously glaciated". Nelson River, Lake Winnipeg and Lake Erie show evidence of genetic bottlenecks (Table 5, 15). Lake Sakakawea and Green Bay do not show evidence of a bottleneck, my data suggest these populations were colonized prior to the Wisconsin glaciation, as will be discussed in the following section. Lake Pepin and Lake Ontario also do not show evidence of genetic bottlenecks, and were likely colonized by multiple sources. Bottlenecks and founder effects are common for freshwater fishes occupying previously glaciated areas as discussed by Bernatchez and Wilson (1998). Nelson River, Lake Winnipeg and Lake Erie populations significantly diverge from the majority of the other populations, as expected as these populations occupy previously glaciated regions (Table 8, 18). Populations in regions north of the last glacial maximum typically exhibit decreased genetic diversity (Bernatchez and Wilson, 1998).

Lake Sakakawea and Lake Pepin, although occupying areas near or within the maximum extent of Wisconsin glaciation (Figure 33), display moderate to high amounts of genetic diversity. The Lake Sakakawea population appears to have existed for a period of time sufficient to develop greater genetic diversity. The contemporary population of Lake Sakakawea is most strongly correlated with the Ohio River according to F_{ST} analysis, even though it lies along the Missouri River (Table 18). This may suggest gene flow via the Mississippi River. However, observation of haplotype groups and the minimum spanning network show the populations of Lake Sakakawea and the Missouri River to be highly similar (Figure 24, 26, 27; Table 9, 10). Many haplotypes

found in the Lake Sakakawea and Missouri Rivers are found within the same group, particularly group C in parsimony analysis and group E in Bayesian analysis. Additionally, ND2 analysis reveals three haplotypes of the Lake Sakakawea population are shared with the Missouri River, where only two are shared with the Ohio River (Table 3, 13). Lake Pepin is strongly correlated with the previously glaciated populations of Lake Ontario and Lake Winnipeg, and displays great haplotype diversity. The populations of the Missouri and Ohio Rivers display the highest level of diversity and are likely the source populations for freshwater drum, as will be discussed in the following section. The historical relationships between these populations largely contributed to the contemporary structure.

Historical relationships and postglacial dispersal of freshwater drum

Fossil records of freshwater drum dating to the Pliocene have been found in Nebraska and in the lower Ohio – upper Mississippi River region, dating the presence of freshwater drum in North America to at least 2.5 million years ago (Burr and Page, 1986; Starnes and Etnier, 1986). Based upon the observed distribution of freshwater drum populations, Barney (1926) proposed that freshwater drum originated from Sciaenids in the Gulf of Mexico and colonized North America via the Mississippi River. Subsequent to entering freshwater North America through the Mississippi River it is likely that freshwater drum colonized the Teays-Mahomet river system to the east and the Missouri River to the west (Figure 6, 7, 9). Multiple species of freshwater fishes were widely distributed in the Teays, upper Mississippi and old Missouri Rivers, all of which were early centres of evolution and diversification for many fish species (Crossman and McAllister, 1986; Hocutt et al., 1986; Robison, 1986). A separation of freshwater fishes into western and eastern clades is commonly observed in phylogeographic studies (Berendzen et al.,

2003; Borden and Krebs, 2009). This east/west division is referred to as the pre-Pleistocene vicariance hypothesis (Mayden, 1988).

The great genetic diversity observed in my analyses of the Ohio River (a remnant of the ancient Teays system), and the Missouri /Mississippi River suggests that drum have inhabited these regions continually since the Pliocene (Table 2, 12). The inference of freshwater drum inhabiting the Teays (Ohio) River and Missouri/Mississippi is supported by fossil evidence (Burr and Page, 1986; Starnes and Etnier, 1986). Genetic variation accumulates in populations which have persisted over time and correlations are expected to be exhibited between founding and descendant populations (Hewitt, 1996). The Ohio and Missouri River populations share many haplotypes with other populations, consistent with the hypothesis that these are the primary colonizing populations for other sites sampled in this study, in accordance with other freshwater fish phylogeographic studies including those of hogsucker (Berendzen et al., 2003) and gilt darter (Near et al., 2001). Glaciation restricted the range of formerly widespread species which likely included freshwater drum. Freshwater fish populations became concentrated in unglaciated regions. Populations that were previously isolated by distance were now in geographic proximity and gene flow between these populations could occur. The freshwater drum population likely became more heterozygous over this period. As ice retreated species were able to disperse and reestablish themselves in newly available habitats (Crossman and McAllister, 1986; Cross et al., 1986.).

Throughout the Pleistocene the majority of fishes, including freshwater drum, persisted in the Mississippi River or tributaries south of the ice sheet (Hocutt et al., 1996, 2000). Refugia south of the ice sheet within the Mississippi drainage included regions of Ohio, Indiana, Illinois and Wisconsin (Figure 5) (Crossman and McAllister, 1986). The Teays River (the flow of which was

diverted throughout the Pleistocene to become the Ohio River) and Missouri River south of the ice sheet were also areas of refuge for fishes (Figure 8, 10). Analysis derived from my study is consistent with the hypotheses of Mayden (1988) of pre-Pleistocene vicariance, as evidenced by the observed suture line splitting the freshwater drum population into eastern and western haplotype groups (Figure 36). Prior to the Pleistocene the Central Highlands region (which is comprised of the Ouachita and Ozark mountains west of the Mississippi River and the Eastern Highlands to the east of the Mississippi River) was one continuous region (Mayden, 1988). Glaciation altered the topography of this region, both by till deposited by glacial advances, and isostatic uplift (Hewitt, 1996, 2000). This split the Central Highlands region, preventing gene flow between populations to the east and west of the Mississippi River. Isolation of populations east and west of the Mississippi resulted in genetic divergence, the effect of vicariance. Berendzen et al. (2008) hypothesized an event of vicariance occurring approximately three million years ago in hogsucker populations to the east and west of the Mississippi River. Data from the hogsucker study indicated two clades were contained within the western population and three clades were found to the east of the Mississippi (Berendzen et al. 2008). My study is consistent with this hypothesis, adding to our understanding of the patterns of postglacial biogeography of freshwater fishes in central North America. My data and other studies suggest the Ohio River was likely a primary colonization source for many freshwater fishes (Mayden, 1988; Mandrak and Crossman, 1992; Borden and Krebs, 2009). Freshwater drum originating from the Teays/Ohio River colonized the lower Great Lakes via the Allegheny and Monongahela Rivers and the Fort Wayne outlet, and colonized the Pickwick Reservoir via the Tennessee River. The Teays/Ohio River also contributed to the population of Lake Maloney via the Mississippi River. The Missouri River population (in my study located at the confluence of the Mississippi River) was also a primary colonization source, as identified in previous phylogeographic studies

(Near et al., 2001; Ray et al., 2006). Freshwater drum populations in Lake Pepin, Lake Winnipeg and Nelson River were colonized directly via the Mississippi River. Lake Sakakawea was colonized from the south via the Missouri River, and was also colonized by the Mississippi River via Lake Agassiz. Green Bay, Lake Maloney and Pickwick Reservoir were also colonized by this central population.

Near et al. (2001) determined a major clade within the population of gilt darter was found within the Missouri River drainage and the upper Mississippi drainage. Ray et al. (2006) observed a rainbow darter clade composed of populations inhabiting tributaries of the Missouri River and the upper Midwest (including Hudson Bay, the upper Mississippi River and western Lake Michigan). The primary source of the populations within these clades is the Missouri River. Zoogeographic studies, including that of Cross et al. (1986), determined that the Missouri River has a high level of species diversity and displays a great affinity with the Central Lowlands. Together these data suggest that the Missouri River/Mississippi River drainage population is a colonizing source for populations existing primarily to the west of the Mississippi River.

Populations sampled from the Ohio and Missouri Rivers are the most diverse of all populations sampled in this study, and appear to be ancestral as evidenced by the variety of haplotypes, number of shared haplotypes and vast dispersal of haplotype groups (Figure 24, 26, 29, 31, 34, 35, 37, 38). These populations also do not show evidence of genetic bottlenecks or recent population expansions, as would be expected for source populations that have persisted over time (Table 4, 5, 14, 15). The Ohio and Missouri River populations were likely the primary colonization source of other populations sampled in my study. This is supported not only by biogeographic and zoogeographic data, along with fossil evidence (Bailey and Smith, 1981; Burr and Page, 1986; Crossman and McAllister, 1986; Hocutt et al., 1986; Robison, 1986; Starnes and

Etnier, 1986) but also numerous phylogeographic studies. Each population in my study was ultimately colonized by a central population derived from the Missouri/Mississippi River including the Teays/Ohio River. However, the Teays/Ohio River was a major colonization source of eastern populations and the Missouri/Mississippi River was the primary source for colonization of western populations. An excellent example is provided by the study of gilt darter by Near et al. (2001) in which they hypothesized that an eastern clade was colonized by the Ohio River and a western clade was colonized by the Mississippi and Missouri Rivers. My data indicate the Missouri/Mississippi River is the primary source population, contributing haplotypes to each population, including the Ohio River (Figure 34, 35). Data obtained in my study further suggest the Ohio River, which was likely colonized by members of the Mississippi River population, is a secondary source for populations in the Great Lakes and Pickwick Reservoir (Table 8, 9). It is likely the Mississippi River is the colonizing source of Lake Maloney, Pickwick Reservoir and the Ohio River, which are the only members of a distinct haplotype group (Figure 27, Table 9, 10). A common haplotype in the Missouri/Mississippi River may have been lost at some point in time or was not detected in our sample. These data confirm my hypothesis "Ohio River and Missouri River are source populations for other sampled populations". This is further consistent with data produced in numerous studies including that of Borden and Krebs (2009) who studied the smallmouth bass population. They hypothesized the source populations which colonized the western and eastern distributions of the species to be the Mississippi and Ohio Rivers respectively, consistent with my data.

The freshwater drum population of Lake Maloney displays increased nucleotide and haplotype diversity suggesting the population has survived for sufficient time to significantly diverge from other sampled populations (Table 2, 12). Tau values suggest this population existed prior to the end of the last glacial maximum and may have served as a refugium. The Niobrara River valley

in Nebraska (a tributary of the Missouri) has been hypothesized as a refugium in the Pleistocene (Figure 10) (Kaul et al., 1988). Lake Maloney is a reservoir formed by damming of the South Platte River, also a tributary of the Missouri River. Freshwater drum occupied regions of the south and central Great Plains prior to the Pleistocene glaciations (Cross et al., 1986). Cross et al. (1986) hypothesized that fishes in this region were founded by populations originating from the Mississippi and Teays Rivers. Data from my study agree with this hypothesis as many haplotypes are shared between the sampled Missouri/ Mississippi River population. There is also a distinct relationship between the freshwater drum populations of the Ohio River (which is derived from the ancient Teays River) and Lake Maloney (Figure 27; Table 9, 10). Lake Maloney significantly differs from virtually all other populations and the majority of haplotypes are not found in any other sampled populations, supporting the proposal that the Lake Maloney population was isolated from the other populations, consistent with the refugium hypothesis (Table 5, 15). Examination of the control region minimum spanning network validates the inference of vicariance prior to the Pleistocene followed by a period of isolation, as great genetic distance is observed between the primary control region haplotype of Lake Maloney (h0007) and the main population comprised primarily of populations within the Mississippi drainage system (Figure 27, 32). This is also observed in the ND2 minimum spanning network, a large proportion of the Lake Maloney population is found within the second haplotype group, which is also separated from the main population. Members of the modern day Lake Maloney population likely survived the Wisconsin glaciation in a small isolated region of a Missourian refugium, likely the Niobrara River (Figure 10), resulting in the population differentiating over time producing the unique haplotype distribution revealed in my study. This interpretation is supported by Kaul et al. (1988) who reported on numerous atypical plant and animal communities found in the Niobrara River valley. Plant and animal distribution within the

Niobrara River valley along with fossil evidence suggest that historical factors account for the observed faunal distribution leading to the proposal of the Niobrara River valley as a glacial and postglacial refugium within the central plains (Kaul et al., 1998). The lake chub (*Couesius plumbeus*) and blacknose shiner (*Notropis heterolepis*) are thought to be glacial relicts which survived in the Niobrara refugium throughout the Pleistocene (Kaul et al., 1988). A phylogeographic study of killifish also determined the haplotypes found within the Platte River to be confined to a smaller clade, consistent with my data (Kreiser et al, 2001). My hypothesis "the Lake Maloney population was colonized by members of the Mississippi/Missouri River drainage" must be refined to include the Ohio River population. The hypothesis must be further refined to include the isolation of freshwater drum within the Niobrara refugium, over which time the population genetically differentiated becoming unique and displaying the effects of vicariance.

The freshwater drum population of Lake Sakakawea exhibits increased nucleotide and haplotype diversity (Table 2, 12), tau values further suggest this population existed prior to the end of the last glacial maximum. Lake Sakakawea shows evidence of a demographic expansion occurring prior to the end of the Wisconsin glaciation 18000 years ago (Table 5). ND2 analysis indicates Lake Sakakawea shares the greatest number of haplotypes with the Missouri River population and was sourced from the Missouri River population south of Lake Sakakawea (Table 3). Three ND2 haplotypes (h1, h15, h110) were observed in the Lake Sakakawea population, all three of which are shared with the Missouri River population. This illustrates a historical relationship between the Missouri River population and Lake Sakakawea. Glacial retreat in the Great Plains would have provided further access for fishes from the Missouri River. A pattern of colonization from a central region is indicated by shared haplotypes of the control region (Figure 35, Appendix 3). Lake Agassiz extended into North Dakota, making this the most likely route for

fish originating from Mississippian refugia (Crossman and McAllister, 1986). Lake Agassiz extended into North Dakota and was connected to the Mississippi River via the Warren outlet between 12800 and 10000 years ago. This provided a dispersal route for fishes from the Mississippi River into what is now Lake Sakakawea. My data indicating colonization from Mississippian populations agree with the hypotheses of Grzybowski et al. (2009) for populations of yellow perch. My hypothesis "the Lake Sakakawea population was colonized by members of the Mississippi/Missouri River drainage" is supported.

Lake Pepin is located along the Mississippi River in Minnesota. The Lake Pepin population shares control region haplotypes with every other population, and a negative value of Tajima's D indicates population expansion. Control region F_{ST} values indicate Lake Pepin is correlated with Lake Ontario, Lake Winnipeg and Pickwick Reservoir (Table 8) whereas ND2 analysis shows Lake Pepin to be correlated with every population other than Green Bay, Lake Maloney and the Missouri River (Table 18). Together these data indicate a high degree of genetic diversity originating from the Mississippi River (Lake Pepin is a widening of the Mississippi River). Analyses of data collected in my study indicate the source population of the Mississippi drainages expanded its range post-glaciation to invade newly available habitats such as Lake Pepin (Table 5; Appendix 4, 5). This inference differs from results of Borden and Krebs (2009) smallmouth bass study where they hypothesized the upper Mississippi is an isolated region. The older, less genetically diverse smallmouth bass population of the upper Mississippi River is bound geographically with the upper Great Lakes via numerous tributaries, including the Illinois and Wisconsin River (Benke and Cushing, 2005; Borden and Krebs, 2009). Borden and Krebs (2009) proposed the observed structure of separate clades east and west of the Mississippi River in smallmouth bass populations is a result of vicariance. Data observed in my study of freshwater drum suggests post-glacial colonization by the genetically diverse source population

of the Mississippi River, particularly evidenced by the large number of control region haplotypes (Figure 43; Appendix 3). My hypothesis "the Lake Pepin population was colonized by members of the Mississippi/Missouri River drainage" is supported.

Crossman and McAllister (1986) hypothesized that freshwater drum in the Hudson Bay drainage system originated from the Mississippi River. Evidence from my study is consistent with this hypothesis as Lake Winnipeg shares haplotypes with populations originating from the Mississippi River. The Nelson River and Lake Winnipeg populations are strongly correlated with each other, both populations share haplotypes with the Missouri/Mississippi River and Lake Pepin (Figure 34, 35; Table 3, 13). Additionally, F_{ST} values indicate a particularly strong correlation between Lake Pepin and Lake Winnipeg, consistent with dispersal from the Mississippi River, including geographically proximal Lake Pepin. Lake Winnipeg was likely colonized by members of the Mississippi River, including Lake Pepin when the Mississippi drainage connected to Lake Agassiz. The Nelson River was subsequently colonized by members of the Lake Winnipeg population. After the Mississippi River/Lake Agassiz connection disconnected approximately 10000 years ago Lake Winnipeg and Nelson River became isolated. Both populations show evidence of a recent population bottleneck based upon Tajima's test of neutrality and The Nelson River population in particular displays limited genetic diversity (Table 5, 15). Limited genetic diversity has been observed in walleye populations from Lake Winnipeg drainage. Stepien et al. (2009) observed lower than average heterozygosity (0.587 vs. 0.698) and fewer alleles than average (48 vs. 139) in the walleye population sampled from the Lake Winnipeg drainage (Stepien et al., 2009). Genetic data from my study and that of Stepien et al. (2009) suggest Nelson River and Lake Winnipeg were recently colonized, not unexpected given the fact this region was completely glaciated until 12800 years ago (Crossman and McAllister, 1986). My data, which indicate recent colonization by a small number of individuals, are

consistent with both genetic data (Stepien et al., 2009) and biogeographic data indicating the Lake Winnipeg and Nelson River was glaciated until 12800 years ago (Bailey and Smith, 1981). My hypothesis "populations of the Hudson Bay drainage were colonized by members of the Mississippi/Missouri River drainage" is supported.

Lake Erie and Lake Ontario have many commonalities. Colonization likely occurred shortly after the ice of the last glacial maximum began to recede. Lake Maumee, from which Lake Erie is derived, began to form 14000 years ago and Lake Iroquois, the predecessor to Lake Ontario, began to form 12700 years ago (Bailey and Smith, 1981). There are multiple routes by which Lake Erie and Lake Ontario were colonized. The first route into the lower Great Lakes was the Allegheny and Monongahela Rivers which led from the Ohio River to Lake Erie (Bailey and Smith, 1981). The Upper Allegheny and Monongahela Rivers were tributaries of the preglacial Pittsburgh River which flowed into the Lake Erie basin (Hocutt et al., 1986) (Figure 12). Entry through either of these routes is plausible, and may have been a source of entry for at least a portion of the Lake Erie and Lake Ontario freshwater drum population. However, the most likely source of entry is the Fort Wayne outlet as this connected Lake Erie to the Wabash valley at various times throughout the glacial period and is hypothesized as the route of colonization for many freshwater fishes, including smallmouth bass and walleye (Mandrak and Crossman, 1992; Borden and Krebs, 2009; Stepien et al., 2009). The Fort Wayne outlet has also been identified as a dispersal route for freshwater mussels, freshwater drum is the only known host for the larval form of some species of freshwater mussels and the only means by which mussels can disperse upstream. As the Fort Wayne outlet is a dispersal route for freshwater mussels this leads greater credibility to my inference of the Fort Wayne outlet as the primary entry for freshwater drum into Lake Erie (Stewart and Watkinson, 2004; Borden and Krebs, 2009). Lake Erie and Lake Ontario are correlated with many other populations, particularly populations associated with

the Mississippi River (Table 8, 18). Tajima's D values indicate population expansion, as do mismatch distribution values for the Lake Ontario population (Table 14, 15). I hypothesize that Lake Ontario increased in population numbers as the ice receded. Lake Erie displays lower diversity, likely a result of geographic isolation from Lake Ontario, particularly before the Welland canal was built, resulting in genetic differentiation. Previously, the Lake Erie population could not be accessed by freshwater drum inhabiting Lake Ontario, however entry to Lake Ontario could occur via Niagara Falls, particularly since freshwater drum eggs are planktonic (Berra, 2001). Multiple populations likely originating from the fragmented Teays River (and present day Ohio River) colonized Lake Erie and Lake Ontario. The primary route of entry was likely the Fort Wayne outlet which drained Lake Erie into the Wabash and Maumee Rivers across the states of Illinois, Indiana and Ohio. Lake Erie and Lake Ontario were also potentially accessed by the Monongahela River and the Allegheny River draining from the Ohio River. This inference is supported by other phylogeographic studies, particularly that of walleye (Stepien et al., 2009) which observed a close genetic relationship between Lake Erie and Lake Ontario based on shared haplotypes.

The freshwater drum population of Green Bay exhibits increased nucleotide and haplotype diversity (Table 2, 12). Furthermore, higher than average tau values also suggest this population has existed for a period of time much greater than observed in the other Great Lakes populations, suggesting colonization occurred prior to the end of the Wisconsin glaciation (Table 4, 14). Glacial Lake Wisconsin, located near present day Green Bay, may have served as a refugium throughout the Pleistocene. The advancing Green Bay lobe in the Wisconsin glaciation met the Baraboo Hills blocking the Wisconsin River to form glacial Lake Wisconsin. Lacustrine deposits were first recognized by Warren in 1874, a report of the shoreline and drainage outlets of glacial Lake Wisconsin based on geological evidence was proposed by Clayton and Attig

(1989) (Clayton and Knox, 2008). There is great evidence for the existence of glacial Lake Wisconsin over the course of the Wisconsin glaciation, it is reasonable that this could have been a refugium for freshwater drum inhabiting the Wisconsin River (which is located in an unglaciated area) (Figure 5). Freshwater drum may have survived in this region prior to the Wisconsin glaciation and subsequently diverged, resulting in the unique gene composition observed and genetic differentiation (Table 8, 18) (Clayton and Knox, 2008). The Green Bay freshwater drum population is different from others, not based only on the three unique control region haplotypes but the degree of polymorphism (Figure 27,32). The distinct genetic composition of the Green Bay population is clearly observed in parsimony and Bayesian analysis. A large proportion of haplotypes (control region, 74.95%; ND2 region, 69.2%) are contained within the rarer haplotype groups (Table 9, 10, 19, 20). Although Green Bay shares one control region haplotype with Lake Ontario, and two haplotypes with Lake Erie, the frequency in which these haplotypes are represented greatly differs. The predominant haplotype of Green Bay (frequency 0.562) is shared with only Lake Erie (frequency 0.1), indicating founder effect resulting from colonization by a small group with limited genetic variation. The population of Green Bay also shares the predominant control region haplotype of Lake Ontario (0.429) and Lake Erie (0.8) at a frequency of 0.188. Shared haplotypes exist, yet overall the control region genetic composition of Green Bay differs greatly in haplotype frequency from the lower Great Lakes populations. Two haplotypes are unique to the Great Lakes freshwater drum population, indicating gene flow has occurred, however there is a high degree of genetic differentiation observed between Green Bay and the lower Great Lakes (Table 8). Control region analyses of the minimum spanning networks (Figure 27, 32) reveal that the predominant haplotypes of Green Bay show great genetic distance from the main population, primarily composed of populations within the Mississippi drainage basin. ND2 analysis also

suggests the Green Bay population is highly distinct from all other populations based on F_{ST} analysis and haplotype distribution (Table 18, Appendix 5). Haplotype group analysis shows the population of Green Bay to be primarily contained within the second group. Parsimony analysis indicates 69.2% of the Green Bay population is in this haplotype group, which also includes Lake Erie (frequency 0.056), Lake Ontario (0.143) and Lake Pepin (0.0714) at low levels. Missouri River and Lake Maloney are also found within this haplotype group at frequencies of 0.3335 and 0.384 respectively (Table 19). Bayesian analysis of ND2 haplotypes is similar to that of parsimony analysis. Green Bay, Lake Erie, Lake Ontario, Lake Pepin and Missouri River populations are distributed in this haplotype group at the same frequency as in parsimony analysis. Pickwick Reservoir is also included in this haplotype group based on Bayesian analysis, Lake Maloney is included at a much higher frequency (0.846) than observed in parsimony analysis. The population of Green Bay is correlated with both Mississippian and Missourian refugia as, ultimately, are all the other populations. The degree of differentiation, including haplotype frequencies unlike any other population and increased tau values, indicates the population of Green Bay was colonized prior to the ice age. It is most likely that freshwater drum existed in glacial Lake Wisconsin throughout the ice age and subsequently entered Lake Chicago via the Fox and Wisconsin Rivers (Figure 11) (Clayton and Knox, 2008). As glacial retreat continued, Lake Michigan connected to Lake Erie via the Indian River lowlands allowing for gene flow between these populations (Underhill, 1986). Limited connections exist(ed) between the upper and lower Great Lakes which had the effect of restricting gene flow (Stepien et al., 2009). This may have also affected genetic composition of the Great Lakes freshwater drum populations, as populations have further diverged. Borden and Krebs (2009) hypothesized that the Lake Michigan smallmouth bass population was colonized via the Chicago outlet and further determined the Lake Michigan smallmouth bass population differs greatly from populations

found in the lower Great Lakes. Their data support my interpretation of the Green Bay population being distinct from the Lake Erie and Lake Ontario populations. Borden and Krebs (2009) further theorized that in the Great Lakes system a shared drainage basin is relatively unimportant in explaining shared alleles, rather shared alleles result from sequential dispersal events from multiple refugia. They determined that smallmouth bass in the upper Great Lakes and upper Mississippi are older and not very diverse, and are connected geographically. Borden and Krebs (2009) theorized that the smallmouth bass populations of Lake Erie and Lake Huron were colonized by multiple sources, Lake Ontario was subsequently colonized by the Lake Erie population. These results concur with my data indicating separate colonization sources for the upper and lower Great Lakes. My hypothesis "Great Lakes populations were colonized by members of the Mississippi/Teays (Ohio) River drainage" must be refined to state "the lower Great Lakes populations were primarily colonized by members of the Mississippi/Teays (Ohio) River drainage, gene flow also occurs between the lower and upper Great Lakes provides " and "the Green Bay population was primarily colonized by members of the Mississippi and Missouri drainages, gene flow also occurs between the lower and upper Great Lakes ".

Late in the Pliocene (more than 2.5 million years ago) seawater rose to a level where it extended to near Alabama (Robison, 1986). In this time period, as reflected in Figure 6, the Mississippi embayment extended north of the Tennessee border. This may be the event which sourced drum from the marine system. There would have been a large amount of brackish water which then receded, isolating populations and resulting in the existing freshwater drum population. As Pickwick Reservoir is located on the northwestern coast of Alabama (Figure 15) it is highly probable that a Sciaenid population, which became freshwater drum, was colonized in this area. At some point in time over the Pleistocene the freshwater drum population Tennessee River became extinct. The Pickwick Reservoir is located along the Tennessee River, which is part of

the contemporary Ohio River basin. The Pickwick Reservoir freshwater drum population contains two unique control region haplotypes (h0609, h6134) and shares control region haplotypes with all populations other than Lake Ontario and Lake Erie (Table 3). The Tennessee River is a major tributary of the Mississippi River. The haplotypes which Pickwick Reservoir shares with other populations create a pattern reflecting colonization originating from a central source (the Mississippi River) (Figure 35). Control region analysis show Pickwick to be spread across a number of haplotype groupings, representative of multiple colonizing sources (Figure 24, 26; Table 9, 10). Shared haplotypes between Green Bay and the pattern of shared haplotypes along the Mississippi River suggests it to be a colonizing source, further analysis of shared haplotypes reveals shared haplotypes between Pickwick Reservoir and the Ohio River. Tau values (Table 4, 14) suggest_Pickwick Reservoir was recently colonized, in the same time period as the glaciated regions of the Great Lakes and Hudson Bay were colonized. The most plausible explanation for my data is that the freshwater drum population was colonized after the last glacial maximum, via the Ohio and Mississippi Rivers (Burr and Page, 1986). The primary source population appears to be the central Mississippi population, like that observed in previously glaciated regions (Figure 34, 35). Shared haplotypes were also observed between Pickwick Reservoir and the Ohio River. The large number of haplotypes in this population indicates colonization by multiple sources, or a genetically diverse population (such as the Mississippi River) (Figure 47). This inference is similar to results of the study of bigeye chub phylogeography by Berendzen et al. (2008). Based upon the discovery of a haplotype determined to be a member of a clade of northeastern haplotypes, Berendzen et al. (2008) proposed that postglacial dispersal was not only northward, but also southward. This is correlated concurs with my data indicating a genetic relationship between Pickwick Reservoir and populations existing in regions that were previously glaciated. My hypothesis "the Pickwick

Reservoir population was colonized by members of the Mississippi/Teays River drainage" should be refined to include the Missouri/Mississippi River.

In conclusion, Barney (1926) suggested freshwater drum originated from Sciaenids in the Gulf of Mexico. Data from this study suggests that the Mississippi River was the centre of the freshwater drum population which subsequently colonized the Teays and old Missouri Rivers. As ice advanced during the Pleistocene, populations to the east and west or within the Mississippi River became concentrated in regions south of the last glacial maximum. Some populations became isolated from this central population in refugia of the Niobrara River and glacial Lake Wisconsin. The populations within these refugia were isolated during glaciation and genetically diverged as a result of vicariance. Evidence of a Mississippi River discontinuity has been observed in many phylogeographic studies as summarized by Soltis et al. (2006). Soltis et al. (2006) observed that many phylogeographic studies of freshwater fishes exhibited evidence of a vicariant event which separated populations into eastern and western lineages, which is in agreement with my data. The routes used to invade these new habitats were inferred by genetic analysis. Upon glacial retreat the main population occupying central North America expanded its range into Lake Pepin increasing in numbers and dispersed into new areas exposed by glacial retreat. The lower Great Lakes were primarily colonized by populations originating from the Teays-Mississippi River, who entered Lake Ontario and Lake Erie primarily through the Fort Wayne outlet beginning 14000 years ago. The Allegheny and Monongahela Rivers and Lake Michigan also likely contributed colonizing members to the lower Great Lakes. The freshwater drum population of Lake Sakakawea was colonized by members of the Missouri/Mississippi River population. The Lake Pepin population was colonized by freshwater drum upon glacial retreat directly via the Mississippi River (no sooner than 18000 years ago) as this central population expanded its range, Missourian populations may have also colonized Lake Pepin

through Lake Agassiz no sooner than 12800 years ago. Lake Winnipeg was colonized primarily via the central Mississippi River population (including Lake Pepin), and the Nelson River was subsequently colonized by dispersal from the Lake Winnipeg population. The population of Green Bay was likely colonized initially by members of the Missouri River population which subsequently entered glacial Lake Wisconsin early in the Pleistocene and survived the remaining glacial period in this refugium. The freshwater drum population inhabiting Lake Maloney was also isolated throughout the Pleistocene, likely in a Nebraska refugium. The vast genetic variation observed in freshwater drum is explained by a complex pattern of pre-Pleistocene vicariance to the east and west of the Mississippi River followed by post-glacial dispersal originating from the Mississippi, Missouri and Ohio Rivers.

FUTURE STUDIES

Future research may involve the comparison of the phylogeography of freshwater drum and other North American fish species to elucidate patterns of dispersal among freshwater fishes and determine the degree of differentiation between species. This could enhance knowledge of the biogeography of North America prior to the Pleistocene, particularly in regards to drainage systems and cryptic refugia. There have been a multitude of phylogeographic studies of various freshwater fish species. Studies of species with geographic ranges similar to that of freshwater drum, including walleye, would provide an excellent comparative basis of colonization patterns. Species with different geographic ranges, habitat preferences and environmental tolerances will likely show significant differentiation but will provide a comparative basis to determine widespread patterns amongst freshwater fishes. Wider interspecies studies, including other phyla, will provide an even broader range of comparison.

Certain geographic regions in this study presenting a unique genetic composition should be investigated further. For example, the area around the Niobrara River valley appears to have an atypical composition across a variety of fauna, a complete lack of phylogeographic studies in this area exists. The region of Lake Sakakawea also merits further investigation. A Missourian refugia in Montana, near Lake Sakakawea, has been hypothesized by Crossman and McAllister (1986) with fauna originating from Beringia. However, data in my study suggest there may have been a refugium colonized from the Missouri River. The populations of Green Bay and Pickwick Reservoir also appear to be atypical. The Green Bay population appears to have existed in a refugium in glacial Lake Wisconsin, and gene flow appears to be limited between the upper and lower Great Lakes. Conversely, the population of Pickwick Reservoir appears to have been colonized subsequent to glacial retreat, even though this region was not glaciated during the Wisconsin glaciation.

Additionally, studies could be employed to determine the marine sister taxon of freshwater drum. Evidence suggests this species arose from Sciaenids in the Gulf of Mexico, at this time it is unknown what species of Sciaenidae shares a common ancestor with *A. grunniens*. Genetic analysis, including sequencing of slowly evolving regions of mitochondrial DNA such as cytochrome oxidase subunit 1, could result in inferring the sister taxon to freshwater drum (Lopez et al., 1997).

Further phylogeographic studies could be performed using nuclear DNA or microsatellites. This would broaden analysis to include paternal DNA. This may be significant as female freshwater drum have been shown to be highly motile in comparison to males of this species (Rypel, 2007). Different relationships between populations and patterns of dispersal may be revealed as a result.

Lastly, genetic analysis may be used to determine if there is evidence of co-evolution between freshwater drum and parasitic species. In many cases freshwater drum is the only known host for certain obligate parasites (Stewart and Watkinson, 2004). This may increase knowledge of colonization patterns of parasites, and of mutation in the face of co-evolution.

LITERATURE CITED

Abd-Elsalam, K.A. 2003. Bioinformatic tools and guidelines for PCR primer design. African Journal of Biotechnology, **5:** 91-95.

Altukhov, Y.P. 2006. Intraspecific genetic diversity: monitoring, conservation and management. Springer Berlin Heidelberg, New York, NY, p. 16-146.

Avise, J.C. 1998. The history and purview of phylogeography: a personal reflection. Molecular Ecology, **7:** 371-379.

Avise, J.C., J. Arnold, R.M. Ball, E. Bermingham, T. Lamb, J.E. Neigel, C.A. Reeb and N.C. Saunders. 1987. Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematics. Annual Review of Ecology and Systematics, **18**: 489-522.

Avise, J.C. 2000. Phylogeography, the history and formation of species. Harvard University Press, Cambridge, MA, p. 9-32.

Avise, J.C. 2006. Evolutionary pathways in nature: a phylogenetic approach. Cambridge University Press, New York, NY, p. 10-29, 44-78.

Bailey, R.M. and G.R. Smith. 1981. Origin and geography of the fish fauna of the Laurentian Great Lakes basin. Canadian Journal of Fisheries and Aquatic Sciences, **38:** 1539-1561.

Barney, R.L. 1926. The distribution of the freshwater sheepshead, *Aplodinotus grunniens* Rafinesque, in respect to the glacial history of North America. Ecology, **7:** 351-364.

Barr, C.M., M. Neiman and D.R. Taylor. 2005. Inheritance and recombination of mitochondrial genomes in plants, fungi and animals. New Phytologist, **168:** 39-50.

Benke, A.C. and C.E. Cushing. 2005. Rivers of North America. Elsevier, Burlington, MA, p. 231-482, 853-902.

Berendzen, P.B., A.M. Simons, and R.M. Wood. 2001. Phylogeography of the northern hogsucker *Hypentelium nigricans* (Teleostei: Cypriniformes): genetic evidence for the existence of the ancient Teays River. Journal of Biogeography, **30:** 1139-1152.

Berendzen, P.B., Gamble, T. and 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.

Bernatchez, L. and C.C. Wilson. 1998. Comparative phylogeography of Nearctic and Palearctic fishes. Molecular Ecology, **7**: 432-452.

Berra, T.M. 2001. Freshwater fish distribution. Academic Press, San Diego, CA. p. 417-420.

Borden, W.C. and R.A. Krebs. 2009. Phylogeography and postglacial dispersal of smallmouth bass (*Micropterus dolomieu*) into the Great Lakes. Canadian Journal of Fisheries and Aquatic Sciences, **66:** 2142-2156.

Brito, H.B. and S.V. Edwards. 2009. Multilocus phylogeography and phylogenetics using sequence based markers. Genetica, **135**: 439-455.

Broughton, R.E. and P.C. Reneau. 2006. Spatial covariation of mutation and nonsynonymous substitution rates in vertebrate mitochondrial genomes. Molecular Biology and Evolution, **23:** 1516-1524.

Brown, J.R., A.T. Beckenbach and M.J. Smith. 1993. Intraspecific DNA sequence variation of the mitochondrial control region of white sturgeon (*Acipenser transmontanus*). Molecular Biology and Evolution, **10:** 326-341.

Burr, B.M and R.L. Mayden. 1992. Phylogenetics and North American freshwater fishes. Systematics, historical ecology, and North American freshwater fishes, edited by R.L. Mayden. Stanford University Press, Stanford, CA. p. 18-75.

Burr, B.M. and L.M. Page. 1986. Zoogeography of fishes of the lower Ohio – upper Mississippi basin. The zoogeography of North American freshwater fishes, edited by Hocutt, C.H and E.O. Wiley. John Wiley & Sons, Inc. New York, NY. p. 287-324.

Cantatore P., M. Roberti, G. Pesole, A. Ludovico, F. Milella, M.N. Gadaleta and C. Saccone. 1994. Evolutionary analysis of cytochrome b sequences in some Perciformes: evidence for a slower rate of evolution than in mammals. Journal of Molecular Evolution, **39:** 589-597.

Cheng, Y.Z., T.J. Xu, X.X. Jin and R.X. Wang. 2011. Complete mitochondrial gemome of the yellow drum *Nibea albiflora* (Perciformes, Sciaenidae). Mitochondrial DNA, **4:** 80-82.

Clayton, J.A and J.C. Knox. 2008. Catastrophic flooding from glacial Lake Wisconsin. Geomorphology, **93:** 384-397.

Coad, B.W. 1995. Encyclopedia of Canadian fishes. Canadian Museum of Nature and Canadian Sportfishing Productions, Singapore. p. 299-300.

Cross, F.B., R.L. Mayden and J.D. Stewart. 1986. Fishes in the western Mississippi basin (Missouri, Arkansas and Red Rivers. The zoogeography of North American freshwater fishes, edited by Hocutt, C.H and E.O. Wiley. John Wiley & Sons, Inc. New York, NY. p. 363-412.

Crossman, E.J. and D.E. McAllister. 1986. Zoogeography of freshwater fishes of the Hudson Bay drainage, Ungava Bay and the Arctic archipelago. The zoogeography of North American freshwater fishes, edited by Hocutt, C.H and E.O. Wiley. John Wiley & Sons, Inc. New York, NY. p. 53-85.

Cui, Z., Y. Liu, C.P. Li, F. You, K.H. Chu. 2009. The complete mitochondrial genome of the large yellow croaker, *Larimichthys crocea* (Perciformes, Sciaenidae): Unusual features of its control region and the phylogenetic position of the Sciaenidae. Gene, **432:** 33-43.

Currat, M., L. Excoffier, W. Maddison, S.P. Otto, N. Ray, M.C. Whitlock and S. Yeaman. 2006. Comment on: "Ongoing adaptive evolution of ASPM, a brain size determinant in Homo sapiens" and Microcephalin, a gene regulating brain size , continues to evolve adaptively in humans". Science, **313:** 172.

Daiber, F.C. 1950. The life history and ecology of the sheepshead, *Aplodinotus grunniens* Rafinesque, in western Lake Erie. Ph.D. dissertation. Columbus, OH, The Ohio State University, 88p.

Day, J.J., Santini, S. and J. Garcia-Moreno. 2007. Phylogenetic relationships of the Lake Tanganyika cichlid tribe Lamproligini: the story from mitochondrial DNA. Molecular Phylogenetics and Evolution, **45:** 629-642.

Dieffenbach, C.W., T.M. Lowe and G.S. Dveksler. 1993. General concepts for PCR primer design. Genome Research, **3:** 30-37.

Ehlers, J. and P.L. Gibbard. 2004. Quaternary glaciations: extent and chronology, part II: North America. Elsevier, Burlington, MA, p. 411-428.

Emerson, B.C. and G.M. Hewitt. 2005. Phylogeography. Current Biology, **15:** 367-371.

Engels, W.R. 1993. Contributing software to the internet: the Amplify program. Trends in Biochemical Sciences, **18:** 448-450.

Excoffier, L . 2004. Patterns of DNA sequence diversity and genetic structure after a range expansion: lessons from the infinite-island model. Molecular Ecology, **13:** 853-864.

Excoffier, L., P. Smouse and J. Quattro. 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. and H.E.L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Molecular Ecology Resources, **10:** 564-567.

Felsenstein, J. 2009. Phylip (Phylogeny Inference Package) version 3.6.9. Distributed by the author. Department of Genome Sciences, University of Washington, Seattle.

Freeman, S. and J.C. Herron. 2007. Evolutionary analysis, fourth edition. Pearson Education, Upper Saddle River, NJ, p. 73-94, 143-355.

Gaggiotti, O.E. and L. Excoffier. 2000. A simple method of removing the effect of a bottleneck and unequal population sizes on pairwise genetic differences. Proceedings of the Royal Society of London, **267:** 81-87.

Gibbons, A. 1998. Calibrating the mitochondrial clock. Science, **279:** 28-29.

Gold, J.R., L.R. Richardson, C. Furman and T.L. King. 1993. Mitochondrial DNA differentiation and population structure in red drum (*Sciaenops ocellatus*) from the Gulf of Mexico and Atlantic Ocean. Marine Biology, **116:** 175-185.

Goldthwait, R.P. 1991. The Teays Valley problem; a historical perspective. Geology and hydrogeology of the Teays-Mahomet River bedrock valley system, edited by Melhorn, W.N. and Kempton, J.P. The Geological Society of North America, Boulder, CO. p. 3-8.

Hartl, D.L. and A.G. Clark. 1997. Principles of population genetics, fourth edition. Sinauer Associates, Inc., Sunderland, MA, p. 62-235, 325-411

Hasegawa, M., H. Kishino and T. Yano. 1985. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in homonoidea. Journal of Molecular Evolution, **29:** 170-179.

Hewitt, G.M. 1996. Some genetic consequences of ice ages and their role in divergence and speciation. Biological journal of the Linnaen society, **58:** 247-276.

Hewitt, G. 2000. The genetic legacy of the Quaternary ice ages. Nature, **405:** 907-913.

Hocutt, C.H., R.E. Jenkins and J.R. Stauffer Jr. 1986. Zoogeography of the fishes of the central Appalachians and central Atlantic coastal plain. The zoogeography of North American freshwater fishes, edited by Hocutt, C.H and E.O. Wiley. John Wiley & Sons, Inc. New York, NY. p. 161-211.

Kaul, R.B., G.E. Kantak and S.P. Churchill. 1988. The Niobrara River valley, a postglacial migration corridor and refugium of forest plants and animals in the grasslands of central North America. The Botanical Review, **54:** 44-81.

Klug, W.S and M.R. Cummings. 2005. Essentials of Genetics, Fifth Edition. Pearson Education, Upper Saddle River, NJ, p. 38-84, 278, 322-328, 501-548.

Koblmüller, S., B. Egger, C. Sturmbauer and K.M. Sefc. 2010. Rapid radiation, ancient incomplete linage sorting and ancient hybridization in the endemic Lake Tanganyika cichlid tribe Tropheini. Molecular Phylogenetics and Evolution. **55:** 318-334.

Kocher,T.D. and White, T.J. 1989. Evolutionary analysis via PCR. In PCR Technology, Principles and Applications for DNA Amplification. Stockton Press, New York, NY. pp. 138-139.

Kreiser,B.R., J.B. Mitton and J.D. Woodling. 2001. Phylogeography of the plains killifish, *Fundulus zebrinus*. Evolution, **55:** 339-350.

Lee, W., J. Conroy, W.H. Howell and T.D. Kocher. 1995. Structure and evolution of teleost mitochondrial control regions. Journal of Molecular Evolution, **41**: 54-66.

Li, W. and Graur, D. 1991. Fundamentals of molecular evolution. Sinauer Associates, Inc. Sunderland, MA. pp. 86-87, 132-133.

Lopez, J.V., M. Culver, J.C. Stephens, W.E. Johnson and S.J. O'Brien. 1997. Rates of nuclear and cytoplasmic mitochondrial DNA sequence divergence in mammals. Molecular Biology and Evolution, **14:** 277-286.

Mandrak, N.E. and E.J. Crossman. 1992. Postglacial dispersal of freshwater fishes into Ontario. Canadian Journal of Zoology, **70:** 2247-2259.

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

Mettler, L.E., T.G. Gregg and H.E. Schaffer. 1988. Population genetics and evolution, second edition. Prentice-Hall, Englewood Cliffs, NJ, pp. 62-90.

Morrone, J.J. 2009. Evolutionary biogeography: an integrative approach with case studies, Columbia University Press. p. 1-113.

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

Nester, W.E., D.G. Anderson, C.E. Roberts, and M.T. Nester. 2007. Microbiology: a human perspective, fifth edition. McGraw-Hill, New York, NY. p. 219-242.

Okasha, S. 2008. Population genetics. The Stanford encyclopedia of philosophy, edited by E.N. Zalta, URL = <http://plato.stanford.edu/archives/fall2008/entries/population-genetics/>.

Posada, D. and K.A. Crandall. 2001. Intraspecific gene genealogies: trees grafting into networks. Trends in Ecology and Evolution, **16**: 37-45.

QIAGEN DNeasy tissue handbook. 2004. Mississauga, ON. pp. 18-20.

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

Raymond, M. and F. Rousset. 1995. An exact test for population differentiation. Evolution, **49:** 1280-1283.

Robison, H.W. 1986. Zoogeographic implications of the Mississippi River basin. The zoogeography of North American freshwater fishes, edited by Hocutt, C.H and E.O. Wiley. John Wiley & Sons, Inc. New York, NY. p. 267-285.

Ronquist, F. and J.P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics, **19:** 1572-1574.

Rypel, A.L. 2007. Sexual dimorphism in growth of freshwater drum. Southeastern Naturalist, **6:** 333-342.

Schneider, S. and L. Excoffier. 1999. Estimation of demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: Application to human mitochondrial DNA. Genetics, **152:** 1079-1089.

Siegert, M.J. 2001. Ice sheets and late Quaternary environmental change. John Wiley and Sons, Ltd., New York, NY, p. 165-193.

Slatkin, M. and L. Voelm. 1991. F_{ST} in a hierarchal island model. Genetics, **127:** 627-629.

Starnes, W.C. and D.A. Etnier. 1986. Drainage evolution and fish biogeography of the Tennessee and Cumberland rivers drainage realm. The zoogeography of North American freshwater fishes, edited by Hocutt, C.H and E.O. Wiley. John Wiley & Sons, Inc. New York, NY. p. 325-361.

Stepien, C.A., D.J. Murphy, R.N. Lohner, O.J. Sepulveda-Villet and A.E. Haponski. 2009. Signatures of vicariance, postglacial dispersal and spawning philopatry: population genetics of the walleye *Sander vitreus*. Molecular Ecology, **18:** 3411-3428.

Stewart, K.W. and D.A. Watkinson. 2004. The freshwater fishes of Manitoba. University of Manitoba Press, Winnipeg, MB. p. 245-247.

Tajima, F. 1993. Measurement of DNA polymorphism. Mechanisms of modern evolution, edited by Takahata, N. and Clark, A.G. Japan Scientific Societies Press, Sinauer Associates, Inc., Sunderland, MA. p. 37-59.

Templeton, A.R. 1998. Nested clade analyses of phylogeographic data: testing hypotheses about gene flow and population history. Molecular Ecology, **7:** 381-397.

Templeton, A.R. 2004. Statistical phylogeography: methods of evaluating and minimizing inference errors. Molecular Ecology, **13**: 789-809.

Templeton, A.R. 2009. Statistical hypothesis testing in intraspecific phylogeography: nested clade phylogeographic analysis vs. approximate Bayesian computation. Molecular Ecology, **18**: 319-331.

Tomelleri, J.R. and Eberle, M.E. 1990. Fishes of the central United States. University Press of Kansas, Lawrence, Kansas. p. 207.

Trenhaile, A.S. 1998. Geomorphology: a Canadian perspective. Oxford University Press, New York, NY, p. 100-199.

Underhill, J.C. 1986. The fish fauna of the Laurentian great lakes, the St. Lawrence lowlands, Newfoundland and Labrador. The zoogeography of North American freshwater fishes, edited by Hocutt, C.H and E.O. Wiley. John Wiley & Sons, Inc. New York, NY. p. 105-136.

Villablanca, F. X. 1994. Spatial and temporal aspects of populations revealed by mitochondrial DNA. Ancient DNA. Edited by Hermann, B and Hummel, S. Springer-Verlag, New York, NY. pp. 31-58.

White, D.J., J.N. Wolff, M. Pierson and N.J. Gemmell. 2008. Revealing the hidden complexities of mtDNA inheritance. Molecular Ecology, **17:** 4925-4942.

Wilson, C.C. and P.D.N. Hebert. 1998. Phylogeography and post-glacial dispersal of lake trout (*Salvelinus namaycush*) in North America. Canadian Journal of Fisheries and Aquatic Sciences, **55:** 1010-1024.

Wilson, C.C., P.D.N. Hebert, J.D. Reist and J.B. Dempson. 1996. Phylogeography and postglacial dispersal of arctic charr *Salvelinus alpinus* in North America. Molecular Ecology, **5:** 187-197.

Zink, R.M. 2002. Methods in comparative phylogeography and their application to studying evolution in the North American aridlands. Integrative and Comparative Biology, **42**: 953-959.

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

Appendix 1. Sample locations. Numbers were arbitrarily assigned to each sample.

Sample locations	Number of specimens sampled	Numbers assigned to each sample
Lake Ontario	7	5, 6, 0630-0634
Lake Erie	20	17, 18, 0611-0620, 0670-0677
Green Bay	16	11, 12, 0626-0629, 0694-06103
Nelson River	15	15, 16, 0659-0663, 0686-0693
Lake Winnipeg	18	1, 2, 0653-0658, 06115-06124
Lake Sakakawea	17	23, 24, 0621-0625, 06145-06154
Lake Pepin	16	19, 20, 0635-0639, 06125-06133
Lake Maloney	16	7-9, 0647-0652, 0678-0684
Missouri River	20	21, 22, 0640-0646, 06104-06114
Ohio River	8	13, 14, 0664-0669
Pickwick Reservoir	19	3, 4, 0605-0610, 06134-06144

Appendix 2. Concentration of DNA eluent for freshwater drum samples. Concentration of DNA eluent after extraction in ng/µL. Each sample was measured five times using the Qubit fluorometer, the high and low values discarded (marked with an asterisk*), and the average concentration calculated. See Appendix 1 for sample location associated with each sample number.

Sample number	Conc. #1	Conc. #2	Conc. #3	Conc. #4	Conc.#5	Average
1	10.6*	10.9	11.1	11.7*	11.5	11.2
2	12.7*	13.1	13.3	14.8	15.3*	13.7
3	8.40*	8.79	8.76	8.88*	8.81	8.79
4	15.3*	15.6	15.9	16.3	16.4*	15.6
5	10.9*	11.2	11.2	11.7	11.9*	11.4
6	14.1*	14.4	14.6	14.8	16.0*	14.6
7	11.0*	11.5	11.4	11.7	11.8*	11.5
9	9.22*	9.40	9.47*	9.47	9.35	9.41
11	10.8*	11.2	11.9	12.4	12.5*	11.8
12	4.17	4.31*	4.30	4.12*	4.21	4.40
13	2.48	2.49	2.50*	2.31*	2.34	2.44
14	5.56	5.68*	5.66	5.55*	5.56	5.59
15	13.7*	14.0	14.7	14.7	15.0*	14.5
16	10.1*	10.7	10.7	10.9	11.4*	10.8
17	13.6*	13.9	15.2	16.1*	15.8	15.0
18	16.8	15.4*	15.6	16.4	16.8*	16.3
19	10.8*	11.2	11.3	11.5	12.8*	11.3

20	7.36	7.41*	7.40	7.31	7.10*	7.36
21	11.8*	12.3	12.4	12.4	12.5*	12.4
22	13.2*	13.7	13.9	14.5*	14.4	14
23	3.01	3.07	3.12*	2.99*	3.05	3.04
24	4.32*	4.27	4.27	4.20	4.11*	4.25
0605	0.415*	0.403	0.378	0.362	0.342*	0.382
0606	11.6	11.6*	11.2*	11.3	11.3	11.4
0607	5.11*	4.98	5.05	5.07	4.69*	3.53
0608	5.39	5.45*	5.45	5.17*	5.18	5.34
0609	10.0*	10.1	10.1	10.3	10.4*	10.2
0610	3.75	3.80*	3.77	3.67	3.62*	3.73
0611	5.88*	5.95	6.05	6.25*	6.25	6.08
0612	13.4	13.4*	13.5	13.9	14.0*	13.6
0613	5.54*	5.30	5.26	5.12	4.65*	5.23
0614	0.522	0.518	0.502*	0.504	0.526*	0.515
0615	12.1	11.9*	12.1	12.7*	12.7	12.3
0616	6.40	6.45	6.45*	6.30*	6.32	6.39
0617	4.07*	3.99	3.98	3.53	3.52*	3.83
0618	6.56*	6.67	6.81	7.03*	6.87	6.78
0619	15.6*	16.0	18.4	18.5*	18.4	17.6
0620	6.43*	6.64	6.82	6.97*	6.97	6.81
0621	6.10*	6.19	6.23	6.25	6.25*	6.22
0622	8.36*	8.37	8.83	8.88	9.04*	8.69
0623	7.82*	7.94	8.05	8.29	8.46*	8.09

0624	9.78*	9.94	10.1	10.7*	10.7	10.2
0625	8.01	6.77*	8.19	8.58*	7.73	7.98
0626	8.05*	8.06	8.23	8.48	8.57*	8.26
0627	8.35*	8.49	8.69	8.91	9.04*	8.70
0628	3.29*	3.33	3.33	3.36*	3.35	3.34
0629	5.19*	5.23	5.31	5.44*	5.42	5.32
0630	4.74*	4.81	4.87*	4.84	4.84	4.83
0631	1.20*	1.21	1.22*	1.21	1.20	1.21
0632	6.99*	7.05	7.11	7.23	7.29*	7.13
0633	2.90*	2.93	2.95*	2.94	2.94	2.94
0634	8.30*	8.34	8.52	8.63	8.68*	8.50
0635	4.79	4.82	4.82	4.87*	4.76*	4.81
0636	5.91*	6.09	6.15	6.16*	6.11	6.12
0637	2.65*	2.63	2.60	2.47	2.39*	2.57
0638	3.68	3.66	3.68	3.74*	3.63*	3.67
0639	2.51	2.53*	2.49	2.46	2.44*	2.49
0640	0.621*	0.613	0.597*	0.599	0.597	0.603
0641	0.987	1.00*	0.981	0.943	0.940*	0.970
0642	1.28	1.28*	1.31	1.34	1.35*	1.31
0643	0.322*	0.319	0.314	0.302*	0.310	0.314
0644	0.644*	0.655	0.659	0.675*	0.662	0.659
0645	0.508	0.526*	0.512	0.511	0.462*	0.510
0646	0.708	0.741*	0.709	0.718	0.698*	0.712
0647	1.33	1.28*	1.31	1.35	1.35*	1.33

0648	1.80	1.81	1.79	1.81*	1.76*	1.80
0649	3.00	3.03	3.03*	2.99	2.93*	3.01
0650	9.83*	9.94	9.95	10.1	10.1*	10.0
0651	4.45*	4.49	4.55*	4.55	4.48	4.51
0652	14.4*	15.1	15.2*	15.1	15.0	15.1
0653	4.87*	4.92	5.01	5.02*	4.99	4.97
0654	6.95*	7.06	7.15	7.23	7.32*	7.15
0655	6.15*	6.45	6.77	6.88	6.92*	6.70
0656	3.88*	3.92	3.91	3.94*	3.93	3.92
0657	6.53*	6.70	6.88	6.94*	6.89	6.89
0658	5.78*	5.87	5.99*	5.94	5.95	5.92
0659	11.7*	11.9	12.1	12.4*	12.3	12.1
0660	10.5*	11.1	11.2	11.6*	11.6	11.3
0661	3.68	3.68*	3.64	3.65	3.58*	3.66
0662	8.58*	8.65	8.84	8.99*	8.89	8.79
0663	4.16*	4.06*	4.11	4.06	4.08	4.08
0664	2.48*	2.36	2.32	2.28	2.25*	2.32
0665	7.88*	7.91	7.96*	7.89	7.90	7.90
0666	7.03	7.05	7.09*	6.97*	7.04	7.04
0667	24.0	24.0	24.0*	24.0*	24.0	24.0
0668	2.67*	2.78	2.78	2.76	2.81*	2.77
0669	4.13	4.15	4.08*	4.27*	4.23	4.17
0670	14.7*	15.1	15.0	15.4	15.8*	15.2
0671	15.5*	15.6	15.9	16.4	16.9*	16.0

0672	10.4	10.4*	10.5	11.1*	10.7	10.5
0673	11.8*	12.0	11.9	12.2	12.7*	12.0
0674	20.0*	22.0	22.0	23.0*	23.0	22.3
0675	0.117*	22.0	19.0*	24.0	24.0	23.3
0676	21.0	20.0*	21.0*	21.0	21.0	21.0
0677	17.3*	18.2	19.2	19.2	20.0*	18.9
0678	3.97*	3.84	3.75*	3.77	3.80	3.80
0679	21.0*	21.0*	21.0	21.0	21.0	21.0
0680	17.9	17.9*	17.7	17.4*	17.7	17.8
0681	4.59*	4.49	4.47	4.37*	4.46	4.47
0682	7.22*	7.14	7.08	6.91*	7.03	7.08
0683	9.46*	9.29	9.04*	9.10	9.12	9.17
0684	13.7*	13.6	13.5	13.3	13.3*	13.5
0685	8.48*	8.39	8.45	8.42	8.39*	8.42
0686	7.71*	7.63	7.39	7.39	7.39*	7.47
0687	21.0	21.0*	21.0*	21.0	21.0	21.0
0688	8.45	8.45*	8.41	8.40*	8.44	8.43
0689	6.77*	6.66	6.64	6.61	6.55*	6.64
0690	4.06*	3.98	3.95	3.92	3.89*	3.95
0691	8.57	8.57*	8.43	8.46	8.21*	8.49
0692	5.31	5.31	5.28*	5.29	5.34*	5.30
0693	14.8*	14.3	14.8*	14.5	14.6	14.5
0694	13.7	13.3*	15.1	15.7	15.9*	14.8
0695	11.1*	11.4	11.5	11.7	12.2*	11.5

0696	12.6	12.2*	12.9	14.0	14.7*	13.2
0697	11.0	10.9*	12.1	12.1	12.2*	11.7
0698	18.0*	18.4	20.0	21.0*	20.0	19.5
0699	12.9*	13.6	14.2	14.6	15.1*	14.1
06100	10.3*	11.9	13.4	14.6	15.4*	13.3
06101	8.15	8.29	8.14*	8.37*	8.25	8.23
06102	5.68*	5.86	6.06	6.32*	6.27	6.06
06103	12.3*	13.6	13.6	14.2	14.3*	13.8
06104	16.4*	17.0	19.8*	19.7	19.1	18.6
06105	15.3*	18.0	19.3	21.0	22.0*	19.4
06106	18.2*	18.5	21.0	22.0	24.0*	17.2
06107	10.1*	10.2	10.8	10.9	11.3*	10.6
06108	2.36*	2.41	2.45	2.48	2.56*	2.45
06109	19.7*	19.9	21.0	21.0*	21.0	20.6
06110	12.7*	13.5	15.2	15.5	16.0*	14.7
06111	17.4*	18.4	19.3	20.0	21.0*	19.2
06112	11.0	11.0*	11.5	11.7	11.8*	11.4
06113	11.5*	11.8	12.4	12.6	12.9*	12.3
06114	10.9*	11.0	11.4	11.2	11.5*	11.2
06115	12.8*	13.0	13.2	13.2	13.6*	13.1
06116	3.08*	3.16	3.27	3.25	3.36*	3.23
06117	8.24*	8.30	8.47	8.49	8.59*	8.42
06118	23.0	22.0*	23.0	23.0*	23.0	23.0
06119	22.0	22.0*	23.0	23.0*	23.0	22.7

06120	16.7*	16.9	17.5*	17.4	17.3	17.2
06121	0.486*	0.448	0.484	0.413*	0.448	0.460
06122	3.04*	3.09	3.20	3.33*	3.26	3.18
06123	1.91*	1.86*	1.89	1.88	1.87	1.88
06124	2.62	2.77*	2.76	2.60*	2.70	2.69
06125	10.4	10.3*	10.6	11.4*	11.1	10.7
06126	9.14*	9.00	8.99*	9.01	9.02	9.01
06127	10.3*	10.8	11.6	12.0	12.1*	11.5
06128	3.59*	3.56	3.55	3.49	3.44*	3.53
06129	3.10*	3.21	3.25	3.35*	3.30	3.25
06130	10.5*	10.6	10.8	11.2*	11.2	10.9
06131	8.45*	8.74	8.81	9.19*	9.07	8.87
06132	8.43*	8.49	8.46	8.93	9.23*	8.63
06133	9.00	8.95*	9.20	9.33*	9.19	9.13
06134	13.2*	14.0	14.2*	13.9	14.1	14.0
06135	7.53*	7.37	7.27	7.22*	7.39	7.34
06136	8.95*	9.17	9.03	9.32	9.36*	9.17
06137	10.7	10.8	10.6*	11.3	11.5*	10.9
06138	7.28	7.22*	7.41	7.44	7.43*	7.38
06139	10.0	10.2*	9.85*	9.93	9.90	9.94
06140	6.51*	6.44	6.50	6.29*	6.43	6.46
06141	13.0*	13.0	13.3	13.1	14.2*	13.1
06142	9.97	9.90*	9.93	9.98*	9.91	9.94
06143	3.20*	3.17	3.08*	3.11	3.12	3.13

06144	6.03	6.01*	6.34*	6.17	6.32	6.17
06145	1.33*	1.35*	1.34	1.33	1.33	1.33
06146	2.98	2.97*	3.08	3.05	3.17*	3.04
06147	1.72*	1.74	1.75	1.74	1.78*	1.74
06148	3.55*	3.69	3.88	3.98	4.05*	3.85
06149	1.27	1.27	1.27	1.27*	1.25*	1.27
06150	2.88*	2.91	2.90	2.90	2.94*	2.90
06151	4.09	4.11	4.05*	4.14	4.18*	4.11
06152	3.34*	3.45	3.50	3.72	3.90*	3.56
06153	3.70*	4.09	4.21	4.24	4.56*	4.18
06154	3.18*	3.19	3.39	3.44	3.47*	3.34

Appendix 3. Design of primers specific to the mitochondrial control region of freshwater drum.

Primer design was dependent on whether the genetic relationship between freshwater drum and red drum (*Sciaenops ocellatus*) was sufficient to amplify the freshwater drum mitochondrial control region using red drum primers. Certain principles had to be applied to ensure specific and efficient amplification.

The following principles were used to design suitable primers. The GC (guanine/cytosine) content must be between 40% and 60% to ensure adequate binding. This is due to the fact that three hydrogen bonds exist between guanine and cytosine as opposed to two hydrogen bonds between adenine and thymine. Guanine or cytosine at the 3' end ensure specific binding and increase reaction efficiency. The ideal primer length is 18 to 26 base pairs in length, length ensures specificity, but should not be so large as to decrease efficiency. The melting temperature (T_m) should be similar between both primers, ranging from approximately 55°C to 60°C as annealing temperature is affected by melting temperature. Annealing temperature which is too high will result in insufficient primer-template binding and too low of an annealing temperature results in non-specific product. Secondary structures or primer-dimer interactions should be minimal or absent as this decreases primer-template binding (Dieffenbach et al., 1993; Abd-Elsalam, 2003).

Following these guidelines, suitable primers were designed and tested using the Operon OLIGO tool to identify any primer interactions. Primers were also checked for specificity using NCBI Basic Local Alignment Search Tool (BLAST). This confirmed that the primers were specific to members of the family Sciaenidae. Amplify v1.2 (Engels, 1993) was used to run a virtual PCR to

predict the outcome of amplification and identify the probability of a match and stability of primer-template binding.

Prior to my study, primers specific to the mitochondrial control region of freshwater drum had not been developed. *Sciaenops ocellatus* (red drum) is distributed throughout estuaries of the Gulf of Mexico and southeast Atlantic coast of North America (Gold et al. 1993). The genetic relationship between these two members of the Sciaenid family should be sufficient to amplify mitochondrial control region sequences of freshwater drum with red drum primers and develop freshwater drum specific primers.

Appendix 4. Control region haplotype frequencies of freshwater drum in sampled populations. Frequency of haplotypes in sampled populations of freshwater drum. Column 1 refers to the name arbitrarily assigned to a particular haplotype. All other columns refer to the frequency of each haplotype in a population.

Sample site / Haplotype	Lake Ontario (n=7)	Lake Erie (n=20)	Green Bay (n=16)	Nelson River (n=15)	Lake Winnipeg (n=18)	Lake Sakakawea (n=17)	Lake Pepin (n=16)	Lake Maloney (n=16)	Missouri River (n=20)	Ohio River (n=8)	Pickwick Reservoir (n=19)
h0005	0.429	0.8	0.188	0	0	0	0.0714	0	0.125	0.143	0
h0612	0.143	0.05	0	0	0	0	0	0	0	0	0
h0631	0.143	0	0	0	0	0	0	0	0	0	0
h0633	0.286	0.05	0	0	0	0	0	0	0	0	0
h0001	0	0	0	0.8	0.625	0	0.357	0.0625	0.188	0	0.176
h0007	0	0	0	0	0	0	0	0.5	0	0	0
h0638	0	0	0	0	0.0625	0	0.0714	0.0625	0	0	0.235
h0647	0	0	0	0	0	0	0	0.375	0	0	0
h0011	0	0	0.0625	0.0667	0.188	0.357	0	0	0	0	0.0588
h0655	0	0	0	0.133	0.0625	0	0	0	0	0	0

h0012	0	0.1	0.562	0	0	0	0	0	0	0	0
h0654	0	0	0	0	0.0625	0	0	0	0.125	0	0
h6134	0	0	0	0	0	0	0	0	0	0	0.0588
h6137	0	0	0	0	0	0.214	0.0714	0	0	0	0.176
h6133	0	0	0	0	0	0	0.0714	0	0	0	0.0588
h0609	0	0	0	0	0	0	0	0	0	0	0.0588
h0607	0	0	0	0	0	0	0	0	0	0.286	0.0588
h0003	0	0	0	0	0	0	0.143	0	0	0	0.118
h0627	0	0	0.0625	0	0	0	0	0	0	0	0
h0696	0	0	0.0625	0	0	0	0	0	0	0	0
h6100	0	0	0.0625	0	0	0	0	0	0	0	0
h0668	0	0	0	0	0	0.0714	0	0	0	0.143	0
h0665	0	0	0	0	0	0	0	0	0	0.143	0
h0014	0	0	0	0	0	0.0714	0	0	0	0.143	0
h0013	0	0	0	0	0	0	0	0	0	0.143	0
h0636	0	0	0	0	0	0	0.0714	0	0	0	0
h0639	0	0	0	0	0	0	0.0714	0	0	0	0
h6130	0	0	0	0	0	0	0.0714	0	0	0	0
h0021	0	0	0	0	0	0	0	0	0.0625	0	0

h0022	0	0	0	0	0	0	0	0	0.0625	0	0
h0641	0	0	0	0	0	0	0	0	0.0625	0	0
h0642	0	0	0	0	0	0	0	0	0.0625	0	0
h0644	0	0	0	0	0	0	0	0	0.125	0	0
h6105	0	0	0	0	0	0	0	0	0.0625	0	0
h6108	0	0	0	0	0	0	0	0	0.0625	0	0
h6104	0	0	0	0	0	0	0	0	0.0625	0	0
h6149	0	0	0	0	0	0.0714	0	0	0	0	0
h0621	0	0	0	0	0	0.0714	0	0	0	0	0
h0024	0	0	0	0	0	0.143	0	0	0	0	0

Appendix 5. ND2 region haplotype frequencies of freshwater drum in sampled populations. Frequency of haplotypes in sampled populations of freshwater drum. Column 1 refers to the name arbitrarily assigned to a particular haplotype. All other columns refer to the frequency of each haplotype in a population.

Sample site / Haplotype	Lake Ontario (n=7)	Lake Erie (n=20)	Green Bay (n=16)	Nelson River (n=15)	Lake Winnipeg (n=18)	Lake Sakakawea (n=17)	Lake Pepin (n=16)	Lake Maloney (n=16)	Missouri River (n=20)	Ohio River (n=8)	Pickwick Reservoir (n=19)
h1	0.857	0.8	0.267	0.867	0.944	0.769	0.867	0.125	0.5	0.5	0.882
h12	0.143	0.05	0.6	0	0.0556	0	0	0.312	0.0556	0	0
h82	0	0	0	0	0	0	0	0.0625	0.0556	0	0
h7	0	0	0	0	0	0	0	0.5	0	0	0.0588
h15	0	0.15	0	0.133	0	0.0769	0.0667	0	0.0556	0.5	0.0588
h26	0	0	0.133	0	0	0	0	0	0.0556	0	0
h131	0	0	0	0	0	0	0.0667	0	0	0	0
h110	0	0	0	0	0	0.154	0	0	0.167	0	0
h41	0	0	0	0	0	0	0	0	0.0556	0	0
h46	0	0	0	0	0	0	0	0	0.0556	0	0