Mitochondrial DNA and microsatellite analysis of post-glacial dispersal and hybridization of longnose suckers (*Catostomus catostomus*) in Labrador: implications for conservation and management near the Trans-Labrador Highway

by

©Barbara Langille

A thesis submitted to the
School of Graduate Studies
in partial fulfilment of the
requirements for the degree of
Master of Science

Department of Biology, Faculty of Science
Memorial University of Newfoundland

October 2014
St. John's, Newfoundland and Labrador

ABSTRACT

Understanding phylogeographic origins and hybridization of species are crucial when proposing effective management strategies in disturbed habitats. The longnose sucker in Labrador is an interesting species as the eastern part of the range in North America has not been studied and Labrador has recently undergone the construction of the Trans-Labrador Highway (TLH) with no assessment of the influence of this disturbance on the longnose sucker. To document genetic diversity, post-glacial dispersal, and putative hybridization with the white sucker, mitochondrial DNA and microsatellites were analysed, leading to evidence for one main glacial refugial source of longnose suckers within Labrador, the Atlantic refugium, with possible minor contributions from the Beringian and Mississippian. Significant population structure among regions and among lakes was inferred. Only four putative hybrid suckers were identified suggesting that hybridization is lower than among other catostomid species. Hence the TLH is likely not influencing hybridization of longnose and white suckers in Labrador.

ACKNOWLEDGEMENTS

I first would like to thank my supervisor Dr. H. Dawn Marshall for all her support and guidance throughout this entire project. Her patient and understanding nature has made it an absolute pleasure to be a part of her lab. Without all her constant hours of reading, proof-reading, editing, data analysis, endless emails and office visits, this thesis would never have been completed.

Secondly I would like to thank Robert Perry and Donald Keefe of the Wildlife Division (Newfoundland and Labrador Department of Environment and Conservation) for acquiring samples from Oliver Barker from Wildlife in the Yukon, from Jessica Courtier from British Columbia as well as collecting samples from Labrador. I would also like to express my gratitude for being offered the chance to go to Labrador to collect samples with Robert Perry, Donald Keefe, and Tim Andrews of the Forestry Division. The trip was an unforgettable experience. I received invaluable insight and gained a newfound appreciation for field work and the difficulties as well as the rewards. This project would not have been possible without financial support from the Newfoundland and Labrador Department of Environment and Conservation.

I would also like to thank Dr. Beth Perry of the Genomics and Proteomics (GaP) facility (CREAIT Network, Memorial University of Newfoundland) for all of her help with microsatellite setup as well as troubleshooting and easing the transition into looking at this new data set. As a committee member, I am also grateful for the time she spent reading and revising my thesis to make it as good as it could be. Her help and patience as

well as constant support and encouragement have been much appreciated and will not be forgotten.

Also to Brettney Pilgrim and Lidan Tao of the GaP Lab, thank you for all your help with using the sequencing instruments and data collection from the instruments. My constant hours of work in the GaP lab went smoothly due to all the help available, and your pleasant nature always made me feel comfortable. All my questions that were asked and subsequently answered and all the troubleshooting help was much appreciated.

I would like to thank my other committee member, Dr. David Innes for taking the time to read over and give such thoughtful revisions so as to make my thesis better. I am truly appreciative of the support.

I would also like to acknowledge the various lab mates I have had over the years who have encouraged me as well as helped me with my lab work. Notably, Sarah Predham, Braydi Rice, Kim Carew and Ainsley Latour have all played a role in getting me to the end of my thesis in one piece. Last but certainly not least, I am eternally grateful to my friends and family who have supported me and encouraged me to do the best I could. Their constant love and patience got me through all stages of this degree and I know it will continue throughout my future endeavours.

TABLE OF CONTENTS

Abstract……..ii

Acknowledgements……iii

List of Tables……x

List of Figures……...xiv

Abbreviations and Symbols…….……xvii

Chapter 1: General Introduction……...1

Phylogeography……...…1

Conservation genetics……..3

Mitochondrial DNA and microsatellite loci…….…...5

Freshwater fish in North America……..….7

Longnose Sucker post-glacial dispersal……..……8

Longnose sucker genetics……..….11

Global importance of longnose suckers and local concerns……12

Goals of thesis……...14

References……..15

Chapter 2: Post-glacial dispersal and genetic structure of Labrador populations of longnose suckers assessed with mitochondrial DNA: the evolutionary genetics of a poorly-understood species……..24

Abstract……..24

Introduction……26

Materials and Methods……...29

Study area and sample collection……...29

DNA isolation, polymerase chain reaction (PCR), and DNA sequencing…….…….29

Sequence analysis…….....33

Genetic diversity and tests of selection……..33

Population structure……...34

Population expansion..……35

Phylogeny construction and network analysis of haplotypes…….35

Results……36

Sequence analysis……...36

Genetic diversity and tests of selection……..38

Population structure……..42

Population expansion..……46

Phylogenetic analysis…….50

Discussion……...……...53

Mitochondrial diversity in the longnose sucker…….55

Strong regional structure with evidence of longitudinal variation of longnose suckers in Canada……..57

Evidence for post-glacial recolonization of Labrador from one major and two minor glacial refugia……..…58

Putative hybridization between longnose suckers and white suckers in southeastern Labrador……..60

Conclusions and future directions……62

References…….63

Chapter 3: Hybridization, gene flow, and post-glacial recolonization of longnose suckers in Labrador inferred with microsatellite markers……...69

Abstract……..69

Introduction……71

Materials and Methods……...75

Study area and sample collection……...75

Microsatellite analysis……...77

Microsatellite alleles and per-locus measures……...79

Hybridization analysis……...80

Population genetic analysis and structure……81

Results……82

Loci and alleles……..82

Bayesian assessment of hybrid individuals……84

Population and regional diversity and structure……...88

Discussion……....95

Bayesian admixture analysis reveal hybridization between longnose and white suckers in southeastern Labrador……96

Regional genetic diversity and structure among longnose sucker populations in Labrador……98

Refugial origins of Labrador suckers inferred from microsatellite structure……99

Major findings and conclusions……..…..101

References……102

Chapter 4: General Conclusions……...109

Refugial origins and genetic structure of longnose suckers in Labrador……...109

Hybridization and population structure ……..…….……...111

The impact of the Trans-Labrador Highway on longnose sucker populations……..114

Comparison of marker types and utility of using both microsatellites and mtDNA…..…….**.....**115

Conclusions……..116

References……118

Appendix: Longnose sucker primer publication……..121

LIST OF TABLES

Table 1.1. Timeline of the Pleistocene glacial and interglacial cycles……10

Table 2.1. Haplotypes determined among 362 putative longnose sucker individuals, with haplotype frequency (N) and region of location for each……..37

Table 2.2. Variable sites found in a 1244 bp combined fragment of cytochrome b, NADH dehydrogenase 4, and cytochrome c oxidase 1 in 28 haplotypes identified among 358 longnose suckers. The nucleotide substitutions are numbered according to position in codon as indicated by the top row. N indicates an unknown base and haplotype 9 has been omitted from this table.……...……39

Table 2.3. Number of individuals in each location used in all population comparisons with corresponding measures of genetic diversity and numbers of haplotypes per location - sample size (*N*), number of haplotypes (*k*), haplotype diversity (*h*), nucleotide diversity (π), and standard deviation (SD), locality abbreviations are: MER Mercier Lake, WAL Walkabout Lake, BIR Birchy Lake, JOE Lac Joe, LOB Lobstick Lake, STR Strange Lake, KON Konrad Lake, IKA Ika Lake, RIG The Right Lake, BRE Brettney Lake, MIS Mistake Lake, CAB Cabot Lake, CRO Crooked Lake, GEN Genetics H Lake, BCA British Columbia Grizzly Lake, BCB British Columbia Teardrop Lake, YUK Yukon.……41

Table 2.4. SAMOVA analysis of mtDNA sequences from 17 populations located in Labrador, Yukon, and British Columbia. K indicates the predefined number of groups into which all populations are separated. F_{CT} is the differentiation among regions, F_{SC} is the differentiation among populations within groups, and F_{ST} is the differentiation among populations. Location abbreviations are defined in Table 2.3……..……..…….44

Table 2.5. Hierarchical analysis of SAMOVA for 17 populations of longnose suckers performed using K = 3 as the number of groups. The F-statistic for among groups is represented by F_{CT}, among populations within groups is represented by F_{SC}, and within populations is represented by F_{ST}……46

Table 2.6. Measures of pairwise population differentiation using F_{ST} values among all Labrador longnose sucker populations. Values followed by * are significant at α = 0.05 after sequential Bonferroni correction for multiple comparisons……..…..……..47

Table 2.7. Measures of pairwise population differentiation using F_{ST} values among all regions in Labrador. Values followed by * are significant at α = 0.05 after sequential Bonferroni correction for multiple comparisons….. ….48

Table 2.8. Neutrality tests and mismatch expansion estimates for longnose suckers in each of four regions in Labrador excluding haplotype 9; SDD is the sum of squared deviations in the mismatch distribution……49

Table 3.1. Primer sequences and PCR annealing temperatures for six microsatellite loci, originally developed for *Xyrauchen texanus* (razorback sucker), *Chasmistes liorus* (June sucker), *Catostomus ardens* (Utah sucker) and *Deltistes luxatus* (lost river sucker). Abbreviations: F, forward primer; R, reverse primer; T_A, annealing temperature……78

Table 3.2. Measures of diversity of six microsatellite loci across 302 longnose suckers (*Catotstomus catostomus*) and 14 white suckers (*C. commersonii*), including number of observed alleles (N_A), and allele size range……..83

Table 3.3. Assortment of the 302 longnose suckers into categories of parental longnose sucker (LS Parental), F_2 hybrid, or Mislabelled, l as determined using NewHybrids, based on six microsatellite loci. *N* is the sample size; LS Parentals are fish identified as longnose suckers with longnose sucker microsatellite genotypes; Mislabelled are fish identified as longnose suckers phenotypically but with white sucker microsatellite and mtDNA……..87

Table 3.4. Sample size (*N*), observed (H_O) and expected (H_E) heterozygosity, and allelic richness (*A*) of longnose suckers (total of 299 individuals without the three individuals with white sucker microsatellites) across six microsatellite loci for 14 lakes in four regions of Labrador. Abbreviations for each lake are given in brackets. Regional numbers are shown in bold type……89

Table 3.5. Overall measures of F_{IS} among longnose suckers in each of 14 lakes among four regions in Labrador. There is no significant heterozygote excess/deficiency at $\alpha < 0.05$ after sequential Bonferroni correction …..90

Table 3.6. Pairwise measures of F_{ST} among populations (lakes) of longnose suckers, inferred from six microsatellite loci using allelic differences. Significance level at $\alpha = 0.05$ after sequential Bonferroni correction for multiple comparisons denoted by non-bolded values…….…….91

LIST OF FIGURES

Figure 2.1. Geographic locations of the 14 longnose sucker populations investigated in Labrador with numbers 1 through 4 indicating each defined region (1, Northern Labrador; 2, Western Labrador; 3, Central Labrador; and 4, Southeastern Labrador). Exact coordinates for locations Ikadlivik Brook Watershed in Northern Labrador and Birchy Lake in Southeastern Labrador are not known and are shown on the map with initials m and n respectively; locality abbreviations are: a Genetics H Lake, b Strange Lake, c Mistake Lake, d Walkabout Lake, e Konrad Lake, f Cabot Lake, g Lobstick Lake, h Lac Joe, i The Right Lake, j Mercier Lake, k Brettney Lake, l Crooked Lake ……...30

Figure 2.2. Geographic locations of the three populations of longnose suckers studied in Canada, Lake Laberge in Yukon (YK), Grizzly Lake (BCA) and Teardrop Lake (BCB) in British Columbia……..31

Figure 2.3. A plot of the F_{CT} parameter (differentiation between groups) for different values of *K*, the number of population groups, generated using SAMOVA. Analysis was performed using all 17 longnose sucker populations from Labrador, Yukon, and British Columbia……...43

Figure 2.4. Maximum likelihood phylogeny of phylogenetic relationships among longnose sucker mtDNA haplotypes. The HKY model of DNA

substitution was used and bootstrap values greater than 50% of 1000 replicates are shown at nodes……51

Figure 2.5. Phylogenetic relationships among longnose suckers mtDNA haplotypes inferred using a Bayesian analysis. Support is shown only for nodes with > 50% posterior probability. ……..……..52

Figure 2.6. Median spanning network depicting the phylogenetic relationships between different haplotypes, drawn using the program Network. The size of each circle represents the haplotype frequency and length of each branch indicates mutational steps. Different colors represent different localities and regions: green British Columbia, blue Yukon, purple Northern, orange Central, brown Western, yellow Southern Labrador…….……....54

Figure 3.1. Geographic locations of the 14 longnose sucker populations investigated in Labrador with numbers 1 through 4 indicating each defined region (1, Northern Labrador; 2, Western Labrador; 3, Central Labrador; and 4, Southeastern Labrador). Exact coordinates for locations Ikadlivik Brook Watershed in Northern Labrador and Birchy Lake in Southeastern Labrador are not known and are shown on the map with initials m and n respectively; locality abbreviations are: a Genetics H Lake, b Strange Lake, c Mistake Lake, d Walkabout Lake, e Konrad Lake, f Cabot Lake, g Lobstick Lake, h Lac Joe, i The Right Lake, j Mercier Lake, k Brettney Lake, l Crooked Lake …….……...76

Figure 3.2. A plot of the estimated probability of *K* (LnP(D)) from the STRUCTURE analysis, for *K*=1-20……...85

Figure 3.3. STRUCTURE plot of hybridization between longnose sucker (red, green, yellow) and white sucker (represented in blue) run at *K*=4. Each individual is represented as a vertical bar, whose length is proportional to the membership into each K cluster. Hybrid individuals are indicated by a black dot below the bar in the cut-out and potentially misidentified individuals are indicated by a black dot above the bar in the cut-out. Region labels are: NL, Northern Labrador; WL, Western Labrador; CL, Central Labrador; SE, Southeastern Labrador.WS refers to white sucker ……..86

Figure 3.4. Hierarchical STRUCTURE bar plot of admixture proportion (Q) of each individual where *K*=4 contains all Labrador regions and white sucker, and *K*=3 contains all Labrador regions without white sucker, both using an ancestry model that assumes admixture where each vertical line represents an individual fish. The colours correspond to each of the genetic clusters (K), with the proportion of colour in each individual corresponds to its proportion of that genetic cluster. Region labels are: NL, Northern Labrador; WL, Western Labrador; CL, Central Labrador; SE, Southeastern Labrador.WS refers to white sucker……..93

Figure 3.5 A plot of the estimated probability of K (LnP(D)) from the STRUCTURE analysis, for *K*=1-20 without white sucker individuals……94

LIST OF ABBREVIATIONS AND SYMBOLS

A or *A*	Allelic richness
B_0	Backcross with parental species 1
B_1	Backcross with parental species 2
BC	British Columbia
BCA	Grizzly Lake
BCB	Teardrop Lake
BIR	Birchy Lake
bp	base pairs
BRE	Brettney Lake
°C	Degrees celsius
CAB	Cabot Lake
CL	Central Labrador
CO1	Cytochrome c oxidase one
CRO	Crooked Lake
CYT b	Cytochrome b
DNA	Deoxyribonucleic acid

dNTP	Deoxyribonucleotide triphosphate
F	forward primer
F_1	First-generation hybrid
F_2	Second-generation hybrid
F_{IS} or F_{IS}	Inbreeding coefficient
F_{ST} or F_{ST}	Population differentiation due to genetic structure
GaP	Genomics and Proteomics
GEN	Genetics H Lake
h or *h*	Haplotype diversity
HKY	Hasegawa-Kishino-Yano
H_E or H_E	Expected heterozygosity
H_O or H_O	Observed heterozygosity
IKA	Ikadlivik Lake
JOE	Lac Joe
k or *k*	Number of haplotypes
K or *K*	The number of genetic clusters as inferred by STRUCTURE
KON	Konrad Lake

LD	Linkage disequilibrium
LGM	Last glacial maximum
ln P(D)	log likelihood of the data
LNS	Longnose sucker
LOB	Lobstick Lake
MER	Mercier Lake
MIS	Mistake Lake
mm	Millimetre
mtDNA	Mitochondrial Deoxyribonucleic acid
$_n$	Number of repeats
N_A or N_A	Number of alleles
N or *N*	Sample size
ND4	Nicotinamide adenine dinucleotide dehydrogenase subunit four
NL	Northern Labrador
PCR	Polymerase chain reaction
Q or *Q*	Admixture proportion
R	Reverse primer

RIG	The Right Lake
s	Seconds
SAMOVA	Spatial Analysis of Molecular Variance
SE	Southeastern Labrador
SSRs	Simple sequence repeats
STR	Strange Lake
T_a	Annealing temperature
Taq	*Thermus aquaticus*
WAL	Walkabout Lake
WL	Western Labrador
WS	White sucker
YK	Yukon
µL	Microlitre
π	Nucleotide diversity

CHAPTER 1: GENERAL INTRODUCTION

The protection and preservation of species is an important aspect of biological conservation. To fully understand the best methods for protection and conservation of species, a multitude of biological questions must be addressed. These questions include but are not limited to topics of habitat use and destruction, population sizes, trophic interactions and diet restrictions, life histories, introgression, genetic lineages and taxonomy, as well as genetic population structure and genetic diversity. The focus of the present thesis research is on genetic aspects of these biological questions with emphasis on historical inferences and phylogeography as well as level of introgression with closely-related species in a poorly known freshwater fish species, the longnose sucker (*Catostomus catostomus*).

Phylogeography

Phylogeography is the study of the geographical distribution of genetic lineages (Avise 2000) or more specifically the understanding of current geographic distributions of individuals and genetic lineages based on historical processes and past geographic distributions (Damien 2010). The demographic and evolutionary history of a species can be inferred by assessing how genetic diversity is distributed throughout the range (Avise *et al*. 2004; Slatkin *et al*. 1987). This is because gene lineages coalesce or merge together through time and space to a common ancestor in such a way as to reflect the biogeographic history of a particular species or population (Avise 2000).

Phylogeography involves an understanding of glacial events, among other biogeographic events, as these events have a direct effect on present-day distributions of species. During glaciation periods, species were confined to refugia or, in the case of freshwater fish, pockets of water free from ice cover, where they were able to survive while environmental conditions elsewhere were unsuitable (Liu *et al*. 2012). Subsequent deglaciation produced temporary spillways allowing fish a way out for eventual colonization of surrounding areas and beyond (Rempel & Smith 1998). The genetic variation of each individual in conjunction with population structure hold the answer to historical questions which can aid in the evaluation of current populations or groups leading to the future assessment and subsequent management of species.

What is a species? The most rudimentary question asked by the scientific community was given new life by phylogeographic approaches and concepts (Hickerson *et al*. 2010). Current species definitions have evolved to include some criteria of a phylogeographic nature, including species delimitation which identifies how a species is, in terms of boundaries and population size, and not just what (De Queiroz 2007). For the purpose of this study, the phylogenetic species concept is the concept I will be using to define separate species as it allows for the possibility of mating between separate species and maintains less restrictive parameters than the biological species concept (Baum 1992). As we understand more about each species, factors working to confuse one species with another are found more often, including species introductions and hybridization. These factors can cause difficulty when attempting to place each species into an existing lineage as they represent a distortion between two or more distinct species. From a

conservation standpoint, phylogeography can help identify and target distinct taxonomic groups created by one or several historical or contemporary events which may merit eventual assessment of conservation and subsequent management (Crandall *et al*. 2000).

The introduction of concept and methods of phylogeography in 1987 by John Avise revolutionized evolutionary genetics, causing an explosion in the field (Avise 2009). Phylogeographic methods offer the possibility to answer many different types of questions in areas ranging from conservation studies to evolutionary hotspots, to introgression and hybridization (Anamthawat-Jónsson 2012). Phylogeography lies at the center of several macro- and micro-evolutionary disciplines such as species conservation, ecology, and evolutionary biology and connects ideas of all sides into a unifying concept (Hewitt 2004).

Conservation genetics

Like phylogeography, conservation biology encompasses a variety of different subjects from island biogeography to environmental monitoring and population biology. Conservation biology is the study of biodiversity, and unlike phylogeography it is specifically biodiversity that has been negatively impacted by humans, and the measures taken to restore or protect different species and their habitats (Soulé 1985). Human development has caused many environmental issues such as habitat loss and degradation which negatively affects worldwide biodiversity by causing numerous species to become extinct at an alarming rate (Frankham 2003). One method to attempt to understand

biodiversity is through conservation genetics, which uses genetic methods to assess diversity within and between populations as well as mating patterns, kinship, gene flow, and hybridization (Avise 2004), and differs from phylogeography as the intent of conservation genetics is to assess the conservation status of any given population. Genetic variability within and among populations is essential for the persistence of most species (Frankham 2005).

An obvious yet important and sometimes forgotten concept is that biodiversity is intertwined with genetic diversity, and the goal of conservation genetics is to preserve genetic diversity which can be accomplished by preserving biodiversity and vice versa (National Science Foundation, 2007). Genetic diversity is maintained and even increased by evolutionary processes such as gene flow, natural selection, speciation, and even in some cases by hybridization among related species. In contrast, genetic diversity can be decreased by factors such as genetic drift, isolation, low reproductive rates, bottleneck events, and extensive hybridization (Avise 2004; Frasier *et al*. 2013).

A highly debated topic within conservation genetics is choosing which populations to consider for conservation research. In general, two main population types exist: central and marginal. Central populations found in the middle of their species range arguably have a higher genetic diversity, and some believe it is more important to conserve this high level of diversity (Hardie & Hutchings 2010). However, marginal populations found at the edge of their species range are thought to require conservation priority as they are likely to represent distinct populations with a higher possibility of extinction (Hampe & Petit 2005). Hardie and Hutchings (2010), and Hampte and Petit

(2005) both present detailed analysis with strong conclusions, however it seems to be most advantageous to obtain a variety of results from both population types. This approach would allow the most balanced yield of detailed population data with genetic diversity comparisons for future conservation and management.

Mitochondrial DNA and microsatellite loci

Assessment of phylogeography and conservation genetics requires the use of genetic markers such as mitochondrial DNA, microsatellites, chloroplast DNA and/or RNA (Wan *et al.* 2004). High genetic variability and other properties of mitochondrial DNA (mtDNA) allow evolutionary questions such as population history, patterns of gene flow, and genetic structure to be answered in a straightforward manner (Zink *et al*. 2008). MtDNA is present in all eukaryotic cells and has a small genome size (~17, 000 base pairs [bp]) relative to the millions of base pairs present in nuclear DNA. MtDNA in most animals follows a maternal pattern of inheritance of a single mtDNA genome and is therefore effectively haploid with no recombination. Therefore each distinct mtDNA sequence or haplotype accumulates genetic variation within and among species through mutation. MtDNA has a high evolutionary rate (Brown *et al*. 1979) which initially seemed to contradict fundamental principles of molecular evolution as it was thought that genes that retain conserved functions (like those in mtDNA) evolved at a slow rate. The high rate can be attributed to an inefficient repair mechanism (Avise *et al*. 2007), the oxygen-rich environment that the organelle is surrounded by, and the lack of histone

proteins that keep evolutionary rates in nuclear DNA relatively low (Avise 2009). MtDNA can evolve on the order of ~100 times faster than nuclear DNA and contains an expanded codon-codon recognition pattern wherein third position nucleotide changes are silent in most cases. All of these features make it useful when researching evolutionary divergent lineages (Avise 1989).

Despite the advantages of mtDNA, recent studies that rely completely on mtDNA have been questioned and criticized as this genetic marker has several outstanding limitations (Edwards & Bensch 2009). Primarily, mtDNA is inherited as a single non-recombining unit which allows for a high occurrence of selective sweeps, variable mutation rates, and non-neutrality which all make mtDNA a little less reliable in terms of population history than other genetic markers (Galtier *et al*. 2009; Hurst *et al*. 2005). As mentioned previously, the distinct division between species has been blurred with varying levels of hybrid species and speciation events, which among other things, create differences in lineage sorting between mtDNA trees and species trees (Funk & Omland 2003). To address these limitations, the analysis of multi-locus data, such as microsatellites, in conjunction with the single-locus data provided by mtDNA has been regarded as a better approach.

Microsatellites are tandem repeats of simple sequences (SSRs) that are found throughout the eukaryotic genome (Roy *et al*. 1994) and are usually selectively neutral. The repeating unit length is 2-6 bp (e.g. a dinucleotide of $[CA]_n$, where n is the number of times the unit is repeated) with the usefulness depending on the number of repeats at a particular locus. They are highly variable loci that generally occur in non-coding regions

(Field & Wills 1998). They are also codominant markers (Lepais *et al*. 2011) that can reveal information pertinent to gene flow and hybridization studies (Roy *et al*. 1994). There are many types of studies for which microsatellites are useful: kinship/paternity, gene flow/genetic structure, marker-assisted selection, hybridization, and fingerprinting (Oliveira *et al*. 2005) to name a few.

Freshwater fish in North America

Based on a collection of work by Damien (2010), Matschiner and colleagues note that in rapidly diversifying groups, such as freshwater fish, phylogeographic analysis can identify species as well as determine different modes of speciation (Matschiner *et al.* 2010). Compared to marine fishes, freshwater fishes have smaller ranges and low dispersal potential. Specifically, freshwater environments have borders to limit gene flow such as land and waterfalls, with habitats that are easy to depict based upon available routes of dispersal as they must follow routes such as coastlines or rivers. Finally the biology of freshwater is relatively easy to study when compared to marine systems which means that there is a large resource of known biological history relating to factors such as vagility and generation time (Damien 2010). That being said, freshwater lakes tend to be young relative to marine environments and can undergo dramatic cyclic changes which can lead to hybridization, fragmentation of populations or even small founder populations (Damien 2010). This makes identifying separate populations difficult which in turn also creates difficulties obtaining resolved phylogenies.

A major influence on present-day distributions of freshwater fish in North America was glacial periods and the extent of ice coverage during each glaciation cycle. The Wisconsinan glaciation (~75 000 to 10 000 years ago) had the largest impact on North American fishes as the glaciers extended the furthest south, covering almost all of Canada and northern United States in two massive ice sheets, the Laurentide and the Cordilleran (Rempel and Smith 1988). Repeated expansions and reductions of populations in association with the glacial period played a huge role on the phylogeography of the fish as they had to displace into locations free from ice cover (Ruzzante *et al*. 2006). Based on geological evidence, we know that there were at least five major refugia for fish in North America, the Beringian, Mississippian, Pacific, Nahanni, and Atlantic (Rempel & Smith 1998).

Longnose sucker post-glacial dispersal

As previously mentioned the extensive glacial history of North America has played a major role in present day distributions and dispersal routes of freshwater fish (Bernatchez & Dodson 1994). Range expansions, followed by reductions as each glacial cycle began, gave rise to the fragmented refugial populations along with the re-arranged water systems seen Post-Pleistocene (Weider & Hobaek 1997). Curiously, longnose suckers are absent from drainages leading into Siberia (including the Chukotka Peninsula) from North America that drained the Bering land bridge (~17,000 years ago), however they can be found in Arctic coastal drainages (Kolyma to Yana river) which are more

western in location (McPhail & Taylor 1999). This may suggest longnose sucker distribution in Siberia is based on Illinoian glacial dispersal (~2.5 to 0.2 million years ago; Lindsey & McPhail 1986) rather than the Wisconsinan dispersal pattern consistent with longnose sucker dispersal across North America (McPhail & Taylor 1999).

In North America, deposits of longnose suckers fossils were first found in Southern Indiana and dated to ~200,000 years ago (Pre-Illinoian period) by the characterization of the sediment and isoleucine epimerization in fossil proteins found at the site (Miller *et al*. 1993). This indicates that longnose suckers have maintained similar geographic coordinates to historic populations (McPhail & Taylor 1999) and appear to have flourished prior to the Wisconsinan glacial period in terms of dispersal and colonization ability. The uncovering of fossils in the Yukon dating to ~60,000 years ago (during the Sangamon interglacial period, which followed the Illinoian glaciation; Cumbaa *et al.* 1981), revealed the ability of longnose suckers to be able to not only re-establish populations after harsh glaciation conditions but thrive as well (McPhail & Taylor 1999). Distributions of fish during the Sangamon interglacial period were once again disrupted and fragmented during the Wisconsinan glaciation period when longnose suckers were able to survive in at least three refugia: the Beringian, Pacific, and the Mississippian (McPhail & Taylor 1999; Dillinger *et al*. 1991). This last glacial period was the most recent prior to the current interglacial period (Table 1.1).

Table 1.1 Timeline of the Pleistocene glacial and interglacial cycles

Series/Epoch	Stage/Age	Glacial or interglacial	Years ago (thousands)
Holocene	Recent	Interglacial	12 – present
Pleistocene	Wisconsinan	Glacial	90 – 12
	Sangamon	Interglacial	140 – 90
	Illinoian	Glacial	240 – 140
	Yarmouth	Interglacial	460 – 240
	Pre-Illinoian	Glacial	>780 – 460

Longnose sucker genetics

The Family Catostomidae, like the Salmonidae has undergone a genome duplication event and contains duplicate copies of various genes (Bart *et al*. 2010). This type of event has the potential to play an important role in the evolution of a species and even speciation as these duplicate genes may diverge into new functional units (Bart *et al*. 2010). Therefore catostomids are tetraploids (4n) instead of diploids like most fish (Uyeno and Smith 1972). Uyeno and Smith (1972) karyotyped various sucker species and determined that the DNA content in each cell was about double the regular size of a cell. However, a further study by Ferris and Whitt (1980) determined that only about half of the chromosomes were being expressed while the remaining half were silenced by genic diploidization and therefore suckers were attaining the genetic state of a regular diploid fish.

Interestingly, catostomids including the longnose sucker are thought to be allopolyploids, containing chromosomes from different species, as they are known to readily hybridize with other sucker species (Hubbs 1955). First noted by Carl Hubbs (1955), hybridization in the suckers has since been described as a destructive event that may threaten the genetic variability of native species (McDonald *et al*. 2008). When introduced (and possibly as a native species as well), the white sucker (*Catostomus commersonii*) will readily hybridize with a variety of sucker species (McDonald *et al*. 2008) including the longnose sucker (Nelson 1973). Perry and colleagues (2002) express the concern that hybridization between closely-related species may represent a serious threat of extinction of one or more native species as hybrids may overrun the parental

species and it is crucial to identify the species that do engage in hybridization to prevent loss of genetic diversity and of the species themselves.

Global importance of longnose suckers and local concerns

Important game fish such as lake trout and northern pike, among others, rely on longnose sucker juveniles as a food source (Perry & Casselman 2012) whereas adult longnose suckers are generally eaten by osprey, eagles, and mammals such as otters and bears (Scott *et al*, 1973). Crait and colleagues (2006) discovered that the North American river otter found in Yellowstone National Park were eating longnose suckers as their second most popular food choice; longnose suckers were found in 43% of otter fecal samples. This level of reliance on longnose suckers by other species is significant and points to a need to preserve longnose sucker populations.

The food resource of the longnose sucker and lake whitefish are similar, as are their ranges and distribution. When whitefish are caught in commercial nets, longnose suckers are also caught as by-catch, but as they are not as palatable (Harris *et al*. 1962) they are generally discarded. Longnose suckers do hold a small commercial market as a dog food ingredient (Morrow 1980; Sumida & Anderson 1990) and as they have an interesting and pleasing appearance, longnose suckers can also be found on display in public aquaria. They are caught for sport as they can grow up to 0.609 meters and weigh over 4.5 kilograms.

Most commonly, longnose suckers are caught as subsistence fish in small communities across North America. A report by Sumida and Anderson (1990) described longnose suckers as being harvested at a rate of ~600 fish in a year in Fort Yukon where they make up ~6% of the diet of the community. Hopper and Power (1991) also described how the longnose sucker harvest for the Ojibwa community in Northern Ontario makes up ~19% of total fish catches. Not eaten as much, suckers were generally caught to be used in the trapping of larger mammals in this Ojibwa community. Labrador communities also rely on longnose suckers for food as they are relatively easy to catch and are plentiful (Mackey and Orr 1987).

The recent completion of the Trans-Labrador Highway (TLH) has sparked interest in survival and dispersal of longnose suckers across Labrador. They are more abundant in the west of Labrador, declining in abundance towards the east (Perry & Casselman 2012). The TLH runs from Québec/Labrador City to Cartwright junction, effectively cutting Labrador into two sections. The native range of longnose sucker fully overlaps with the placement of the Trans-Labrador Highway in Labrador. There are two major concerns with the highway: isolation of various populations of longnose suckers; and increased fishing pressure based on the increased accessibility to previously unreachable locations. These concerns are well founded as many Labrador communities rely on this important subsistence fish and without it may suffer from lack of food.

Goals of thesis

Very little is known about the phylogeography or extent of hybridization of longnose suckers in Newfoundland and Labrador, yet these are crucial components of an effective management strategy for these important subsistence fish. Hence I propose to use mtDNA markers along with microsatellite markers to describe population diversity and structure of populations of longnose suckers in Labrador and compare it to fish from other regions in North America such as the Yukon. I wish to make inferences about: Refugial origin and route(s) of post-glacial re-colonization of *Catostomus catostomus* into Labrador, present-day patterns of gene flow and population structure in Labrador, and the extent of hybridization with the white sucker.

REFERENCES

Anamthawat-Jónsson, K. (2012). Current topics in phylogenetics and phylogeography of terrestrial and aquatic systems. Rijeka: In Tech Publishers, ISBN 978-953-51-0217-5.

Avise, J.C. (1989). A role for molecular genetics in the recognition and conservation of endangered species. *Trends in Ecology and Evolution*, 4, 279-281.

Avise, J.C. (2000). Phylogeography: The history and formation of species. Cambridge: Harvard university press, ISBN 0-674-66638-0.

Avise, J.C. (2004). What is the field of biogeography, and where is it going? *Taxon*, 53, 893-898.

Avise, J.C. (2007). Twenty-five key evolutionary insights from the phylogeographic revolution in population genetics. *Phylogeography of southern European refugia*, 7-21.

Avise, J.C. (2009). Phylogeography: Retrospect and prospect. *Journal of Biogeography*, 36, 3-15.

Bart, H.L., Reneau, P.C., Doosey, M.H., Bell, C.D. (2010). Evolutionary divergence of duplicate copies of the growth hormone gene in suckers (Actinopterygii: Catostomidae). *International Journal of Molecular Sciences*, 11, 1090-1102.

Baum, D. (1992). Phylogenetic species concepts. *Trends in Ecology and Evolution*, 7, 1-2.

Bernatchez, L., & Dodson, J. J. (1994). Phylogenetic relationships among Palearctic and Nearctic whitefish populations as revealed by mtDNA variation. *Canadian Journal of Fisheries Ecology*, 51, 240-251.

Brown, W.M., George, M. Jr., Wilson, A.C. (1979). Rapid evolution of animal mitochondrial DNA. *Proceedings of the National Academy of Sciences (PNAS)*, 76, 1967-1971.

Crait, J.R., & Ben-David, M. (2006). River otters in Yellowstone Lake depend on a declining cutthroat trout population. *Journal of Mammalogy*, 87, 485-494.

Crandall, K.A., Bininda-Emonds, O.R.P., Mace, G.M., Wayne, R.K. (2000). Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution*, 15, 291- 295.

Cumbaa, S.L., McAllister, D.E., Morlan, R.E. (1981). Late Pleistocene fish fossils of *Coregonus*, *Stenodus*, *Thymallus*, *Catostomus*, *Lota*, and *Cottus* from the Old Crow basin, northern Yukon, Canada. *Canadian Journal of Earth Sciences*, 18, 1740- 1754.

Rutgers, D. S. (2010). Phylogeography: concepts, intraspecific patterns, and speciation processes. Nova Science Publishers, Hauppauge, NY.

De Queiroz, K. (2007). Species concepts and species delimitation. *Systems Biology*, 56, 879-886.

Dillinger, R.E. Jr., Birt, T.P., Green, J.M., Davidson, W.S. (1991). Post-glacial dispersal of longnose suckers, *Catostomus catostomus*, in the Mackenzie and Yukon drainages. *Biochemical Systematics and Ecology*, 19, 651-657.

Edwards, S., & Bensch, S. (2009). Looking forwards or looking backwards in avian phylogeography? A comment on Zink and Barrowclough 2008. *Molecular Ecology*, 18, 2930-2933.

Ferris, S.D., & Whitt, G.S. (1980). Genetic variability in species with extensive gene duplication: the tetraploid Catostomid fishes. *The American Naturalist*, 115, 650-666.

Field, D., & Wills, C. (1998). Abundant microsatellite polymorphism in *Saccharomyces cerevisiae,* and the different distributions of microsatellites in eight prokaryotes and *S. cerevisiae*, result from strong mutation pressures and a variety of selective forces. *Proceedings of the National Academy of Sciences*, 95, 1647-1652.

Frankham, R. (2003). Genetics and conservation biology. *Comptes Rendus Biologies*, 326, 22-29.

Frankham, R. (2005). Genetics and extinction. *Biological Conservation*, 126, 131-140.

Frasier, T.R., Gillett, R.M., Hamilton, P.K., Brown, M.W., Kraus, S.D., White, B.N. (2013). Postcopulatory selection for dissimilar gametes maintains heterozygosity in the endangered North Atlantic right whale. *Ecology and Evolution*, 3, 3483-3494.

Funk, D.J., & Omland, K.E. (2003). Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal Mitochondrial DNA. *Ecology, Evolution, and Systematics*, 34, 397-423.

Galtier, N., Nabholz, B., Glémin, S., Hurst, G.D.D. (2009). Mitochondrial DNA as a marker of molecular diversity: a reappraisal. *Molecular Ecology*, 18, 4541-4550.

Hampe, A., & Petit, R.J. (2005). Conserving biodiversity under climate change: the rear edge matters. *Ecology Letters*, 8, 461-467.

Hardie, D.C., & Hutchings, J.A. (2010). Evolutionary ecology at the extremes of species ranges. *Environmental Reviews*, 18, 1-20.

Harris, R.H.D. (1962). Growth and reproduction of the longnose sucker, *Catostomus catostomus* (Forster), in Great Slave Lake. *Journal of the Fisheries Research Board of Canada*, 19, 113- 126.

Hewitt, G.M. (2004). Genetic consequences of climatic oscillations in the Quaternary. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 359, 183-195.

Hickerson, M.J., Carstens, B.C., Carvender-Bares, J., Crandall, K.A., Graham, C.H., Johnson, J.B., Rissler, L., Victoriano, P.F., Yoder, A.D. (2010). Phylogeography's past, present, and future: 10 years after Avise, 2000. *Molecular Phylogenetics and Evolution*, 54, 291-301.

Hopper, M., & Power, G. (1991). The fisheries of an Ojibwa community in Northern Ontario. *Arctic*, 44, 267-274.

Hubbs, C.L. (1955). Hybridization between fish species in nature. *Systematic Zoology*, 4, 1-20.

Hurst, G.D.D., & Jiggins, F.M. (2005). Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. *Proceedings of the Royal Society of Biological Sciences*, 272, 1525-1534.

Lepais, O., & Bacles, C.F.E. (2011). De novo discovery and multiplexed amplification of microsatellite markers for black alder (*Alnus glutinosa*) and related species using SSR-enriched shotgun pyrosequencing. *Journal of Heredity*, 102, 627-632.

Lindsey, C.C., & McPhail, J.D. (1986). Zoogeography of fishes of the Yukon and Mackenzie basins. *Zoogeography of North American Freshwater Fishes*, New York, 639-674.

Liu, J.Q., Sun, Y.S., Ge, X.J., Gao, L.M., Qiu, Y.X. (2012). Phylogeographic studies of plants in China: advances in the past and directions in the future. *Journal of Systematics and Evolution*, 50, 267-275.

Mackey, M.G.A., & Orr, R.D. (1987). The evaluation of household country food use in Makkovik, Labrador, July 1980-June1981. *Arctic*, 40, 60-65.

Matschiner, M., Hanel, R., Salzburger, W. (2010). Phylogeography and speciation processes in marine fishes and fishes from large freshwater lakes. *Phylogeography*, 1-29.

McDonald, D.B., Parchman, T.L., Bower, M.R., Hubert, W.A., Rahel, F.J. (2008). An introduced and a native vertebrate hybridize to form a genetic bridge to a second native species. *Proceedings of the National Academy of Sciences*, 105, 10837-10842.

McPhail, J.D., & Taylor, E.B. (1999). Morphological and genetic variation in northwestern longnose suckers, *Catostomus catostomus*: The salish sucker problem. *American Society of Ichthyologists and Herpetologists*, 1999, 884-893.

Miller, B.B., Palmer, D.F., McCoy, W.D., Smith, A.J., Colburn, M.L. (1993). A pre-Illinoian Pleistocene fossil assemblage from near Connersville, Southeastern Indiana. *Quaternary Science*, 40, 254-261.

Morrow, J.E. (1980). The freshwater fishes of Alaska. University of B.C. *Animal Resouces Ecology Library*, 248.

National Science Foundation (2007). Loss of genetic diversity threatens species diversity. *Environmental News Network.*

Nelson, J.S. (1973). Occurrence of hybrids between longnose sucker (*Catostomus catostomus*) and white sucker (*Catostomus commersoni*) in Upper Kananaskis Reservoir, Alberta. *Journal of the Fisheries Research Board of Canada*, 30, 557-560.

Oliveira, E.J., Pádua, J.G., Zucchi, M.I., Vencovsky, R., Vieira, M.L.C. (2005). Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*, 29, 294-307.

Perry, W.L., Lodge, D.M., Feder, J.L. (2002). Importance of hybridization between indigenous and nonindigenous freshwater species: An overlooked threat to North American biodiversity. *Systematic Biology*, 51, 255-275.

Perry, R.C., & Casselman, J.M. (2012). Comparisons of precision and bias with two age interpretation techniques for opercular bones of longnose sucker, a long-lived northern fish. *North American Journal of Fisheries Management*, 32, 790-795.

Rempel, L.L., & Smith, D.G. (1998). Post-glacial fish dispersal from the Mississippi refuge to the Mackenzie River basin. *Canadian Journal of Fisheries and Aquatic Sciences*, 55, 893-899.

Roy, M.S., Greffen, E., Smith, D., Ostrander, E.A., Wayne, R.K. (1994). Patterns of differentiation and hybridization in North American wolflike canids, revealed by analysis of microsatellite loci. *Molecular Biology and Evolution*, 11, 553-570.

Ruzzante, D.E., Walde, S.J., Cussac, V.E,. Dalebout, M.L., Seibert, J., Ortubay, S., Habit, E. (2006). Phylogeography of the Percichthyidae (Pisces) in Patagonia: roles of orogeny, glaciation, and volcanism. *Molecular Ecology*, 15, 2949-2968.

Scott, W.B., & Crossman, E.J. (1973). Freshwater fishes of Canada. Department of Fisheries and Oceans. *Fisheries Research Board of Canada.* Ottawa.

Slatkin, M. (1987). Gene flow and the geographic structure of natural populations. *Science*, 239, 787-792.

Soule, M.E. (1985). What is conservation biology? *BioScience*, 35, 727-734.

Sumida, V.A., & Anderson, D.B. (1990). Patterns of fish and wildlife use for subsistence in Fort Yukon, Alaska. Technical Paper No. 179 Alaska Department of Fish and Game. 1-90.

Uyeno, T., & Smith, G.R. (1972). Tetraploid origin of the karyotype of catostomid fishes. *Science*, 175, 644-646.

Wan, Q-H., Wu, H., Fujihara, T., Fang, S-G. (2004) Which genetic marker for which conservation genetic issue. *Electrophoresis*, 25, 2165-2176.

Weider, L.J., & Hobaek, A. (1997). Post-glacial dispersal, glacial refugia, and clonal structure in Russian/Siberian populations of the arctic *Daphnia pulex* complex. *Heredity*, 78, 363-372.

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

CHAPTER 2: POST-GLACIAL DISPERSAL AND GENETIC STRUCTURE OF LABRADOR POPULATIONS OF LONGNOSE SUCKERS ASSESSED WITH MITOCHONDRIAL DNA: THE EVOLUTIONARY GENETICS OF A POORLY-UNDERSTOOD SPECIES

Abstract

Longnose suckers (*Catostomus catostomus*) are hypothesized to have survived the Wisconsinan glaciation period in at least three main refugia: the Beringian, Pacific, and Mississippian. Although longnose suckers are an important subsistence freshwater fish species, there is little understanding of the phylogeography of this species across its full range. Here we compare longnose suckers in Labrador, Canada, to populations in Yukon and British Columbia. Labrador provides a unique sampling area as much of Labrador has been undisturbed by human involvement until recently with the construction of the Trans-Labrador highway, which bisects the region with the potential to hinder the dispersal of longnose suckers. As part of a long-term monitoring plan, we herein document patterns of genetic diversity and population structure, and infer post-glacial re-colonization routes, using the mitochondrial cytochrome b, NADH dehydrogenase 4, and cytochrome c oxidase 1 genes. Among 362 fish, the combined 1244 bp sequence fragment yielded 20 haplotypes within Labrador, five Yukon haplotypes, and four British Columbia haplotypes. No sequences were shared among any of the provinces and strong regional structure was supported by Spatial Analysis of Molecular Variance (SAMOVA) among

western and eastern populations. Based on the SAMOVA and a median spanning network of haplotypes we find evidence for one main glacial refugial source of suckers within Labrador: the Atlantic refugium. However we also have evidence based on the measures of F_{ST} and maximum likelihood phylogeny to suggest a minor influence of multiple refugia in western and northern Labrador, including the Beringian and Mississippian. A recent population expansion was inferred from a pairwise mismatch distribution, and negative Tajima's D and Fu's Fs values, indicating that populations in Labrador are not in equilibrium since re-colonization. Population genetic structure exists among regions within Labrador, but less so among populations within regions, reflecting the historical imprint of re-colonization from different refugia and possibly ongoing gene flow.

INTRODUCTION

Catostomid fishes have wide-ranging diversity in many characters including morphological traits, habitat, biochemical and genetic traits, and evolutionary history. This diversification of species in the family Catostomidae has led to many taxonomic and systematic (Doosey *et al*. 2010), cladistic (Smith & Koehn 1971), phylogenetic (Harris *et al*. 2002) and hybridization studies (Hubbs 1955; McDonald *et al.* 2008). Despite these efforts, historical origins and phylogeography of many catostomids are unknown or unclear and require further investigation. One such catostomid, the longnose sucker (*Catostomus catostomus*), is a long-lived, cold water fish found within freshwater of North America and in Siberia (Doosey *et al*. 2010; McPhail & Taylor 1999; Perry & Casselman 2012).

The extensive glacial history of North America has played a major role in present-day distributions and dispersal routes of freshwater fish (Millette *et al*. 2011). Range expansions followed by glacial reductions gave rise to fragmented refugial populations and re-arranged water systems (Weider & Hobaek 1997) eventually establishing the current landscape. Based on absence of certain clasts in Saskatchewan gravel as well as radiocarbon dating (Young *et al*. 1994), Rempel and Smith (1998) determined that the Wisconsinan glaciation period (~75 to 10 thousand years ago) had the greatest impact on fish fauna as the Laurentide ice sheet extended furthest, covering the high elevations in Alberta for the first time. During this period, longnose suckers survived in three of five known refugia (pockets of water free from ice-cover), the Beringian, Pacific, and the Mississippian (McPhail & Taylor 1999; Dillinger *et al*. 1991) with the remaining two

being the Atlantic and Nahanni (Rempel & Smith 1998). During the subsequent deglaciation period, spillways of water from these refugia, created from the temporary swell of melt-water, provided fish the means to disperse out into new locations previously covered by ice (Rempel & Smith 1998). In this way freshwater fish including the longnose sucker were able to ultimately colonize all of North America.

Only two publications have focussed on the evolutionary genetics of the longnose sucker. Dillinger and colleagues (1991) investigated the post-glacial dispersal of longnose suckers in the Yukon and Northwest Territories using gill raker counts and blood transferrin alleles. They hypothesized that latitudinal clinal variation was occurring based on the difference in lateral line scale counts between locations, and were able to identify dual refugial origins (Beringian and Mississippian) based on protein transferrin alleles. McPhail and Taylor (1999) described a distinctive longnose sucker called the Salish sucker found in the lower Fraser Valley, British Columbia and western Washington, inferred from sequences of the mitochondrial NADH dehydrogenase subunit 2 and cytochrome b genes. Existence of "eastern" and "western" North American longnose suckers was also hypothesized from these data with a sequence divergence of ~2% between eastern and western groups of individuals (McPhail & Taylor 1999). A study of *Salmo* species found within species sequence divergences of 0-1.5%, and inter-species divergences of 2-3.5% (Wilson *et al.* 1985), indicating that the sequence divergence determined by McPhail and Taylor may be considerable. Based on mitochondrial data, McPhail and Taylor (1999) concluded that British Columbia was likely colonized by longnose suckers from three glacial refugia: the Pacific, Beringian and Mississippian.

Analysis of mitochondrial DNA (mtDNA) has proven to be a relevant and reliable evolutionary genetic marker which has made enormous contributions to the conservation and management of a diverse range of species (Nabholz *et al*. 2008). This compact genetic marker is highly conserved in terms of gene content and order, but evolves quickly, rapidly accumulating sequence differences within and among populations which are useful for phylogenetic and phylogeographic studies (Zardoya *et al*. 1995). In addition, mtDNA is strongly impacted by the influences of genetic drift (Hurst *et al*. 1999) and may become genetically fixed between species following speciation events (Curole & Kocher 1999) which can aid in studies of introgression and hybridization (Inoue *et al*. 2000).

Here we compare longnose suckers in Labrador, Canada, the most northeastern part of their range in North America, to northern and western populations in British Columbia and Yukon. Labrador provides a unique landscape within which to study this species as many locations remain untouched, and human involvement occurring only in secluded areas. However, the construction of a Labrador-wide provincial highway, the Trans-Labrador Highway, within the last 10 years, potentially influences future ecological and evolutionary dynamics of freshwater fish in the region, including suckers. As part of a long-term monitoring plan, the provincial Department of Environment and Conservation is interested in documenting patterns of genetic diversity and population structure of some of these species. To aid in this effort and contribute to understanding the evolutionary genetics of a poorly studied species we make inferences about post-glacial dispersal routes and phylogeography as revealed by nucleotide sequence

variability within three mitochondrial genes for the longnose sucker: cytochrome b (CYT b), cytochrome c oxidase I (CO1), and NADH dehydrogenase subunit 4 (ND4).

MATERIALS AND METHODS

Study area and sample collection

Caudal fin clippings (~200 mm²) were taken from longnose suckers captured via gill netting by Newfoundland and Labrador provincial wildlife biologists between the years 2003 and 2012. A total of 424 fish were collected from 14 lakes throughout Labrador. These were divided into four regions as used by the Wildlife Division of the provincial Department of Environment and Conservation in order to document regional variation: northern, central, western and southeastern (Figure 2.1). For comparison to Labrador, 44 fish fins were collected from Lake Laberge (YK) in the Yukon by Oliver Barker, and a total of 60 fish fins from Grizzly Lake (BCA) and Teardrop Lake (BCB) in British Columbia were provided by Jessica Courtier (Figure 2.2). All fish fins were stored at -20°C.

DNA isolation, polymerase chain reaction (PCR), and DNA sequencing

Genomic DNA was extracted from a portion (~4 mm^2) of the caudal fin clipping using the Qiagen QIAamp® DNA Mini Kit (Qiagen Inc., Toronto, Canada) according to the manufacturer's protocol detailing tissue sample extraction.

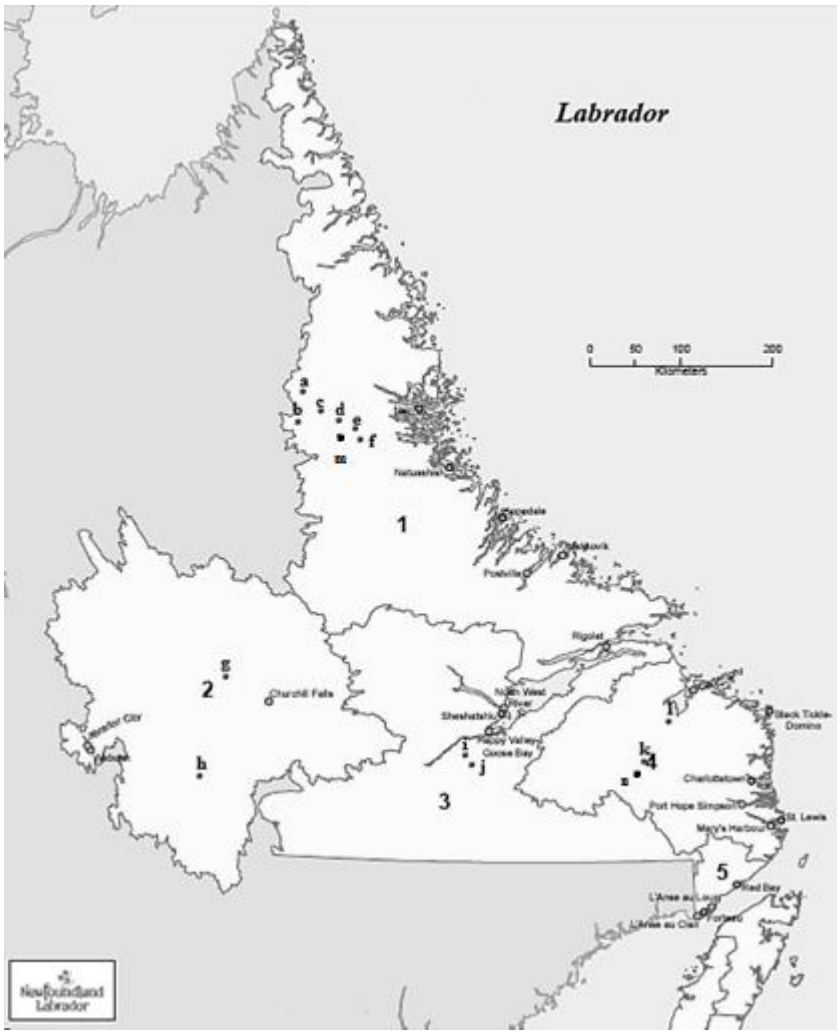

Figure 2.1. Geographic locations of the 14 longnose sucker populations investigated in Labrador with numbers 1 through 4 indicating each defined region (1, Northern Labrador; 2, Western Labrador; 3, Central Labrador; and 4, Southeastern Labrador). Exact coordinates for locations Ikadlivik Brook Watershed in Northern Labrador and Birchy Lake in Southeastern Labrador are not known and are shown on the map with initials m and n respectively; locality abbreviations are: a Genetics H Lake, b Strange Lake, c Mistake Lake, d Walkabout Lake, e Konrad Lake, f Cabot Lake, g Lobstick Lake, h Lac Joe, i The Right Lake, j Mercier Lake, k Brettney Lake, l Crooked Lake

Figure 2.2. Geographic locations of the three populations of longnose suckers studied in northwest Canada, Lake Laberge in Yukon (YK), Grizzly Lake (BCA) and Teardrop Lake (BCB) in British Columbia

Mitochondrial regions CYT b, ND4, and CO1 were chosen as they represent different parts of the genome and may uncover different patterns of selection or have varying levels of informativeness. Data for these regions were also readily available in NCBI and were used in the primer design (http://www.ncbi.nlm.nih.gov/). A 364 base pair (bp) fragment of cytochrome b from the mitochondrial DNA (mtDNA) was amplified using the following primer pair: forward LNSCytB-1F-M13 5'-GTAAAACGACGGCCAGTATGATGAAA-3' and reverse LNSCytB-1R-M13 5'-CAGGAAACAGCTAATATTTGTCCTCA-3' (Langille *et al*. 2014; Appendix A). NADH dehydrogenase 4 (350bp) was targeted by PCR amplification using primers forward LNS-ND4-1F5'-GATTTTGGCCAGCCAGAACCA-3' and reverse LNS-ND4-1R-5'-TCAGGACTCAAGGACAAGGGGT-3' (Langille *et al*. 2014). Finally, cytochrome c oxidase I (630bp) was targeted by PCR amplification using primers forward LNS-CO1-1F- 5' -TCAACCAACCACAAAGACATTGGCAC and reverse LNS-CO1-1R- 5'- TAGACTTCTGGGTGGCCAAAGAATCA (Langille *et al*. 2014). Polymerase chain reaction (PCR) was carried out on all samples following Langille and colleagues' protocol, with a final volume of 25uL containing 1X Qiagen PCR buffer (Qiagen Inc.), 200 µM dNTPs (New England Biolabs Inc., Whitby, Canada), 400 nM each primer, 1U HotStar *Taq* DNA polymerase (Qiagen Inc.), and 2 µL genomic DNA (25-200 ng). The amplification profile for CYT b and ND4 consisted of an initial denaturation for 5 min at 95℃, followed by 30 cycles of 94℃ for 60 sec, 45℃ (cytb) and 50℃ (ND4) for 30 sec and 72℃ for 90 sec, with a final extension at 72℃ for 5 min. The amplification profile for CO1 consisted of an initial denaturation for 2 min at 95℃, followed by 35 cycles of 94℃ for 30 sec, 54℃ for 30 sec and 72℃ for 60 sec, with a final extension at 72℃ for 10 min.

Sequencing was carried out on purified PCR products using BigDye Terminator chemistry, on the ABI Prism 3730 DNA Analyzer (Applied Biosystems Inc., Foster City, USA) using Sequencing Analysis v5.0 software.

Sequence analysis

All sequence reads were aligned and edited using Sequencher v4.9 (Gene Codes Corporation, Ann Arbor, USA). Each individual was represented by forward and reverse sequence reads for each gene region; these were edited together into a combined consensus sequence of all three genes for each individual. All sequences where then checked with the Basic Local Alignment Search Tool (BLAST; http://blast.ncbi.nlm.nih.gov/Blast.cgi) online, for species identification. Distinct haplotypes were identified using MEGA v.4.0 (Tamura *et al*. 2007).

Genetic diversity and tests of selection

We calculated haplotype diversity (h) and nucleotide diversity (π) within populations and overall, using ARLEQUIN v3.11 software. Rates of nonsynonymous substitutions (d_N) and synonymous substitutions (d_S) were compared using Z-tests of selection with the Nei-Gojobori method and Jukes Cantor distance correction (Nei & Gojobori 1986) for the alternate hypotheses of strict neutrality (Ha: $d_N \neq d_S$) using MEGA. When d_N is equal to d_S ($d_N - d_S = 0$), the gene is evolving randomly. However when d_N is in excess ($d_N - d_S > 0$) the gene may be undergoing positive selection and when ds is in excess ($d_N - d_S < 0$), the pattern of variation is consistent with purifying selection.

McDonald-Kreitman's test of neutrality (McDonald & Kreitman 1991) was performed in DNAsp v5.0 (Librado & Rozas 2009) by comparing longnose suckers and white suckers. Nonsynonymous and synonymous sites were scored as fixed (d_N or d_S) or polymorphic (p_N or p_S). Under neutral theory we expect d_N/d_S (fixed sites) = p_N/p_S (polymorphic sites). A $d_N/d_S > p_N/p_S$ would indicate positive selection because the nonsynonymous to synonymous changes within species is lower than the nonsynonymous to synonymous changes between species while $d_N/d_S < p_N/p_S$ would indicate negative or purifying selection because the nonsynonymous to synonymous changes within species is higher than the nonsynonymous to synonymous changes between species.

Population structure

Population genetic structure was assessed using Spatial Analysis of Molecular Variance (SAMOVA; Dupanloup *et al*. 2002) which identifies groups of populations that are geographically proximal, genetically homogeneous, and as differentiated from each other as possible (Dupanloup *et al*. 2002). As recommended, the annealing process was repeated 100 times and the largest statistically significant estimate of among-group differentiation (F_{CT}) was chosen to represent the best sample grouping. We also examined population structure using overall and pairwise measures of standard frequency-based F_{ST} (Nei & Gojobori 1986) as calculated in ARLEQUIN, among populations and among regions in Labrador.

Population expansion

Tajima's D (10,000 replicates; Tajima 1989) and Fu's Fs (10,000 replicates; Fu 1997) tests of neutrality were calculated to evaluate population range expansion using ARLEQUIN. The pairwise mismatch distribution was also computed; this summarizes information pertaining to genetic differences between all paired individuals; a unimodal distribution is an indication of population expansion. The mismatch distribution was compared to a null model of demographic equilibrium using 10,000 bootstrap replicates in ARLEQUIN. SDD is the sum of squared deviations between the observed and expected mismatch and is used as the model test statistic. The distribution was also characterized with the raggedness index (r; Harpending 1994), a statistic which reflects the modality of the curve, also tested against a null model of population range expansion.

Phylogeny construction and network analysis of haplotypes

The best-fit nucleotide substitution model among haplotypes was estimated using MEGA, by testing the goodness of fit of the data to different models of molecular evolution (Tamura *et al*. 2011). Phylogenies of haplotypes were computed in MEGA using this model with maximum likelihood, maximum parsimony, and neighbour-joining algorithms; bootstrap support of nodes was estimated with 1000 replicates. All phylogenies were rooted using sequences from the white sucker, *Catostomus commersoni* which is a closely related species also found in Labrador. In addition, a Bayesian analysis of phylogeny was constructed with MRBAYES 3.2.0 (Ronquist & Huelsenbeck 2003). We used the Generalized Time Reversible (GTR) model of DNA substitution and γ-distributed rate variation across sites. We used the default search for 1,000,000

generations where every 1 in 10 were sampled with the first 25,000 trees of the sample discarded, resulting in a total of 75,000 trees estimated using Bayesian posterior probabilities. Finally, a median spanning network of haplotypes was computed to describe the relationships and relative frequencies of haplotypes, using the program NETWORK v.4.6.1.2 (Bandelt *et al*. 1999).

RESULTS

Sequence analysis

Sequences of a 290 bp fragment of the cytochrome b gene, 310 bp fragment of NADH dehydrogenase 4, and 645 bp fragment of cytochrome c oxidase 1 were obtained for a total of 362 individuals. Of these, 139 individuals were from northern Labrador (NL), 38 from central Labrador (CL), 40 from western Labrador (WL), 70 from southeastern Labrador (SE), 32 from the Yukon (YK), and 43 from British Columbia (BC). A total of 29 haplotypes were identified where 20 represent Labrador haplotypes, 5 represent Yukon haplotypes, and 4 represent British Columbia haplotypes, with no shared haplotypes among these locations (Table 2.1). Initial examination of sequences indicated that haplotype 9 represented by four individuals found in southeastern Labrador, was substantially different (i.e. 8.43% variable) from the other longnose sucker haplotypes but closely matched the reference white sucker (*Catostomus commersonii*) sequence. All haplotypes subsequently were uploaded to the NCBI website using the BLAST tool for species identification. Haplotypes 1 through 29 with the exception of haplotype 9 matched longnose sucker with 99-100% similarity, while haplotype 9 matched white

Table 2.1. Haplotypes determined among 362 putative longnose sucker individuals, with haplotype frequency (N) and region of location for each

Haplotype	N	Region
1	234	Northern, Central, Western, and Southeastern Labrador
2	29	British Columbia
3	22	Yukon
4	11	Northern and Western Labrador
5	9	Northern and Western Labrador
6	9	British Columbia
7	5	Northern and Southeastern Labrador
8	5	Yukon
9	4	Southeastern Labrador
10	3	Yukon
11	3	Southeastern Labrador
12	3	Southeastern Labrador
13	3	British Columbia
14	2	Northern Labrador
15	2	Northern Labrador
16	2	Northern Labrador
17	2	Northern Labrador
18	2	Northern Labrador
19	2	British Columbia
20	1	Central Labrador
21	1	Southeastern Labrador
22	1	Western Labrador
23	1	Western Labrador
24	1	Western Labrador
25	1	Northern Labrador
26	1	Southeastern Labrador
27	1	Northern Labrador
28	1	Yukon
29	1	Yukon

sucker with 99% similarity, suggesting that haplotype 9 may be white suckers rather than longnose suckers. Therefore, unless otherwise stated, haplotype 9 was omitted from further analysis. Of the 29 haplotypes, 10 were singletons while 19 were shared among two or more individuals. Including haplotype 9, 11.1 % of nucleotide sites were variable which translated into an amino acid variability of 1.9%. After removing haplotype 9, the nucleotide variability reduced to 2.6% of sites (Table 2.2), but the amino acid variability only fell to 1.4%, indicating that many of the nonsynonymous differences occur among individuals of the same species. Of these variable sites among longnose suckers, 22 were singleton variable sites while 15 were parsimony informative sites among sample locations. The CO1 gene region was the most conserved with 1.86% variable sites among 13 haplotypes, while the ND4 and CYT b gene regions were the least conserved with 3.55% among 13 haplotypes and 3.10% among10 haplotypes respectively.

Genetic diversity and tests of selection

The number of haplotypes (k) in each region varied from 2 to11 with the lowest occurring in central Labrador and the highest occurring in northern Labrador. These high and low levels are not solely explained by sample size as one of the lakes with the lowest number of haplotypes had one of the largest sample sizes (N=25). Haplotype diversity was $h = 0.309$ over all localities, with the highest occurring in western Labrador ($h = 0.597$) and the lowest in central Labrador ($h = 0.0526$; Table 2.3). Nucleotide diversity

Table 2.2. Variable sites found in a 1244 bp combined fragment of cytochrome b, NADH dehydrogenase 4, and cytochrome c oxidase 1 in 28 haplotypes identified among 358 longnose suckers. The nucleotide substitutions are numbered according to position in codon as indicated by the top row. N indicates an unknown base and haplotype 9 has been omitted from this table.

Haplotype	Position and nucleotide change
	[CO1][ND4][Cytb]
	1333323333 2333333331 3331313333 31
1	GCGCATAGAA TTGTGGGTTG TTCCCGTTTG TG
2	...T..G... .....AA.C. CG.......A C.
3	......G... .....AA.C. CG.......A C.
4	A......... ....
5	.T......G. .....AA.C. C........A CA
6	....AA... .G........ C.
7	.... .G........ ..
8	......G.G. .....AA.C. CG.......A C.
10	......G.G. .....AACC. CG.......A C.
11	....A ..
12	....A.... ..
13	....AAA... .G.......A C.
14	........G. ......A... ....C. C.
15	........GG .....AA... ......CC.. C.
16	........G. .....AA... ......CC.. C.
17	N......... CA........ ....
18	....A ....
19	........G. .....AA... .G........ C.
20	N....C..NN NN........ .G........ ..
21	N........N NNA....... ....

```
22        .T..G...G. ...G.AA.C. C........A CA
23        NT..G...G. NN........ .G.......A CA
24        N......... ...G.AA.C. C......... ..
25        .........N NN........ ..G....... ..
26        .......... .......... ....T..... ..
27        N......... .......... ...T...... ..
28        ..A...G... .....AA.C. CG.......A C.
29        ......GAG. .....AA.C. CG.......A C.
```

Table 2.3. Number of individuals in each location used in all population comparisons with corresponding measures of genetic diversity and numbers of haplotypes per location with the omission of haplotype 9 - sample size (*N*), number of haplotypes (*k*), haplotype diversity (*h*), nucleotide diversity (π), and standard deviation (SD), locality abbreviations are: CAB Cabot Lake, GEN Genetics H Lake, IKA Ika Lake, KON Konrad Lake, MIS Mistake Lake, STR Strange Lake, WAL Walkabout Lake, MER Mercier Lake, RIG The Right Lake, BIR Birchy Lake, BRE Brettney Lake, CRO Crooked Lake, JOE Lac Joe, LOB Lobstick Lake, BCA British Columbia Grizzly Lake, BCB British Columbia Teardrop Lake, YUK Yukon.

Population	Location	*N*	*k*	$h \pm$ SD	$\pi \pm$ SD (x10^{-3})
CAB	NL	25	2	0.080 ± 0.072	0.450 ± 0.424
GEN	NL	21	6	0.609 ± 0.114	1.84 ± 1.17
IKA	NL	13	1	0.000 ± 0.000	0.000 ± 0.000
KON	NL	24	3	0.236 ± 0.109	0.820 ± 0.636
MIS	NL	19	2	0.105 ± 0.092	0.762 ± 0.609
STR	NL	19	4	0.521 ± 0.123	1.54 ± 1.03
WAL	NL	18	2	0.209 ± 0.116	0.168 ± 0.240
MER	CL	13	2	0.154 ± 0.126	0.247 ± 0.306
RIG	CL	25	1	0.000 ± 0.000	0.000 ± 0.000
BIR	SE	22	4	0.178 ± 0.106	0.146 ± 0.219
BRE	SE	28	3	0.320 ± 0.106	0.589 ± 0.503
CRO	SE	16	4	0.425 ± 0.133	0.362 ± 0.378
JOE	WL	23	6	0.656 ± 0.096	3.69 ± 2.10
LOB	WL	17	3	0.522 ± 0.101	4.30 ± 2.47
BCA	BC	29	2	0.246 ± 0.094	0.990 ± 0.720
BCB	BC	14	4	0.780 ± 0.061	2.35 ± 1.47
YUK	YK	32	5	0.508 ± 0.096	0.577 ± 0.494
Overall	N/A	358	28	0.309 ± 0.036	4.76 ± 0.763

was $\pi = 1.11x10^{-3}$ overall, with the highest level occurring in western Labrador ($\pi = 3.94x10^{-3}$) and the lowest level occurring in central Labrador ($\pi = 8.50x10^{-5}$; Table 2.3).

The ratio of nonsynonyous to synonymous variation over all haplotypes was consistent with purifying selection (Z-test: $d_N - d_S = -7.86$, $P < 0.05$). For the McDonald-Kreitman test, there were 24 polymorphic synonymous substitutions within longnose and 102 fixed synonymous differences between longnose and white suckers. There were 8 polymorphic nonsynonymous substitutions and 9 fixed nonsynonymous differences. The resulting ratios of $d_N/d_S = 0.0968 < p_N/p_S = 0.286$ were not significantly different from each other (Fisher exact test $P = 0.0849$) indicating no departure from expectations under neutrality.

Population structure

The SAMOVA analysis showed that as the number of predefined groups (K) increased, the F_{CT} increased until K=3 and decreased steadily thereafter (Figure 2.3), while the F_{SC} value decreased constantly (Table 2.4; Figure 2.3). The decrease of F_{CT} after K=3 suggests that adding extra groups after this point does not improve the model of population structure. Division of populations into the three groups (K=3; Table 2.4) resulted in all Labrador populations clustering together, the Yukon and BCA (Grizzly Lake, British Columbia) grouping together, and BCB (Teardrop Lake, British Columbia) forming its own group. The SAMOVA detected significant levels of variation among regions, among populations within regions, and within populations. Highest levels were among

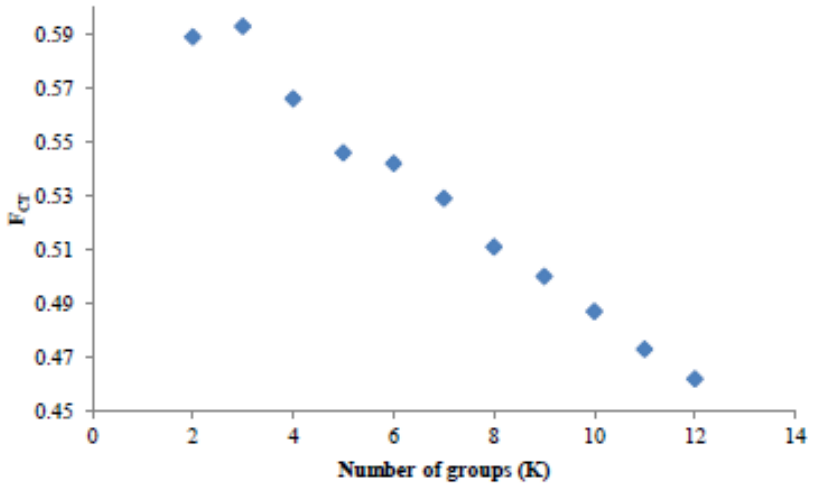

Figure 2.3. A plot of the F_{CT} parameter (differentiation between groups) for different values of *K*, the number of population groups, generated using SAMOVA. Analysis was performed using all 17 longnose sucker populations from Labrador, Yukon, and British Columbia.

Table 2.4. SAMOVA analysis of mtDNA sequences from 17 populations located in Labrador, Yukon, and British Columbia. K indicates the predefined number of groups into which all populations are separated. F_{CT} is the differentiation among regions, F_{SC} is the differentiation among populations within groups, and F_{ST} is the differentiation among populations. Location abbreviations are defined in Table 2.3

K	F_{CT}	F_{SC}	F_{ST}	Population groupings
2	0.589	0.111	0.635	YUK/BC; LAB
3	0.593	0.095	0.631	YUK/BCA; BCB; LAB
4	0.566	0.045	0.586	YUK/BCA; BCB; BIR; LAB
5	0.546	0.085	0.585	YUK; BCA; BCB; CRO; LAB
6	0.542	0.031	0.556	YUK; BCA; BCB; BIR; MER; LAB
7	0.529	-0.016	0.522	YUK; BCA; BCB; BIR; JOE; LOB; LAB
8	0.511	-0.012	0.505	YUK; BCA; BCB; BIR; JOE; LOB; MER; LAB
9	0.500	-0.036	0.482	YUK; BCA; BCB; BIR; JOE; LOB; CRO; STR; LAB
10	0.487	-0.033	0.469	YUK; BCA; BCB; BIR; JOE; LOB; CRO; STR; MER; LAB
11	0.473	-0.037	0.453	YUK; BCA; BCB; BIR; JOE; LOB; CRO; STR; MER; GEN; LAB
12	0.462	-0.033	0.444	YUK; BCA; BCB; BIR; JOE; LOB; CRO; STR; MER; GEN; IKA; LAB

regions (59.3%) and lowest levels were among populations within regions (3.86%; Table 2.5).

The overall measure of F_{ST} calculated from pairwise differences between Labrador longnose suckers and western Canada longnose suckers was $F_{ST=}$ 0.803, which indicate strong population genetic structure among regions in Canada, consistent with the SAMOVA. Pairwise estimates of population differentiation among lakes in Labrador range from -0.039 to 0.302 (Table 2.6). Both western Labrador lakes (JOE and LOB) were significantly differentiated from certain lakes in northern and southeastern Labrador, and Strange Lake (STR) in northern Labrador was differentiated from Brettney Lake (BRE) in southeastern Labrador (Table 2.6). Finally, The Right (RIG) and Genetics H Lake (GEN), both in northern Labrador, were significantly different from each other. Pairwise F_{ST} was also calculated for regional differentiation estimates in Labrador and revealed significant differentiation of western Labrador from the rest of Labrador (Table 2.7).

Population expansion

Tajima's D was significantly less than zero in all regions of Labrador separately, except western Labrador, and overall (Table 2.8), indicating a relatively recent population expansion. Fu's Fs statistic, which is considered more sensitive to population expansion then Tajima's D, was also negative for all regions except western Labrador and significantly so for southeastern and northern Labrador, and overall (Table 2.8) also

Table 2.5. Hierarchical SAMOVA for 17 populations of longnose suckers performed using K = 3 groups as inferred from SAMOVA. The F-statistic for among groups is represented by F_{CT}, among populations within groups is represented by F_{SC}, and within populations is represented by F_{ST}.

Source of variation	Variance components	Percent variation	F-statistic
Among groups (regions)	3.042	59.28	0.5930*
Among populations/ within groups (regions)	0.198	3.860	0.0950*
Within populations	1.891	36.86	0.6310*

*$P < 0.05$

Table 2.6. Measures of pairwise population differentiation using F_{ST} values among Labrador longnose sucker populations. Values followed by * are significant at $\alpha = 0.05$ after sequential Bonferroni correction for multiple comparisons.

	MER	WAL	BIR	JOE	LOB	STR	KON	IKA	RIG	BRE	MIS	CAB	CRO	GEN
MER	0.000													
WAL	-0.039	0.000												
BIR	0.037	0.062	0.000											
JOE	0.096	0.132	0.076	0.000										
LOB	0.200	0.247	0.073	-0.016	0.000									
STR	0.037	0.082	0.061	0.046	0.136	0.000								
KON	0.004	0.042	0.078	0.139*	0.251*	0.075	0.000							
IKA	-0.000	0.033	0.040	0.108	0.212	0.057	0.015	0.000						
RIG	0.054	0.086	0.086	0.169*	0.302*	0.114	0.056	0.000	0.000					
BRE	0.028	0.049	0.091	0.154*	0.272*	0.097*	0.023	0.021	0.060	0.000				
MIS	-0.018	0.009	0.061	0.089	0.182	0.007	0.028	-0.021	0.015	0.033	0.000			
CAB	-0.024	0.009	0.082	0.131	0.250	0.043	0.034	-0.029	0.000	0.039	-0.032	0.000		
CRO	-0.021	0.002	-0.027	0.041	0.064	0.011	0.022	-0.014	0.027	0.033	0.004	0.023	0.000	
GEN	-0.004	0.031	0.064	0.028	0.115	-0.027	0.038	0.007	0.049*	0.051	-0.014	0.007	0.010	0.000

Table 2.7. Measures of pairwise population differentiation (F_{ST}) among regions in Labrador. Values followed by * are significant at $\alpha = 0.05$ after sequential Bonferroni correction for multiple comparisons.

	Central	Western	Southeastern	Northern
Central	0.00			
Western	0.197*	0.00		
Southeastern	0.00230	0.239*	0.00	
Northern	0.00161	0.228*	0.0153	0.00

Table 2.8. Neutrality tests and mismatch expansion estimates for longnose suckers in each of four regions in Labrador excluding haplotype 9; SDD is the sum of squared deviations in the mismatch distribution.

Region	Tajima's D	Fu's Fs	SDD	Raggedness (r)
Northern Labrador	-2.25*	-7.04*	0.0046	0.583
Western Labrador	0.879	4.98	0.0496	0.188
Central Labrador	-1.49*	-0.611	0.0015	0.903
Southeastern Labrador	-1.58*	-2.95*	0.0053	0.288
Overall	-1.94*	-8.44*	0.0142	0.491

*$P < 0.05$

suggesting recent population expansion in at least northern and southeastern Labrador. The sum of squared deviations of the mismatch distribution from a null model of population expansion was not significant (Table 2.8) so we cannot reject the null model. The raggedness statistic reinforces this assessment as r does not differ significantly from expected under the null model of expansion (r = 0.362; P= 0.568). Based on the informative neutrality tests and the mismatch distribution, we can conclude that longnose suckers in Labrador underwent a recent population expansion.

Phylogenetic analysis

The best substitution model for the data (HKY + G; where HKY = Hasegawa-Kishino-Yano [Hasegawa *et al*. 1985]) was chosen based on the lowest BIC scores as this term compensates for the number of parameters in the model and therefore the lower the BIC score the higher the likelihood of the model (Minin *et al.* 2003). The neighbour-joining, maximum parsimony and maximum likelihood phylogenies were all identical in terms of pairings of haplotypes at nodes and so only the maximum likelihood tree is shown here (Figure 2.4). All longnose sucker haplotypes clustered in one group separate from haplotype 9 which was used as a representative of a white sucker sequence. All longnose sucker haplotypes further separated into three main groups: a Labrador group, a Yukon/British Columbia haplotype 2/western Labrador group, and a British Columbia/Northern Labrador group. In addition, a Bayesian analysis of phylogeny (Figure 2.5) strongly supported the grouping of Yukon and British Columbia

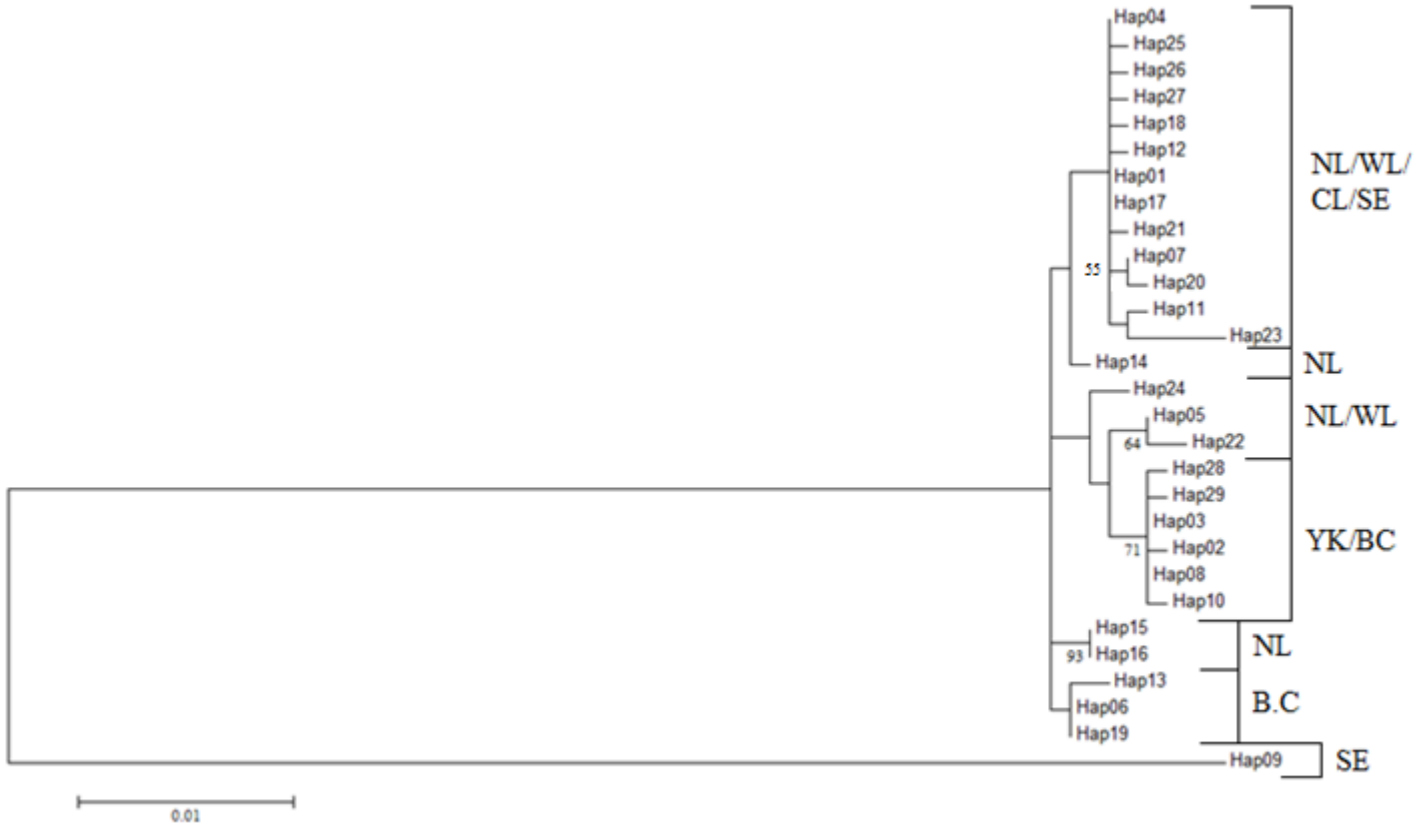

Figure 2.4. Maximum likelihood phylogeny of phylogenetic relationships among longnose sucker mtDNA haplotypes. The HKY model of DNA substitution was used and bootstrap values greater than 50% of 1000 replicates are shown at nodes.

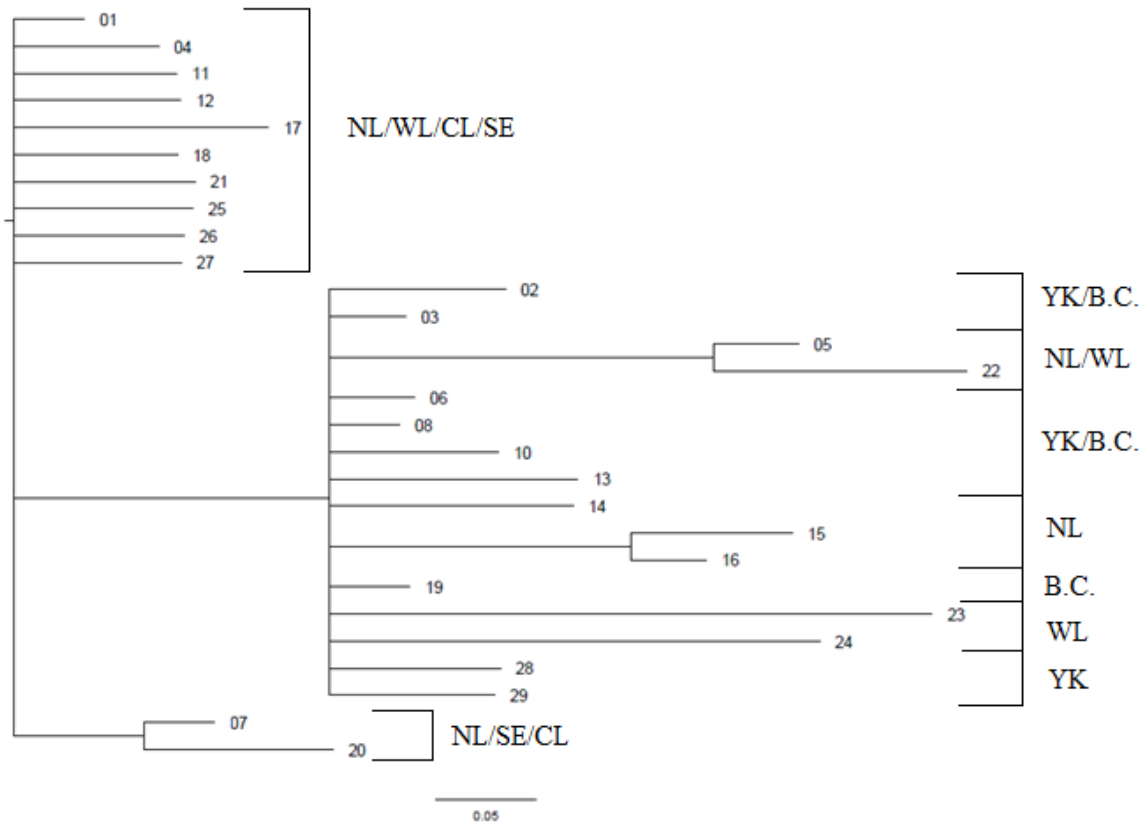

Figure 2.5. Phylogenetic relationships among longnose suckers mtDNA haplotypes inferred using a Bayesian analysis. Support is shown only for nodes with > 50% posterior probability.

together (except for haplotypes 6, 13, and 19 which were placed within the Labrador haplotypes) as well as haplotypes 15 and 16, 7 and 20. These haplotypes also grouped closer to Yukon and British Columbia populations. Haplotypes 5, 22, and 24 (western Labrador haplotypes) were strongly supported as part of the western Canada set of haplotypes.

The median spanning network of all haplotypes (Figure 2.6) indicated the same three groupings as the maximum likelihood phylogeny: a Labrador group, which contains almost 94% of all Labrador individuals, a Yukon/British Columbia/northern/western Labrador group (3.50% of Labrador individuals), and a northern Labrador/British Columbia group (2.80% of all Labrador individuals).

DISCUSSION

The order Cypriniformes is highly diverse, containing fish found on several continents, of many sizes and with different reproductive modes (bisexual and unisexual polyploidy or tetraploidy; Saitoh *et al*. 2006). With respect to marine fish, freshwater fish tend to have limited dispersal ability as a direct consequence of barriers to gene flow resulting in significant levels of genetic differentiation among populations (McGlashan *et al*. 2000). Proctor and colleagues (2004) describe dispersal as a mechanism to decrease inbreeding, and to reduce competition for resources and mates among relatives so as to maximize genetic fitness of progeny. Therefore investigations into historical and contemporary dispersal patterns of different populations can help judge genetic stability

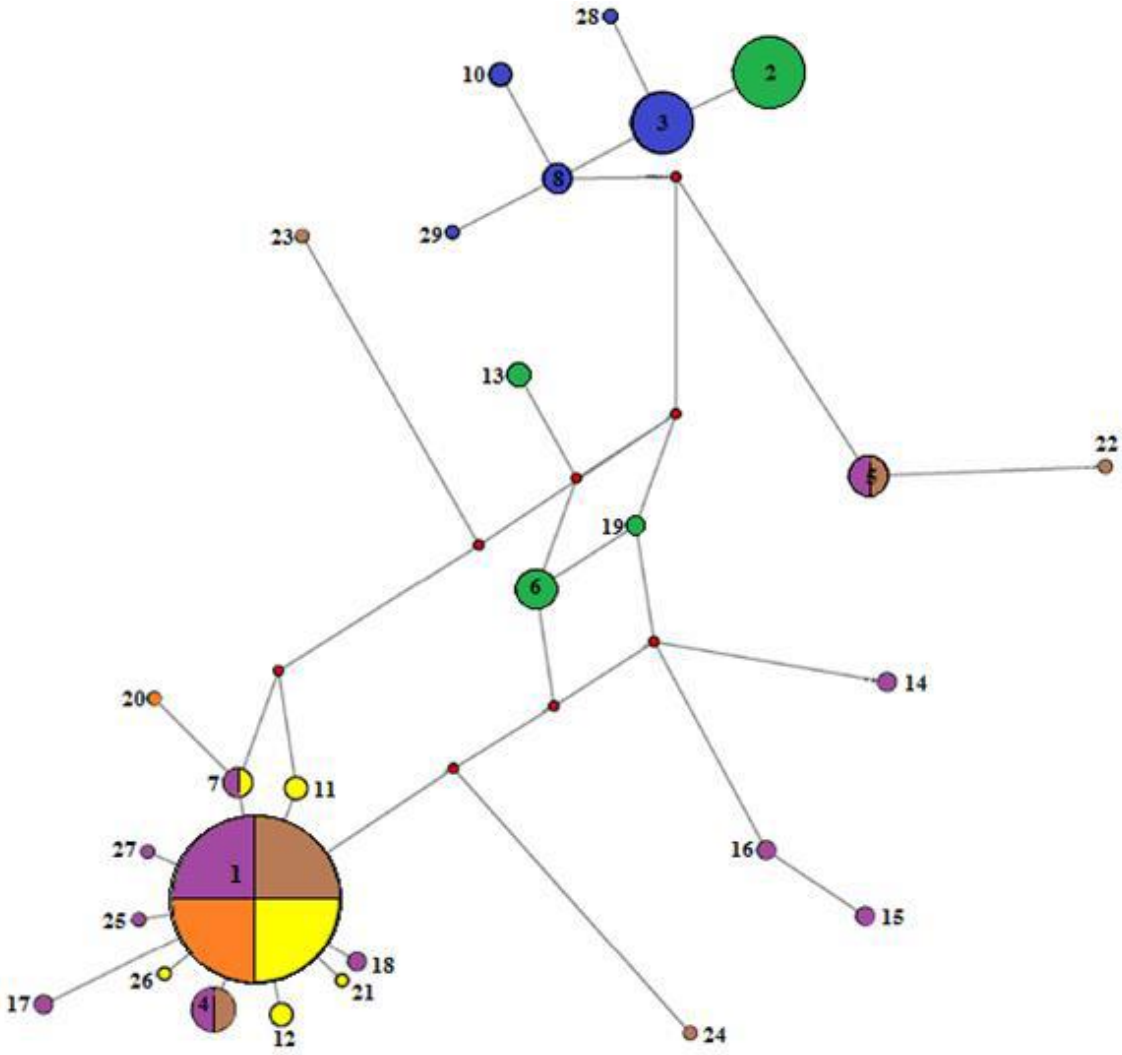

Figure 2.6. Median spanning network depicting the phylogenetic relationships between different haplotypes, drawn using the program Network. The size of each circle represents the haplotype frequency and length of each branch indicates mutational steps. Different colors represent different localities and regions: green British Columbia, blue Yukon, purple Northern, orange Central, brown Western, yellow Southern Labrador

of a species into the future. Here we set out to achieve an understanding of the post glacial dispersal routes with an assessment of genetic diversity and population structure of the longnose sucker in Labrador, a poorly-studied species with respect to evolutionary the longnose sucker in Labrador, a poorly-studied species with respect to evolutionary genetics but potentially a good model organism for freshwater fish species in northern Canada, and a species of regional conservation and management interest.

Mitochondrial DNA diversity in the longnose sucker

The three gene regions we were able to amplify each provides insight into levels of variability and informativeness of different mitochondrial genes. The CO1 gene region was the most conserved with 1.86% variable sites among haplotypes, while the ND4 gene region was the least conserved with 3.55%. Based on mitogenomic work, Anderson and colleagues (1982) determined that the COI gene had one of the highest level of conservation among species (bovine and human comparison; 79.4%) and the highest level of amino acid conservation (91.2%). Another study by Hebert and colleagues (2004) determined that the COI gene contained high levels of diversity between species, however a very small level of diversity within species. Based on this, it is unsurprising that many of our individuals of the same species contained identical COI fragments. Zardoya and Meyer (1996) described the performance of various mitochondrial protein-coding genes in vertebrates and their ability to infer accurate phylogenetic trees of a variety of species. According to these authors, both ND4 and CYT b ranked in the top set of genes as

'reliable evolutionary tracers' independent of their length, also consistent with the greater levels of variability we found in these genes among Labrador longnose suckers.

Including all three gene regions, we found a total of 20 haplotypes among 287 longnose suckers from Labrador with 11 found in the north, seven found in the southeast, six found in the west and two found in central Labrador. A previous study by McPhail and Taylor (1999) resolved 10 different haplotypes (five singletons) for the ND2 gene region among 22 individuals, as well as four haplotypes for the CYT b gene region among 95 individuals, in western Canada. Upon comparison to 10 CYT b haplotypes found in Labrador, this is a similar number of haplotypes per sample size indicating a similar rate of evolution within this gene. However we found 3.1% of sites to be variable among haplotypes while McPhail and Taylor (1999) determined 5.7% of sites were variable in the CYT b gene region. This lower level of variable sites at the CYT b gene region in Labrador could be attributed to the short amount of time these fish have been in Labrador when compared to longnose suckers in western Canada or may also imply more refugia represented in western Canada than the eastern range for longnose suckers. Nonetheless within the Yukon and British Columbia samples in our study, we have only 0.34% variable sites. Hence we likely do not have an adequate sample size (total of 75 among 3 lakes) to identify the variation that McPhail and Taylor observed. Furthermore, there are differences in the location of the lakes that were sampled; McPhail and Taylor sampled the lower portion of British Columbia and into Washington State whereas our samples are closer to the Yukon and more inland.

We were able to identify a main Labrador haplotype which was found in all lakes and all regions as well as several closely-related, less frequent haplotypes which were generally only located in one lake or region. This pattern is consistent with a recent colonization of these longnose suckers from one main refugium to Labrador. Low levels of nucleotide diversity and high haplotypic diversity also indicate a past bottleneck event followed by population expansion, consistent with the recent retreat of glacial ice from the region. The pairwise mismatch distribution, Tajima's D and Fu's Fs support this conclusion as they also suggest a population expansion. Negative values of Tajima's D and Fu's Fs indicate that our populations have not reached equilibrium since the re-colonization event. However, levels of nucleotide diversity were markedly higher in western and northern Labrador samples than central, which indicates the presence of divergent haplotypes in these regions, which likely indicate the presence of multiple refugia in these regions, albeit at minor levels.

Strong regional structure with evidence of longitudinal variation of longnose suckers in Canada

We conclude from the SAMOVA analysis of all 17 populations that there is strong genetic structure among regions in Canada; at the mostly likely K of 3, almost 60% of the sequence variation observed was between provinces. This was expected since no haplotypes were shared between Labrador, Yukon, and British Columbia. Significant variation primarily between provinces may indicate historical populations were similar

within each disparate geographic area but different between them. McPhail and Taylor (1999) also observed significant differentiation between eastern (Québec) and western (British Columbia) longnose suckers. However, they were reserved in this conclusion as they had small sample sizes. Our results point towards the acceptance of western and eastern (longitudinal) division of populations based on the level of differentiation between Labrador and Yukon/British Columbia at K=2. We cannot identify any strong northern-southern (latitudinal) shift in populations based on the genetic divergence between Yukon, British Columbia populations or northern and southeastern Labrador populations. This was expected given our small latitudinal difference between lakes, and our large longitudinal difference. Additional sampling locations in southern/central United States may clarify the possibility of a latitudinal shift in populations.

While the SAMOVA did not identify any structure within Labrador, pairwise F_{ST} among lakes and regions identified western Labrador lakes as being highly differentiated from the other Labrador regions and from many of the individual lakes. Taken together, these results suggest that the structure apparent in Labrador is at a regional rather than fine scale. This probably reflects the imprint of historical colonization and lack of equilibrium between gene flow and genetic drift as well as multiple refugial origins, but it may also suggest ongoing gene flow at a local scale among populations of suckers.

Evidence for post-glacial recolonization of Labrador from one major and two minor glacial refugia

Dillinger and colleagues (1991) suggested that longnose suckers in the Yukon originated from the Beringian and Mississippian refugia after the Wisconsinan glaciation while McPhail and Taylor (1999) concluded that longnose suckers in British Columbia survived in and recolonized from three refugia - Beringia, the Mississippian, and the Pacific. Based on our results of the SAMOVA of all haplotypes in the 17, we conclude that there likely is one main refugium with the possibility of small influence from 2 other refugia in western and northern Labrador. This main refugium is unlikely to encompass any of the above proposed refugia as there are Labrador lakes that are genetically distinct from all other localities (as seen in the median spanning network and in the lack of shared haplotypes between provinces). As proposed by Rempel and Smith (1999) there were an additional two refugia in which freshwater fishes were able to survive the Wisconsinan glaciation: the Nahanni and the Atlantic. Labrador is theorized to have been colonized by both the Mississippian and the Atlantic refugia (Black *et al.* 1986) so the Atlantic is likely the main Labrador refugial source. Once out of the Atlantic refugium, the longnose sucker likely traveled along east coast of the United States as the ice melted and up into the east coast of Canada and straight into southeastern Labrador with further dispersal across Labrador as ice continued to melt.

However, evidence from the measures of $F_{ST,}$ maximum likelihood phylogeny, and median spanning network supports the possibility of some influence of multiple refugial sources in both western and northern Labrador. Based on the presence of both Yukon and British Columbia in the same cluster (top right cluster of the median spanning network) we suggest that this represents the Beringian refugium. Although, within eastern

Canada it was thought that the main refugial origins were the Mississippian, the Atlantic, and the Acadian refugia (Taylor 2004), it is unlikely that fish from these refugia would be found in northern Canada based on freshwater dispersal routes and the retreat of the glacial ice during the Wisconsinan. Therefore, the only explanation to as the Labrador haplotypes clustering closely with the Yukon haplotypes (which compose one of the few highly supported clusters in the maximum likelihood phylogeny) would be from the Beringian refugium. British Columbia is a location where the Mississippian refugium played a large part in the reintroductions of fish when the glaciers receded after the Wisconsinan glaciation (Matthews 1998). This suggests that the cluster in the network containing British Columbia and northern and western Labrador samples (middle cluster of median spanning network) originated from a Mississippian refugium. Northern and western regions were likely re-colonized through a Québec route from the Mississippian and the Beringian refugia. However, these two additional proposed refugia only encompass a total of 6.3% of individuals from Labrador and should be regarded with caution as greater sample sizes from these regions may clarify the relative importance of these minor refugia.

Putative hybridization between longnose suckers and white suckers in southeastern Labrador

Habitat and spawning periods tend to overlap in many species including between the longnose and white suckers. A report by Nelson (1973) recorded the first occurrence

of wild hybridization between these two species in Alberta with a second report by Dion and colleagues (1994) confirming interspecific mating between these suckers in a Québec system. We have possibly uncovered the first documented case of hybrids in Labrador (the four fish with haplotype 9) however as mtDNA is maternally transmitted we can only see evidence of crosses between female white suckers and male longnose suckers and not the reverse. A previous study based on observation reported that white sucker males would readily spawn with longnose females (32% of the time) however the opposite was not observed (Dion *et al*. 1994). Dion and colleagues (1994) theorized that the white sucker males were not in an ideal spawning habitat and may have been displaced by the longnose sucker males who were in their ideal habitat, leading to no observation of mating between a longnose sucker female and white sucker male. The presence of this white sucker mtDNA in longnose suckers indicates that we may have documented the occurrence of spawning between male longnose and female white suckers, opposite to what Dion and colleagues (1994) observed. Using mtDNA we are unable to conclusively identify hybrid individuals, due to the possibility of an ancient hybridization event followed by continual backcrossing. It is also possible that we may have misidentified white sucker individuals as longnose suckers when the fish were caught as they can appear quite similar. In order to clarify these possibilities, investigation into the potential hybridization of longnose suckers with white suckers will be fully addressed in the following chapter with a panel of nuclear microsatellite markers as well as an additional analysis in order to help clarify refugial origins and population structure of longnose suckers in Labrador.

Conclusions and future directions

In conclusion, we find evidence for one major and two minor glacial refugial sources for longnose suckers in Labrador. Individuals from the Atlantic refugia appear to be present commonly throughout all of Labrador with several closely related singleton sequences. However some evidence in both western and northern Labrador, point towards the possibility of multiple refugial sources including the Mississippian and the Beringian. Little structure is present among populations within each region suggesting that populations are not in equilibrium due to the recent colonization, possibly hidden by ongoing local gene flow. Evidence of a recent population expansion supports the former. It will be important to investigate the population structure of Labrador suckers with a panel of nuclear markers such as microsatellites to more conclusively address levels of ongoing gene flow and refugial origins. The extent of any hybridization of longnose with white suckers requires further examination with nuclear markers. A more comprehensive sampling across North America and in Siberia would increase the confidence of refugial origin assessment and should be done in the future to refine our understanding of population structure and historical dispersal patterns of the longnose sucker throughout its range.

REFERENCES

Anderson, S., DeBruijn, M.H.L., Coulson, A.R., Eperon, I.C., Sanger, F., Young, I.G. (1982). Complete sequence of bovine mitochondrial DNA conserved features of the mammalian mitochondrial genome. *Journal of Molecular Biology*, 156, 683-717.

Bandelt, H.J., Forster, P., Röhl, A. (1999). Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution*, 16, 37-48.

Bernatchez, L. (2001). The evolutionary history of brown trout (*Salmo trutta L.*) inferred from phylogeographic, nested clade, and mismatch analysis of mitochondrial DNA variation. *Evolution*, 55, 351-379.

Black, G.A., Dempson, B., Bruce, W.J. (1986). Distribution and post-glacial dispersal of freshwater fishes of Labrador. *Canadian Journal of Zoology*, 64, 21-31.

Curole, J.P., & Kocher, T.D. (1999). Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends in Ecology and Evolution*, 14, 394-398.

Dillinger, R.E., Birt, T.P., Green, J.M., Davidson, W.S. (1991). Post-glacial dispersal of longnose suckers, *Catostomus catostomus*, in the Mackenzie and Yukon drainages. *Biochemical Systematics and Ecology*, 19, 651-657.

Dion, R., Richardson, M., Roy, L., Whoriskey, F.G. (1994). Spawning patterns and interspecific matings of sympatric white (*Catostomus commersoni*) and longnose (*C. catostomus*) suckers from the Gouin reservoir system, Quebec. *Canadian Journal of Zoology*, 72, 195-200.

Doosey, M.H., Bart, H.L., Saitoh, K., Miya, M. (2010). Phylogenetic relationships of catostomid fishes (Actinopterygii: Cypriniformes) based on mitochondrial ND4/ND5 gene sequences. *Molecular Phylogenetics and Evolution*, 54, 1028-1034.

Dupanloup, I., Schneider, S., Excoffier, L. (2002). A simulated annealing approach to define the genetic structure of populations. *Molecular Ecology*, 11, 2571-2581.

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

Fu, Y.X. (1997). Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics*, 147, 915-925.

Harpending, H.C. (1994). Signature of ancient population growth in a low-resolution mitochondrial DNA mismatch distribution. *Human Biology*, 591-600.

Harris, P.M., Mayden, R.L., Espinosa Pérez, H.S., Garcia de Leon, F. (2002). Phylogenetic relationships of *Moxostoma* and *Scartomyzon* (Catostomidae) based on mitochondrial cytochrome b sequence data. *Journal of Fish Biology*, 61, 1433-1452.

Hasegawa, M., Kishino, H., Yano, T. (1985). Dating of the human-ape splitting by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution*, 22, 160-174.

Hebert, P.D.N., Stoeckle, M.Y., Zemlak, T.S., Francis, C.M. (2004). Identification of birds through DNA barcodes. *PLoS biology*, 2, e312.

Hubbs, C.L. (1955). Hybridization between fish species in nature. *Systematic Zoology*, 4, 1-20.

Hurst, C.D., Bartlett, S.E., Davidson, W.S., Bruce, I.J. (1999). The complete mitochondrial DNA sequence of the Atlantic salmon, *Salmo salar*. *Gene*, 239, 237-242.

Inoue, J.G., Miya, M., Tsukamoto, K., Nishida, M. (2000). Complete mitochondrial DNA sequence of the Japanese sardine *Sardinops melanostictus*. *Fisheries Science*, 66, 924-932.

Langille, B.L., Perry, R., Keefe, D., Marshall, H.D. (2014). Novel mitochondrial DNA primers for identification of population trends in longnose suckers (*Catostomus catostomus*) and multispecies identification. *Conservation Genetics Resources*, online access.

Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25, 1451-1452.

Matthews, W.J. (1998). *Patterns in freshwater fish ecology*. Chapman and Hall, New York.

McDonald, J.H., & Kreitman, M. (1991). Adaptive protein evolution at the A*dh* locus in *Drosophila*. *Nature*, 351, 652-354.

McDonald, D.B., Parchman, T.L., Bower, M.R., Hubert, W.A., Rahel, F.J. (2008). An introduced and a native vertebrate hybridize to form a genetic bridge to a second native species. *Proceedings of the National Academy of Sciences*, 105, 10837-10842.

McGlashan, D.J., & Hughes, J.M. (2000). Reconciling patterns of genetic variation with stream structure, earth history and biology in the Australian freshwater fish *Craterocephalus stercusmuscarum* (Atherinidae). *Molecular Ecology*, 9, 1737-1751.

McPhail, J.D., & Taylor, E.B. (1999). Morphological and genetic variation in northwestern longnose suckers, *Catostomus catostomus*: The salish sucker problem. *American Society of Ichthyologists and Herpetologists*, 1999, 884-893.

Millette, K.L., Xu, S., Witt, J.D.S., Cristescu, M.E. (2011). Pleistocene-driven diversification in freshwater zooplankton: Genetic patterns of refugial isolation and post-glacial recolonization in *Leptodora kindtii* (Crustacea, Cladocera). *Limnology and Oceanography*, 56, 1725-1736.

Minin, V., Abdo, Z., Joyce, P., Sullivan, J. (2003) Performance-based selection of likelihood models for phylogeny estimate. *Systematic Biology*, 52, 674-683.

Nabholz, B., Mauffrey, J.F., Bazin, E., Galtier, N., Glemin, S. (2008). Determination of mitochondrial genetic diversity in mammals. *Genetics*, 178, 351-361.

Nei, M., & Gojobori, T. (1986). Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Molecular Biology and Evolution*, 3, 418-426.

Nelson, J.S. (1973). Occurrence of hybrids between longnose sucker (*Catostomus catostomus*) and white sucker (*Catostomus commersoni*) in Upper Kananaskis Reservoir, Alberta. *Journal of the Fisheries Research Board of Canada*, 30, 557-560.

Perry, R.C., & Casselman, J.M. (2012). Comparisons of precision and bias with two age interpretation techniques for opercular bones of longnose sucker, a long-lived northern fish. *North American Journal of Fisheries Management*, 32, 790-795.

Proctor, M.F., McLellan, B.N., Strobeck, C., Barclay, R.M.R. (2004). Gender-specific dispersal distances of grizzly bears estimated by genetic analysis. *Canadian Journal of Zoology*, 82, 1108-1118.

Rempel, L.L., & Smith, D.G. (1998). Post-glacial fish dispersal from the Mississippi refuge to the Mackenzie River basin. *Canadian Journal of Fisheries and Aquatic Sciences*, 55, 893-899.

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

Saitoh, K., Sado, T., Mayden, R.L., Hanzawa, N., Nakamura, K., Nishida, M., Miya, M. (2006). Mitogenomic evolution and interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): The first evidence toward resolution of higher-level relationships of the world's largest freshwater fish clade based on 59 whole mitogenome sequences. *Journal of Molecular Evolution*, 63, 826-841.

Smith, G.R., & Koehn, R.K. (1971). Phenetic and cladistics studies of biochemical and morphological characteristics of *Catostomus*. *Systematic Zoology*, 20, 282- 297.

Tajima, F. (1989). Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, 123, 585-595.

Tamura, K., Dudley, J., Nei, M., Kumar, S. (2007). MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, 24, 1596-1599.

Taylor, E.B. (2004) An analysis of homogenization and differentiation of Canadian freshwater fish faunas with an emphasis on British Columbia. *Canadian Journal of Fisheries and Aquatic Sciences*, 6, 68-79.

Weider, L.J., & Hobaek, A. (1997). Post-glacial dispersal, glacial refugia, and clonal structure in Russian/Siberian populations of the arctic *Daphnia pulex* complex. *Heredity*, 78, 363-372.

Wilson, G.M., Thomas, W.K., Beckenbach, A.T. (1985). Intra- and inter-specific mitochondrial DNA sequence divergence in *Salmo*: rainbow, steelhead, and cutthroat trouts. *Canadian Journal of Zoology*, 63, 2088-2094.

Young, R.R., Burns, J.A., Smith, D.G., Arnold, L.D., Rains, R.B. (1994). A single, late Wisconsin, Laurentide glaciation, Edmonton area and southwestern Alberta. *Geology*, 22, 683-686.

Zardoya, R., Garrido-Pertierra, A., Bautista, J.M. (1995). The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss*,41, 942-951.

Zardoya, R., & Meyer, A. (1996). Phylogenetic performance of mitochondrial protein-coding genes in resolving relationships among vertebrates. *Molecular Biology and Evolution*, 13, 933-942.

CHAPTER 3: HYBRIDIZATION, GENE FLOW AND POST-GLACIAL RECOLONIZATION OF LONGNOSE SUCKERS IN LABRADOR INFERRED WITH MICROSATELLITE MARKERS

Abstract

Hybridization among catostomid species has been documented since the late 1940s and continues to be an active area of research as the incidence of hybrid species continues to increase. The longnose sucker (*Catostomus catostomus*) and white sucker (*C. commersonii*) produce viable offspring throughout many regions of North America, threatening the biodiversity and fitness of the parental species in those areas. With a set of six microsatellite markers we sought to document the occurrence of hybridization between longnose and white suckers in northern, western, central, and southeastern Labrador. We also assessed refugial origins and fine scale population structure of longnose suckers in Labrador and compared them to previous mitochondrial DNA analysis. Among longnose suckers, 1.32% (STRUCTURE and NewHybrids) contained admixed ancestry with all located in southeastern Labrador. These levels of hybridization are lower to those found in catostomids in other studies, and are unlikely related to the recent construction of the Trans-Labrador Highway. STRUCTURE analysis also identified four genetic clusters among 302 longnose and 14 white sucker individuals: a southeast/central group, a western/northern group, a distinct northern group, and a white sucker group. Results from the STRUCTURE analysis therefore suggest the existence of three refugial origins (Atlantic, Beringian, and Mississippian) also found with mtDNA,

while pairwise F_{ST} measures indicated significant differentiation among most lakes, supporting a pattern of limited gene flow both among and within regions.

INTRODUCTION

Natural hybridization in wild fish species was recorded based on morphological variation long before genetic techniques became readily available in the early 1970s. Prior to Carl Hubbs' report (1955) on the extensive occurrence of fish hybridization, it was thought that fish species never hybridized in nature. Hubbs (1955) was able to successfully show that the opposite was more common; fish species hybridize in nature with the frequency of hybridization following a gradual decrease from freshwater to the sea as factors of isolation diminish. Miller and colleagues (1989) suggested that hybridization accounted for at least 38% of fish extinctions in the USA. They hypothesized that the actual percentage may be even higher, as hybrid morphology, which was the main determinant of hybrids, is often not readily apparent (Perry *et al*. 2002).

The suckers (Cypriniformes; Catostomidae) are of particular interest with respect to hybridization as they appear to hybridize freely, with the ability to produce viable offspring (Ferris 1984). Catostomids share several distinguishing characteristics, most obviously the distinctive, fleshy lips from which their name is derived, found on the underside of the head, as well as a row of teeth along the pharyngeal bone found in the throat (Banister 1998). Hubbs (1955) concluded that in the Western USA at least one in every hundred suckers was an interspecific hybrid. High rates of mating between different species of suckers were observed under specific circumstances which may explain the high level of hybridization. For example, extensive hybridization often occurred when there was a low population size of one sucker species relative to another

sucker or if a non-native sucker was introduced (Hubbs 1955). Altered or crowded breeding grounds may also have played a role in the increased incidence of hybrids, as these would cause confusion and lead to an increased chance of accidental fertilization (Nelson 1968). The ability of these species to readily interbreed is detrimental to the genetic integrity of all species involved as it can cause outbreeding depression, which is when offspring from two separate populations have lower fitness when compared to the offspring from same populations (Gilk *et al*. 2004; Pritchard *et al*. 2007).

White suckers (*Catostomus commersonii*) have been documented to act as a "genetic bridge" between other pairs of sucker species as they will readily spawn and produce viable offspring with the other suckers (McDonald *et al*. 2008). Within a controlled experiment, McDonald and colleagues (2008) determined that white suckers formed hybrids with both the flannel mouth sucker (*Catostomus latipinnis*) and the blue sucker (*Catostomus discobolus*), and these F_1 hybrids in turn spawned with a different sucker species to form three-way hybrids. Sucker hybridization has also been documented between the June sucker (*Chasmistes liorus*) and the Utah sucker (*Catostomus ardens*) where hybrids appear as morphological intermediates (Cardall *et al.* 2007). Low numbers of June sucker available for spawning apparently led to the choice to spawn with readily available Utah suckers resulting in morphologically distinct populations of June sucker from the original parent species of previous years (Miller *et al*. 1981).

Longnose suckers are easily distinguishable from all other Catostomidae by their long protruding snout, small scales, and inferior mouth (Nelson 1973; Smith & Koehn 1971). The longnose sucker (*Catostomus catostomus*) has a geographically wide

distribution ranging from North America to Siberia (Doosey *et al*. 2010) and extending into the Arctic (Page & Burr 1991) and can be found in either freshwater or brackish water habitats (McPhail & Taylor 1999). Hybridization between the longnose sucker and white sucker, also distributed across North America, was first recorded in 1973, using morphological data and scale counts to describe the hybrid species as an intermediate between the two parental species (Nelson 1973). Mating of the longnose and white sucker was likely due to a geographic disturbance or an introduction event and the sustained occurrence of hybridization can be attributed to the substantial, stable population size of the white sucker.

In the case of identifying hybrids in populations of fish with very little observable phenotypic differences between parental species, genetic markers are a useful tool. As mitochondrial DNA (mtDNA), a traditional population genetic marker is maternally inherited, it has limited use in this context (Roy *et al*. 1994). MtDNA can indicate the existence of hybridization within a population but only with respect to the maternal parental species, and not whether the hybridization is ongoing or historical. Microsatellites are tandem repeats of short simple sequences (e.g. $(CA)_n$), found commonly in eukaryotic genomes. High mutation rates, likely due to replication slippage, lead to high allelic diversity (Roy *et al*. 1994; Whittaker *et al*. 2003). This slippage occurs when the strands of DNA realign incorrectly after replication, leading to insertions or deletions of the repeat motif, more often in longer microsatellite loci (Whittaker *et al*. 2003). Inherited biparentally, microsatellites provide a co-dominant marker system that can assess heterozygosity, allelic richness, relatedness, assortative mating, and

geographical patterns of gene flow (Cardall *et al*. 2006), among others. Microsatellites can also determine the direction and extent of hybridization and backcrossing to parental species (Roy *et al.* 1994).

The extensive glacial history of North America has played a major role in present-day distributions and dispersal routes of freshwater fish (Millette *et al*. 2011) as glacial cycles forced freshwater fish to expand and contract their geographic range. Rempel and Smith (1998) determined that the Wisconsinan glaciation period (~75 to 10 thousand years ago) had the greatest impact on fish fauna as the Laurentide ice sheet extended furthest and identified five main refugial origins of freshwater fish; the Beringian, Mississippian, Pacific, Nahanni, and Atlantic. It was hypothesized by McPhail and Taylor (1999) that longnose suckers survived the Wisconsinan glaciation in three main refugia: the Beringian, the Pacific, and the Mississippian. We found evidence for one major and 2 minor refugial sources of longnose suckers in Labrador based on mitochondrial DNA analysis, of which the Beringian and the Mississippian overlap with the refugia proposed by McPhail and Taylor (1999) while most common refugial source in Labrador was the Atlantic (Chapter 2).

We also identified longnose suckers with white sucker mtDNA in Chapter 2 which may represent putative hybrids. Labrador provides a unique landscape within which to study this species as many locations have remained relatively undisturbed by human involvement and with overlapping habitat of both the longnose and the white sucker. As part of a long-term monitoring plan, the provincial Department of Environment and Conservation is interested in documenting genetic diversity, population

structure and hybridization of this species. Here we utilize six microsatellite makers to determine the extent of hybridization between the longnose sucker and the geographically abundant white sucker, and to further investigate refugial origins as well as assess fine-scale pattern of gene flow among longnose suckers in Labrador.

MATERIALS AND METHODS

Study area and sample collection

Caudal fin clippings (~200 mm²) were taken from longnose suckers captured via gill netting by Newfoundland and Labrador provincial wildlife biologists between the years 2003 and 2012. A total of 424 samples were collected from 14 lakes throughout Labrador. These were divided into four regions as used by the Wildlife Division of the provincial Department of Environment and Conservation in order to document regional variation: northern, central, western and southeastern (Figure 3.1). A total of 30 white sucker samples were collected from Brettney Lake in southeastern Labrador via gill netting by Newfoundland and Labrador provincial wildlife biologists in 2013. All fish fins were stored at -20°C until needed.

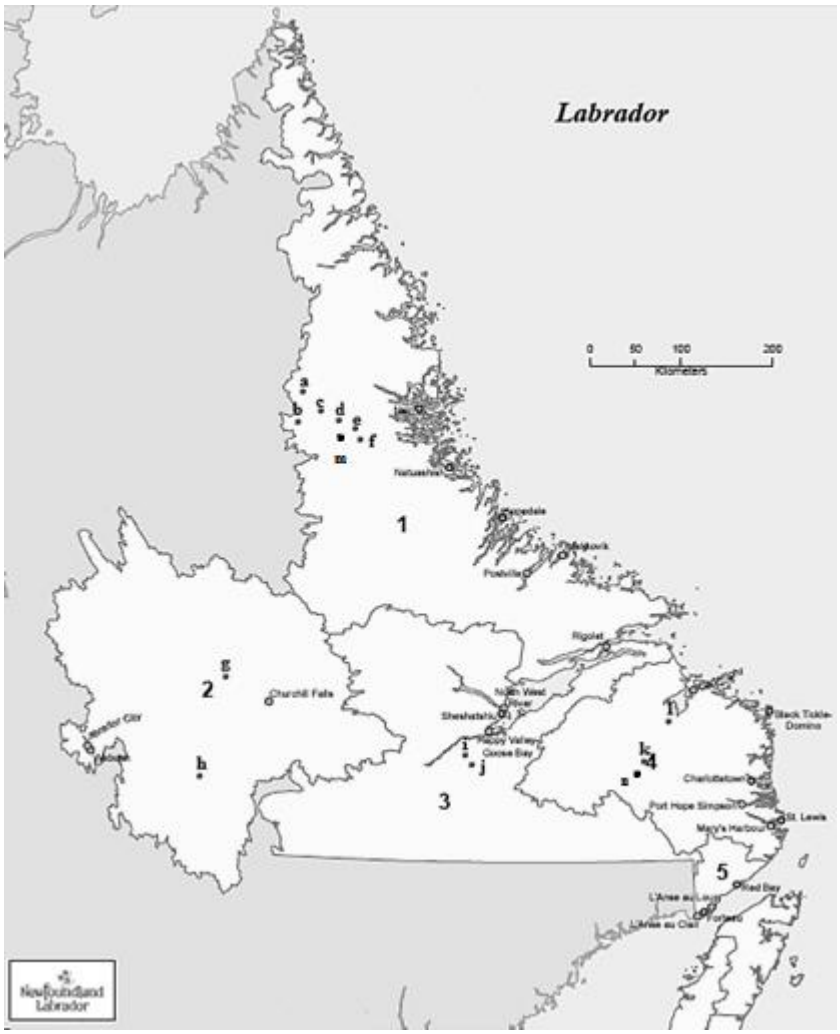

Figure 3.1. Geographic locations of the 14 longnose sucker populations investigated in Labrador with numbers 1 through 4 indicating each defined region (1, Northern Labrador; 2, Western Labrador; 3, Central Labrador; and 4, Southeastern Labrador). Exact coordinates for locations Ikadlivik Brook Watershed in Northern Labrador and Birchy Lake in Southeastern Labrador are not known and are shown on the map with initials m and n respectively; locality abbreviations are: a Genetics H Lake, b Strange Lake, c Mistake Lake, d Walkabout Lake, e Konrad Lake, f Cabot Lake, g Lobstick Lake, h Lac Joe, i The Right Lake, j Mercier Lake, k Brettney Lake, l Crooked Lake

Microsatellite analysis

Genomic DNA was extracted from a portion (~4 mm^2) of the caudal fin clipping using the QiagenQIAamp®DNA Mini Kit (Qiagen Inc., Toronto, Canada) according to the manufacturer's protocol detailing tissue sample extraction.

A total of ten microsatellite loci were initially chosen for study. Four (Xte4, Xte6, Xte7, and Xte11) were from Turner *et al.* (2009) who described these primers for a variety of species including the white sucker (*Catostomus commersonii*) among others. Three microsatellite markers (US3, US4, and US9) were chosen from Cardall *et al.* (2006) who developed their primers for use in landscape studies of speciation and gene flow between sucker species, with emphasis on hybridization and conservation. Finally, three (Dlu409, Dlu4243 and Dlu4283) were chosen from Tranah *et al.* (2001) as they were designed to work in a variety of different suckers species (*Catostomus snyderi*, *Catostomus rimiculus*, *Chasmistesbrevirostris*, *Deltistesluxatus*, and *Catostomus occidental)*. Six loci of the original ten amplified well and were used for subsequent analysis with the fluorescently labelled forward primers (Table 3.1); Xte4, Xte7, Xte11, Dlu409, Dlu4283, and Dlu4243.

PCRs were performed in an Applied Biosystems GeneAmp PCR System 9700 Thermal Cycler (Applied Biosystems Inc., Foster City, USA) and contained 1X PCR Master Mix (Promega Corp., city, USA), 1μL 10 mM of each primer (0.5μL10 mM for primers Dlu409 and Dlu4243) and 2 μL genomic DNA (2– 570 ng). The amplification profile for Turner and colleagues (2009) primers was altered for amplification in longnose suckers and consisted of an initial denaturation for 2 min at 94℃, followed by 35 cycles

Table 3.1. Primer sequences and PCR annealing temperatures for six microsatellite loci, originally developed for *Xyrauchentexanus* (razorback sucker), *Chasmistesliorus* (June sucker), *Catostomus ardens* (Utah sucker) and *Deltistes luxatus* (lost river sucker). Abbreviations: F, forward primer; R, reverse primer; T_A, annealing temperature

Locus	Repeat motif	Dye label	Primer sequence (5'-3')	Fragment length (bp)	$MgCl_2$ (mM)	T_A (°C)
Xte4[a]	AC	PET	F:GGATTGCCTTTATGGTGTCT R:TTCTCTTCAACTGGTCTAAAT	320	2.5	50
Xte7[a]	TC	NED	F:GGAATAATGGTAGAGAAGAACG R:TAATAATGGAAAGAGGGTGAGG	143	2.5	52
Xte11[b]	TG	6-Fam	F:CCACTATAGGGATTACAAAA R:CACCTGAGCAACACACCTT	296	2.5	52
Dlu409[c]	GATA	VIC	F:TGCGATCCTAGAAGGAGTAAAACA R:ATTCCATTTGCTGTCAACTTCAAA	206	2	52
Dlu4243[c]	GATA	6-Fam	F:TGGTTGGATGCTGAAATAAAGTAA R:TGAGCCTCATCATAGATGGATAGA	160	2	57
Dlu4283[c]	GATA	6-Fam	F:CTGAAAGCACCTCCTCCATTAG R:GTTCTCTTCTCCTGTTTCGCTTAT	107	2	57

[a] Indicates primer sequences previously reported in McPhee and Turner (2004)
[b] Indicates primer sequences previously reported in Turner *et al.* (2009)
[c] Indicates primer sequences previously reported in Tranah *et al*. (2001)

of 94℃ for 30 sec, 50- 52℃ for 30 sec and 72℃ for 30 sec, with a final extension at 72℃ for 5 min. Locus Dlu4243 was amplified using an initial denaturation for 2 min at 94℃, followed by 30 cycles of 94℃ for 60 sec, 52- 57℃ for 30 sec and 72℃ for 60 sec, with a final extension at 72℃ for 15 min (revised from Cardall *et al*. 2006). Locus Dlu409 was amplified using an initial denaturation for 2 min at 94℃, followed by 30 cycles of 94℃ for 30 sec, 52℃ for 30 sec and 72℃ for 60 sec, with a final extension at 72℃ for 5 min (Table 3.1). Prior to electrophoresis, microsatellite amplicons were diluted according to concentration in order to be electrophoresed together and 1 μL of combined microsatellites were added to 8.9 μL of formamide and 0.1μL of the LIZ-500 size standard (Applied Biosystems Inc., Foster City, USA). Loci without overlapping size ranges were paired for electrophoresis as follows; Xte4 and Xte11; Xte7; and Dlu409 andDlu4243 were electrophoresed using GeneScan Software on the ABI Prism 3730 DNA Analyzer and alleles were scored using Peak Scanner 2 (Applied Biosystems Inc., Foster City, USA).

Microsatellite alleles and per-locus measures

Measures of allele size ranges and frequency were recorded for each locus. Null alleles, large allele dropout, or PCR artifacts that may have affected subsequent analysis were checked using MICRO-CHECKER (Van Oosterhout *et al*. 2004). Diversity measures including number of alleles (N_A), and allelic richness (A) were computed in ARLEQUIN v.3.11 (Excoffier *et al.* 2005) for each locus. Genotypic distributions were tested for deviation from Hardy-Weinberg equilibrium using Fisher's exact tests (Guo &

Thompson 1992) and linkage disequilibrium, which is a test of non-random association of alleles at different loci, was tested for all samples in ARLEQUIN.

Hybridization analysis

In order to identify hybrid individuals STRUCTURE was employed, which implements a Bayesian clustering algorithm with Markov Chain Monte Carlo (MCMC) randomization to determine the appropriate number of genetic clusters among the data, within minimal linkage or Hardy Weinberg differentiation within clusters (Pritchard *et al.* 2000). The data were assessed using an ancestry model which assumes admixture of populations with correlated allele frequencies and without *a priori* population information (Falush *et al*. 2003). The number of *K*, which is the number of clusters that best fit the data was estimated by the log likelihood of the data given the number of clusters (lnP(D); Pritchard *et al*. 2007). An estimate of admixture proportion (*Q*) was generated for each individual genotype by three independent runs for consistency using 1,000,000 final replicates (after a 500,000 replicate burn-in) with assumed correlated allele frequencies. We used threshold values of $0.1 < Q < 0.9$ in order to identify recent hybridization since this is the low end of the range for the detection of hybrids (Aboim *et al.* 2010).

As a complementary approach to the admixture analysis, a second analysis was performed using the software NewHybrids (Anderson & Thompson 2002). This Bayesian statistical method estimates the probability that each individual segregates into separate parental or hybrid genotype class. There are six genotypic classes which are possible after two generations of hybridization: two parental types (P_0 and P_1), 1st-generation hybrids (F_1), 2nd-generation hybrids (F_2), and backcrosses of F_1 with each of the parental species

(B_0 and B_1). The number of hybrids was estimated by 10 independent replicates for consistency using 100,000 final replicates (with a 50,000 replicate burn-in) and default settings of Jeffreys' distributed prior (Jeffreys 1946) which is based on Fisher information of non-informative priors (Bernardo & Smith 1994).

Population genetic analysis and structure

Diversity measures including number of alleles (N_A), allelic richness (A), and observed and expected heterozygosity (H_O and H_E) were computed in ARLEQUIN for each population. As a measure of inbreeding, F_{IS} (Weir & Cockerham 1984) was calculated for each population using ARLEQUIN.

The level of fine-scale or recent dispersal of longnose suckers among lakes could not be clearly elucidated from the mitochondrial DNA analysis in Chapter 2 so we used the multi-locus microsatellite data obtained here to further investigate the level of structure among our longnose sucker sampling sites. Pairwise F_{ST} values were evaluated, using differences between alleles, for each population (Weir & Cockerham 1984) with 1,000 permutations and a significance level of $\alpha = 0.05$. Additionally, STRUCTURE 2.3.4 (Pritchard *et al*. 2000) was used to investigate the most likely number of genetic clusters among all individuals and describe the distribution of these clusters among populations and regions. The data was assessed using an ancestry model which assumes admixture of populations with correlated allele frequencies and without *a priori* population information (Falush *et al*. 2003). Independent runs of K were tested between 1 and 20 with 1,000,000 iterations for each (after a 500,000 burn-in). The number of *K* that best fit the data was estimated by the log likelihood of the data given the number of

clusters (lnP(D); Pritchard *et al*. 2007). A STRCTURE analysis was also run with white sucker individuals removed and again with just northern and western individuals using the same parameters as above in order to identify all groups present in Labrador.

RESULTS

Loci and alleles

A total of six microsatellite loci were measured for 302 longnose sucker samples and 14 white sucker samples from Labrador. All loci were polymorphic in the longnose sucker populations and all loci save one were polymorphic in the white sucker samples. All populations were evaluated in MICRO-CHECKER individually, resulting in no evidence for scoring error due to stuttering, no evidence for large allele dropout, and no evidence for null alleles for all loci except locus Dlu4283 in 2 lakes (Lobstick Lake and Cabot Lake). As part of MICRO-CHECKER software, it not only identifies but can also correct various genotyping errors including null alleles for use in further analyses by adjusting the observed allele and genotype frequencies (Van Oosterhout *et al*. 2004). Therefore adjusted genotypes were used for these lakes at locus Dlu4283. There was an average of N_A= 17.7 alleles and N_A= 5.0 alleles in the longnose and white sucker species, respectively, across all loci. The longnose sucker alleles per locus ranged between N_A= 7 and N_A= 32 while the white sucker ranged between N_A= 1 and N_A= 13 (Table 3.2). The tests for linkage disequilibrium (LD) determined significant LD for only four out of 225

Table 3.2. Measures of diversity of six microsatellite loci across 302 longnose suckers (*Catotstomus catostomus*) and 14 white suckers (*C. commersonii*), including number of observed alleles (N_A), and allele size range (bp)

Longnose sucker		
Locus	N_A	size (bp)
Xte4	32	334-402
Xte7	7	91-233
Xte11	11	270-295
Dlu409	26	162-298
Dlu4243	7	80-118
Dlu4283	23	133-301
White sucker		
Locus	N_A	size (bp)
Xte4	2	314-326
Xte7	1	146
Xte11	2	295-302
Dlu409	13	278-343
Dlu4243	10	171-200
Dlu4283	2	295-301

pairwise comparisons, after Bonferroni correction for multiple comparisons, suggesting that there are no major departures from linkage equilibrium.

Bayesian assessment of hybrid individuals

To identify the most likely number of genetic clusters (K) among all longnose suckers in Labrador, a plot of the log likelihood values obtained from the STRUCTURE analysis was generated for *K*=1 to *K*=20 (Figure 3.2). The plot indicates a general increase from *K*=1 to K=4 after which it decreases from *K*=5 to K=20. Specifically, there is a large increase from *K*=1 to *K*=3 followed by a small increase to *K*= 4. A plateau in seen from *K*=4 to *K*=7 followed by a large decrease from *K*=8 to *K*=20 (Figure 3.2). *K*=4 was therefore selected as the most likely number of distinct genetic clusters, as recommended in the STRUCTURE manual (Pritchard *et al*. 2010).

Using STRUCTURE set to *K*=4, we identified all hybrid individuals based on the presence of the white sucker in the longnose sample set with $0.1 < Q < 0.9$. This analysis identified four admixed individuals of varying degrees (found in the blue; Figure 3.3). These four individuals were from Birchy Lake in southeastern Labrador and all contained longnose sucker mtDNA, determined in the previous chapter. There were a total of 1.32% admixed individuals out of 302 longnose suckers. We were also able to identify three individuals with complete white sucker microsatellites as they appear as complete blue bars in the plot (Figure 3.3) which were three of the four white sucker mtDNA sequences found in the previous chapter.

The NewHybrids analysis identified the same individuals as were found with STRUCTURE. All four of these individuals were in the F_2 class of hybrids while F_1

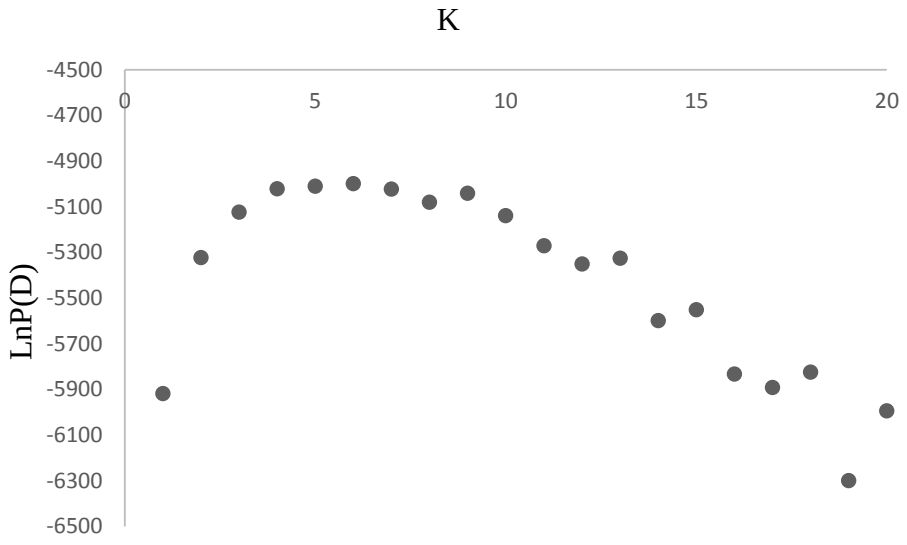

Figure 3.2. A plot of the estimated probability of K (LnP(D)) from the STRUCTURE analysis, for *K*=1-20

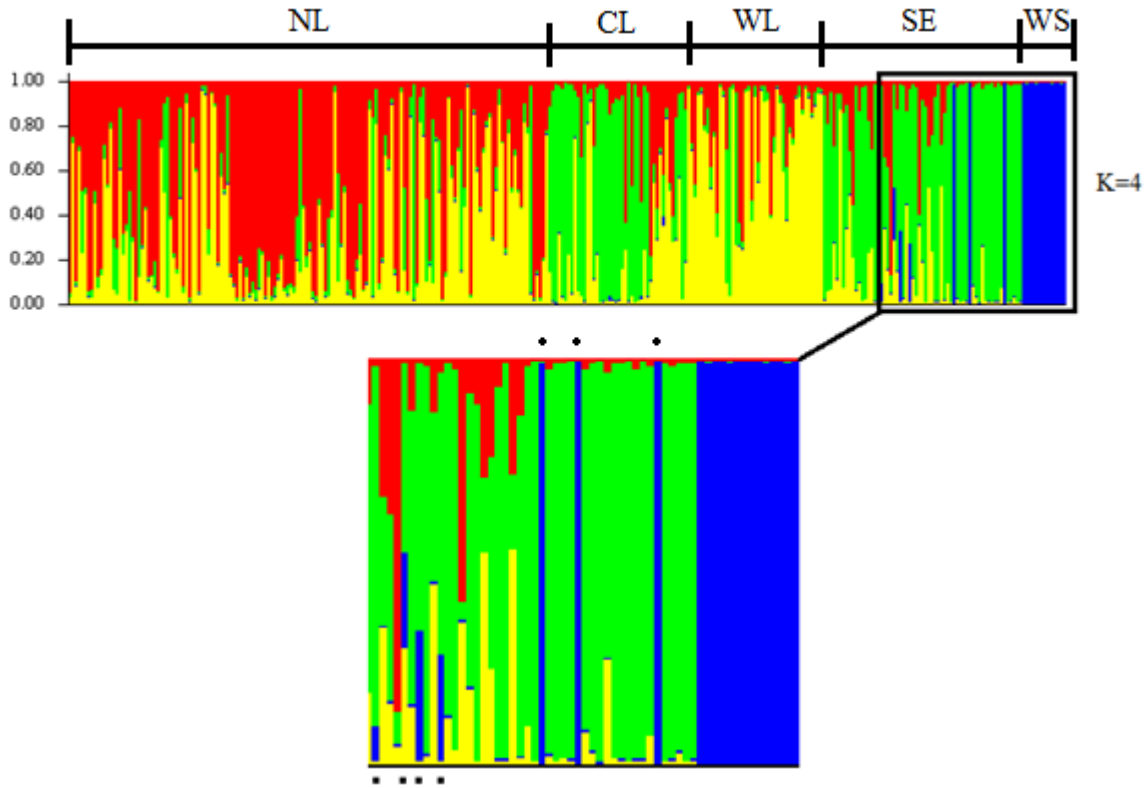

Figure 3.3. STRUCTURE plot of hybridization between longnose sucker (red, green, yellow) and white sucker (represented in blue) run at *K*=4. Each individual is represented as a vertical bar, whose length is proportional to the membership into each K cluster. Hybrid individuals are indicated by a black dot below the bar in the cut-out and potentially misidentified individuals are indicated by a black dot above the bar in the cut-out. Region labels are: NL, Northern Labrador; WL, Western Labrador; CL, Central Labrador; SE, Southeastern Labrador.WS refers to white sucker

Table 3.3. Assortment of the 302 longnose suckers into categories of parental longnose sucker (LS Parental), F_2 hybrid, or Mislabelled, l as determined using NewHybrids, based on six microsatellite loci. *N* is the sample size; LS Parentals are fish identified as longnose suckers with longnose sucker microsatellite genotypes; Mislabelled are fish identified as longnose suckers phenotypically but with white sucker microsatellite and mtDNA.

Labrador region	*N*	LS Parental	F_2	Mislabelled white suckers
Northern	152	152	0	0
Western	54	54	0	0
Central	34	34	0	0
Southeastern	62	55	4	3

hybrids were not present (Table 3.3). There were no individuals belonging to the B_0 class (hybrid backcross with white sucker) and no individuals belonging to B_1 (hybrid backcross with longnose sucker). A total of 1.32% of the 302 fish samples were hybrids between longnose and white suckers (Table 3.3). NewHybrids identified the same three individuals as the STRUCTURE analysis as belonging to the white sucker group.

Population and regional diversity and structure

Microsatellite diversity was characterized regionally and in individual lakes. Allelic richness was highest in northern and western Labrador with A= 12.5 and A=11.8 respectively (Table 3.4). Central Labrador contained the lowest allelic richness with A= 6.83, however when divided by the sample size, western Labrador was highest while northern Labrador was lowest. The allelic richness ranged from A= 4.40 in Genetics H Lake (northern Labrador) to A= 9.33 in Lac Joe (western Labrador). Averaged over all loci, central and southeastern Labrador had the lowest expected heterozygosity H_E= 0.514 and H_E= 0.507 respectively, while western Labrador has the highest H_E= 0.601 (Table 3.4), ranging from H_E= 0.389 in Konrad Lake (northern Labrador) and H_E= 0.676 in Genetics H Lake (northern Labrador). There were no populations of longnose suckers with observed heterozygosity significantly different from the expected heterozygosity and no F_{IS} were significantly different from zero (Table 3.5).

Pairwise F_{ST} values among lakes (Table 3.6), which measure population differentiation, indicated almost all pairwise lake comparisons to be statistically significant after Bonferroni correction for multiple comparisons. Most of the non-significant comparisons were between northern Labrador lakes which means they are not

Table 3.4. Sample size (*N*), observed (H_O) and expected (H_E) heterozygosity, and allelic richness (*A*) of longnose suckers (total of 299 individuals without the three individuals with white sucker microsatellites) across six microsatellite loci for 14 lakes in four regions of Labrador. Regional numbers are shown in bold type.

Region	Lake	*N*	H_O	H_E	*A*
Northern Labrador		**152**	**0.491**	**0.545**	**12.5**
	Walkabout Lake	30	0.544	0.528	6.50
	Mistake Lake	26	0.461	0.521	7.83
	Strange Lake	21	0.574	0.538	6.67
	Konrad Lake	24	0.395	0.444	5.00
	Ikadlivik Brook	25	0.528	0.478	7.60
	Cabot Lake	21	0.533	0.581	7.33
	Genetics H	5	0.640	0.560	4.40
Western Labrador		**54**	**0.562**	**0.601**	**11.8**
	Lac Joe	28	0.538	0.546	9.33
	Lobstick Lake	26	0.568	0.632	8.67
Central Labrador		**34**	**0.404**	**0.514**	**6.83**
	Mercier Lake	16	0.413	0.477	5.33
	The Right Lake	18	0.409	0.482	4.50
Southeastern Labrador		**59**	**0.435**	**0.507**	**9.00**
	Birchy Lake	27	0.485	0.474	7.17
	Brettney Lake	12	0.470	0.447	6.20
	Crooked Lake	20	0.624	0.666	7.50

Table 3.5. Overall measures of F_{IS} among longnose suckers in each of 14 lakes among four regions in Labrador. There is no significant heterozygote excess/deficiency at $\alpha < 0.05$ after sequential Bonferroni correction

Region	Lake	F_{IS}
Northern Labrador		
	Walkabout Lake	-0.045
	Mistake Lake	0.027
	Strange Lake	-0.057
	Konrad Lake	0.075
	Ikadlivik Brook	0.001
	Cabot Lake	-0.013
	Genetics H	0.050
Western Labrador		
	Lac Joe	0.000
	Lobstick Lake	0.065
Central Labrador		
	Mercier Lake	0.096
	The Right Lake	0.057
Southeastern Labrador		
	Birchy Lake	-0.012
	Brettney Lake	-0.087
	Crooked Lake	-0.047

Table 3.6. Pairwise measures of F_{ST} among populations (lakes) of longnose suckers, inferred from six microsatellite loci. Significance at $\alpha = 0.05$ after sequential Bonferroni correction for multiple comparisons denoted by non-bolded values.

	MER	WAL	BIR	JOE	LOB	STR	KON	IKA	RIG	BRE	MIS	CAB	CRO	GEN
MER	0.000													
WAL	0.046	0.000												
BIR	**0.024**	0.042	0.000											
JOE	0.041	0.051	0.057	0.000										
LOB	0.040	0.041	0.059	**0.007**	0.000									
STR	0.058	**0.017**	0.051	0.036	0.044	0.000								
KON	0.085	0.040	0.111	0.089	0.078	0.085	0.000							
IKA	0.048	0.023	0.072	0.062	0.045	0.068	0.047	0.000						
RIG	0.042	0.060	**0.021**	0.089	0.089	0.079	0.103	0.095	0.000					
BRE	**0.017**	0.037	**0.000**	0.045	0.054	0.043	0.100	0.050	**0.044**	0.000				
MIS	0.042	0.017	0.067	0.047	0.032	0.059	**0.019**	**0.016**	0.067	0.056	0.000			
CAB	0.097	0.049	0.116	0.079	0.055	0.099	0.058	0.046	0.127	0.097	**0.025**	0.000		
CRO	0.110	0.120	0.034	0.149	0.147	0.132	0.193	0.166	**0.019**	0.078	0.151	0.203	0.000	
GEN	0.099	**0.029**	0.105	**0.065**	**0.034**	0.087	**0.030**	**0.024**	**0.124**	0.120	**0.005**	**0.015**	0.204	0.000

highly differentiated. However there were no significant comparisons between southeastern and central lakes, which was surprising since these lakes are distant from each other and the central lakes are at a higher elevation than the southeastern lakes. Genetics H Lake in northern Labrador was also not significantly differentiated from most other lakes.

Using the STUCTURE analysis performed earlier we determined K=4 to be the most likely number of genetic clusters. Cluster 1 contained mainly longnose suckers from central and southeastern Labrador; cluster 2 primarily contained longnose suckers from western and northern Labrador; cluster 3 contained longnose suckers from northern Labrador and the final cluster 4 was composed of the white sucker individuals (Figure 3.4). White sucker individuals were then removed and STRUCTURE analysis was run again from which we were able to confirm the same grouping (K=3; Figure 3.5). Every region contained individuals, recent migrants, from each of the other genetic clusters determined by proportion of assignment with Q values, with the highest level of these individuals found in northern Labrador and lowest levels found in southeastern Labrador (Figure 3.4). In northern Labrador there were 9 individuals from the southeastern/central cluster, 58 from the northern/western cluster, and 85 from the distinct northern cluster. In central there were three individuals from the northern/western cluster, four individuals from the distinct northern cluster, and 26 from the southeastern/central cluster. In western Labrador there were five individuals from the southeastern/central cluster, seven from the distinct northern cluster, and 42 from the northern/western cluster. In southeastern Labrador there were four individuals from the northern/western cluster, three individuals from the distinct northern cluster, and 52 from the southeastern/central cluster.

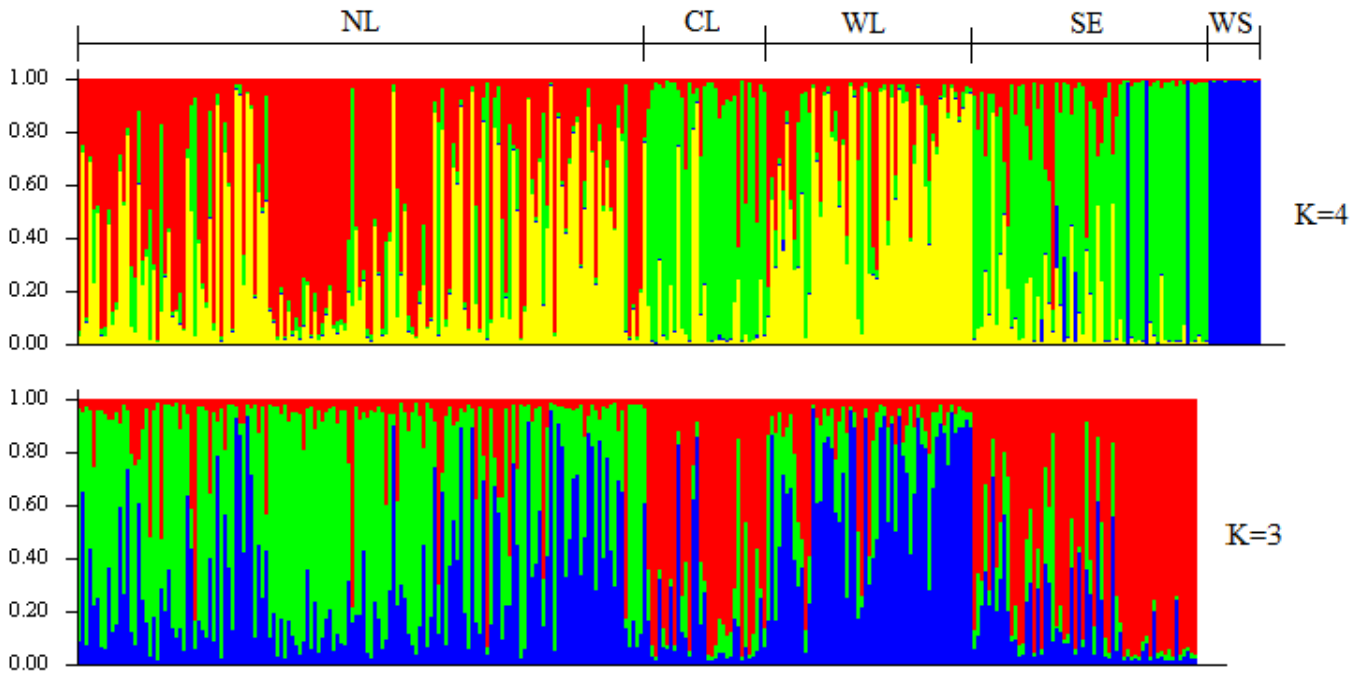

Figure 3.4. Hierarchical STRUCTURE bar plot of admixture proportion (Q) of each individual where *K*=4 contains all Labrador regions and white sucker, and *K*=3 contains all Labrador regions without white sucker, both using an ancestry model that assumes admixture where each vertical line represents an individual fish. The colours correspond to each of the genetic clusters (K), with the proportion of colour in each individual corresponds to its proportion of that genetic cluster. Region labels are: NL, Northern Labrador; WL, Western Labrador; CL, Central Labrador; SE, Southeastern Labrador.WS refers to white sucker.

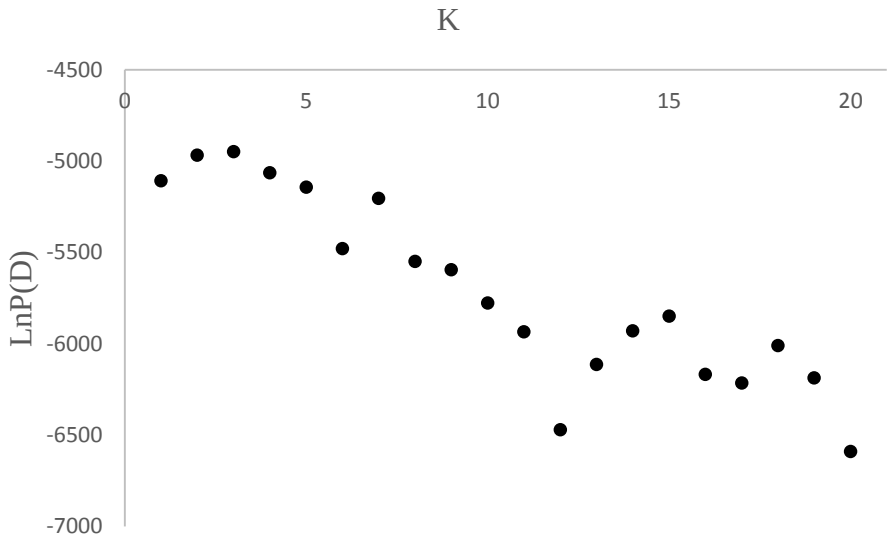

Figure 3.5. A plot of the estimated probability of K (LnP(D)) from the STRUCTURE analysis, for *K*=1-20 without white sucker individuals

All northern and western Labrador samples were then analyzed with STRUCTURE using the same method outlined above. The ideal number of genetic clusters was *K*=1 based on the log likelihood values, indicating that northern and western Labrador are genetically similar. Thus there was no statistical support for the distinct northern group observed in the *K*=3 and *K*=4 analyses at this level (although this cluster was visible in the K=2 plot).

DISCUSSION

Hybridization is increasingly observed among fish species and tends to be more frequent in fish than in any other vertebrate group (Allendorf & Waples 1996). Several factors increase the chance of interspecific hybridization: external fertilization, weak behavioural isolating mechanisms, decreasing habitat complexity, and increased susceptibility to recently evolved forms when in close contact (Scribner *et al*. 2001). In freshwater, multiple species also tend to inhabit the same geographically restricted locations (Barthel *et al*. 2010) which limit the quantity and therefore the choice of conspecific mates leading to an increased incidence of hybrids. Here we report the first genetic assessment of hybrid individuals within longnose sucker populations in Labrador as inferred from variability at six microsatellite loci. We also discuss the patterns of genetic diversity and population structure revealed by the six loci, and compared these conclusions with conclusions obtained from mitochondrial DNA data.

Bayesian admixture analyses reveal hybridization between longnose and white suckers in southeastern Labrador

Interspecific hybridization among catostomids has been documented across western and central Canada based on morphology since Carl Hubbs' report (1955), however the eastern part of the range of the longnose sucker has remained uninvestigated until now. We observed some evidence for hybridization in Labrador between the longnose and white sucker in southeastern Labrador only. Based on the STRUCTURE and NewHybrids analyses, 1.32% (n= 4/302) of longnose suckers have hybrid or admixed ancestry, respectively, all occurring in Birchy Lake. Using meristic characters Dauble and Buschbom (1981) identified ~12-15% hybrids between *Catostomus macrocheilus* and *C. columbianus* while a study by McDonald and colleagues (2008) uncovered ~9% hybrids between *C. Latipinnis* and *C. discobolus* following morphological criteria determined by Baxter and Stone (1995), suggesting that we have lower levels of hybridization in Labrador than can be observed across the sucker family.

Dion and colleagues (1994) observed that white sucker males participated in 32% of female longnose sucker spawnings in the Gouin reservoir, Québec. This high level of interspecific mating was partially attributed to the reduction of reservoir water for maintenance several years previously, which created impassable rapids. It was thought that hybridization between these species was quite rare, however it may be increased in response to a disturbance event (Nelson 1973). In Labrador, the recently constructed Trans-Labrador Highway which runs east to west across the entire land mass unlikely acted as a disturbance event as we have not observed higher levels of hybridization.

Jansson and Öst (1997) described the increased occurrence of hybridization between Atlantic salmon (*Salmo salar*) and brown trout (*Salmo trutta*) after a disturbance event six years prior. This is comparable to the time since the first section of the highway was completed (1992), to the year(s) we have longnose sucker samples from (2003 to 2012) and it therefore does not appear to be affecting hybridization of these two suckers.

The mtDNA data from Chapter 2 identified four putative hybrids between longnose sucker males and white sucker females. Three of these were determined to be white suckers that had been misidentified as longnose sucker based on the microsatellite data while the remaining one was confirmed as a hybrid. This low level of hybridization implies less of an imminent genetic and reproductive threat to the longevity of longnose suckers populations in Labrador as well as to white suckers. In general, homogenization of populations via hybridization creates a risk of loss of the "pure genetic line" of a particular species which could lead to population instability and susceptibility to various diseases or environmental stressors (Rahel 2000). Disturbance events tend to increase the level of hybridization in successive years indicating that the very small percentage of hybrids found (all in one lake) after the construction of the Trans-Labrador highway may indicate that these suckers may not be in the same spawning areas or there may be pre-zygotic isolating barriers not observed in other freshwater systems across North America and therefore it will be important to continue to monitor populations of longnose suckers in Labrador.

Regional genetic diversity and structure among longnose sucker populations in Labrador

The highest levels of allelic richness and expected heterozygosity were found in northern and western Labrador while the lowest values were found in central Labrador. The levels of genetic diversity may be attributed to the dispersal abilities of the longnose sucker throughout each region, both historical and contemporary. The high levels likely indicate the mixture of different refugial sources for suckers in northern and western Labrador where evidence of all three refugia (Beringian, Mississippian, and Atlantic) was present, compared with central and southeastern Labrador whose populations likely originated solely from the Atlantic refugium. Low levels of expected heterozygosity in central Labrador is similar to the pattern observed in brook trout populations in similar locations in each region (Pilgrim 2011). Throughout Labrador, glaciers remained present in higher elevations, such as in central Labrador, for longer post-glacial time periods, which would have given populations of fish less time to accumulate variation once they colonized the region. This similar heterozygosity patterns suggests inland fish likely followed similar dispersal routes of colonization throughout Labrador based on the landscape which includes the Torngat Mountains to the North, deep canyons, higher elevation in central and western Labrador, and different habitat types through each region (from barrens to wetland).

Interpretations of F_{ST} can be based on four classes of values for the assessment of genetic differentiation: 0-0.05 for little, 0.05-0.15 for moderate, 0.15-0.25 for large and

>0.25 for very large (Wright 1978). In Chapter 2, we obtained significant F_{ST} values for population comparisons between western Labrador and other regions, differentiating it from the rest of Labrador. However, all other F_{ST} values were not significant leading us to suggest that the structure in Labrador was at a regional rather than fine scale, and associated with post-glacial recolonization from different refugia. With the microsatellite data, most pairwise comparisons among populations of longnose suckers resulted in a moderate amount of genetic differentiation. The highest level of genetic differentiation occurred between a population in the north and one in the southeast confirming a lack of genetic mixing/movement between these regions, supporting the conclusion that genetic structure in Labrador exists at a regional scale. However, the observation that most lakes are moderately and significantly different from each other indicates a finer level of population structure than detectable with mtDNA, suggesting that within and between regional gene flow may be limited.

Refugial origins of Labrador suckers inferred from microsatellite structure

In Chapter 2 we proposed that longnose suckers in Labrador originated from one main refugial source, the Atlantic, and two smaller refugia, the Beringian and Mississippian, based on an analysis of mitochondrial DNA. We identified two refugial origins in western and northern Labrador (Mississippian and Beringian) and a third found in all regions of Labrador (Atlantic). Based on the microsatellite genetic clustering at K=3 we see Labrador populations segregate into three main clusters: northern and western, central and southeastern, and a distinct northern. The longnose sucker clusters

correspond closely to the network clades identified with mitochondrial DNA. At K=3 we have a red clade that is found in all areas, however almost all of southeastern and central Labrador and composed of it, suggesting this may correspond with the Atlantic refugium. The blue cluster found in northern and western Labrador may represent the Beringian while the distinct northern cluster, in green may represent the Mississippian refugia. Several locations in northern Labrador in this distinct green cluster appear to have experienced recent isolation as they are distinct from the rest of the north and west with respect to the microsatellite loci. However, this observation is not apparent once central and southeastern samples are removed indicating that we have a poorly supported group which may be better supported with additional loci designed for longnose sucker. Three refugial origins were also proposed for brook trout in Labrador, including the Atlantic and the Mississippian as well as one other unknown refugium in northern Labrador (Pilgrim 2011). It appears that different fish with different habitat types ultimately followed some of the same paths into Labrador following the Wisconsinan glaciation period.

The strong structure identified among each region may be largely historical reflecting patterns of post-glacial colonization from different refugial origins. However, the microsatellite STRUCTURE analysis suggests low levels of persistent contemporary gene flow in the form of individuals from non-region associated clusters within each cluster-defined region, which could represent recent migrants. In particular, there are individuals from central and southeast Labrador which have dispersed to northern and western locations and vice versa. MtDNA F_{ST} values, which reflect historical patterns, were low among lakes, but high among the regions. More recent patterns revealed by

STRUCTURE indicate migrants in each region which suggests some level of ongoing gene flow, however may be clarified further by an isolation by distance matrix.

Major findings and conclusions

We were able to use microsatellite loci designed for different species of sucker for hybrid identification within longnose suckers. The level of hybridization observed here is lower than the levels observed in different sucker species and based on the range and abundance of both species there does not appear to be any immediate threat to the longevity of either. However, disturbance events tend to increase the level of hybridization with the potential that hybridization will continue and therefore it will be important to continue to monitor populations of longnose suckers in Labrador.

The refugial origins of longnose suckers in Labrador proposed in Chapter 2 were further supported with microsatellite data; we found potential evidence of Beringian, Mississippian, and Atlantic refugial sources of longnose suckers throughout Labrador. The microsatellite data provides a clearer fine-scale pattern of population structure as it identified limited ongoing gene flow.

REFERENCES

Aboim, M.A., Mavarez, J., Bernatchez, J., Coelho, M.M. (2010). Introgressive hybridization between two Iberian endemic cyprinid fish: a comparison between two independent hybrid zones. *Journal of Evolutionary Biology*, 23, 817-828.

Allendorf, F.W., & Waples, R.S. (1996). Conservation and genetics of salmonid fishes. *Conservation Genetics*, pp 238-280.

Anderson, E.C., & Thompson, E.A. (2002). A model-based method for identifying species hybrids using multilocus genetic data. *Genetics*, 196, 1217-1229.

Banister, K.F. (1998). *Encyclopedia of Fishes*. San Diego: Academic Press. P.100.

Barthel, B.L., Dorothy, O.M., Philipp, D.P. (2010). Molecular genetic confirmation of hybridization between largemouth and smallmouth bass (*Micropterus*) in the wild. *Copeia*, 2010, 671-675.

Baxter, G.T., & Stone, M.D. (1995). *Fishes of Wyoming* (Wyoming Game and Fish Dept, Cheyenne), pp 129-131.

Bernardo, J.M., & Smith, A.F.M. (1994). *Bayesian theory*. John Wiley and Sons, New York, New York, USA.

Cardall, B.L., Bjerregaard, L.S., Mock, K.E. (2006). Microsatellite markers for the June sucker (Chasmistes liorus mictus), Utah sucker (Catostomus ardens), and five other catostomid fishes of western North America. *Molecular Ecology Notes*, 7, 457-460.

Dauble, D.D., & Buschbom, R.L. (1981). Estimates of hybridization between 2 species of catostomids in the Columbia River. *Copeia*, 4, 802-810.

Dion, R., Richardson, M., Roy, L., Whoriskey, F.G. (1994). Spawning patterns and interspecific matings of sympatric white (*Catostomus commersoni*) and longnose (*C. catostomus*) suckers from the Gouin reservoir system, Quebec. *Canadian Journal of Zoology*, 72, 195-200.

Doosey, M.H., Bart, H.L., Saitoh, K., Miya, M. (2010). Phylogenetic relationships of catostomid fishes (Actinopterygii: Cypriniformes) based on mitochondrial ND4/ND5 gene sequences. *Molecular Phylogenetics and Evolution*, 54, 1028-1034.

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

Falush, D., Stephens, M., Pritchard, J.K. (2003). Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, 164, 1567-1587.

Ferris, S.D. (1984). Tetraploidy and the evolution of the catostomid fishes. *Evolutionary Genetics of Fishes, Monographs in Evolutionary Biology*, 55-93.

Gilk, S.E., Wang, I.A., Hoover, C.L., Smoker, W.W., Taylor, S.G., Gray, A.K., Gharrett, A.J. (2004). Outbreeding depression in hybrids between spatially separated pink

salmon, *Oncorhynchus gorbuscha*, populations: marine survival, homing ability, and variability in family size. *Environmental Biology of Fishes*, 69, 287-297.

Guo, S.W., & Thompson, E.A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, 48, 361-372.

Hubbs, C.L. (1955). Hybridization between fish species in nature. *Systematic Zoology*, 4, 1-20.

Jansson, H., & Öst, T. (1997). Hybridization between Atlantic salmon (*Salmo salar*) and brown trout (*S. trutta*) in a restored section of the River Dälalven, Sweden. *Canadian Journal of Fisheries and Aquatic Sciences*, 54, 2033-2039.

Jeffreys, H. (1946). An invariant form for the prior probability in estimation problems. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 186, 453-461.

Jensen, J.L., Bohonak, A.J., Kelly, S.T. (2005). Isolation by distance, web service. *BMC Genetics*, 6, 13. v.3.23 http://ibdws.sdsu.edu/.

Kwak, M., & Gepts, P. (2009). Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae). *Theoretical and Applied Genetics*, 118, 979-992.

McDonald, D.B., Parchman, T.L., Bower, M.R., Hubert, W.A., Rahel, F.J. (2008). An introduced and a native vertebrate hybridize to form a genetic bridge to a second

native species. *Proceedings of the National Academy of Sciences*, 105, 10837-10842.

McPhail, J.D., & Taylor, E.B. (1999). Morphological and genetic variation in northwestern longnose suckers, *Catostomus catostomus*: The salish sucker problem. *American Society of Ichthyologists and Herpetologists*, 1999, 884-893.

Miller, R.R., & Smith, G.R. (1981). Distribution and evolution of Chasmistes (Pisces: Catostomidae) in western North America. *Occasional Papers of the Museum of Zoology*, University of Michigan, 696, 1-46.

Millette, K.L., Xu, S., Witt, J.D.S., Cristescu, M.E. (2011). Pleistocene-driven diversification in freshwater zooplankton: Genetic patterns of refugial isolation and post-glacial recolonization in *Leptodora kindtii* (Crustacea, Cladocera). *Limnology and Oceanography*, 56, 1725-1736.

Nelson, J.S. (1968). Hybridization and isolating mechanisms between *Catostomus commersonii* and *C. macrocheilus* (Pisces: Catostomidae). *Journal of the Fisheries Research Board of Canada*, 25, 101-150.

Nelson, J.S. (1973). Occurrence of hybrids between longnose sucker (*Catostomus catostomus*) and white sucker (*Catostomus commersoni*) in Upper Kananaskis Reservoir, Alberta. *Journal of the Fisheries Research Board of Canada*, 30, 557-560.

Page, L.M., & Burr, B.M. (1991). *A field guide to freshwater fishes of North America north of Mexico*. Houghton Mifflin Company, Boston, 432.

Perry, W.L., Lodge, D.M., Feder, J.L. (2002). Importance of hybridization between indigenous and nonindigenous freshwater species: An overlooked threat to North American biodiversity. *Systematic Biology*, 51, 255-275.

Pilgrim, B.L. (2011). Phylogeography and post-glacial dispersal of brook trout (*Salvelinus fontinalis*) in Labrador, Canada, inferred from microsatellite and mitochondrial DNA variation. Master of Science thesis.

Pritchard, V.L., Jones, K., Cowley, D.E. (2007). Genetic diversity within fragmented Cutthroat trout populations. *Transactions of the American Fisheries Society*, 136, 606-623.

Pritchard, J.K., Wen, X., Falush, D. (2010). Documentation for *structure* software: Version 2.3. http://pritch.bsd.uchicago.edu/structure.html.

Rahel, F.J. (2000). Homogenization of fish faunas across the United States. *Science*, 288, 854-856.

Roy, M.S., Greffen, E., Smith, D., Ostrander, E.A., Wayne, R.K. (1994). Patterns of differentiation and hybridization in North American wolflike canids, revealed by analysis of microsatellite loci. *Molecular Biology and Evolution*, 11, 553-570.

Scribner, K.T., Page, K.S., Barton, M.L. (2001). Hybridization in freshwater fishes: a review of case studies and cytonuclear methods of biological inference. *Reviews in Fish Biology and Fisheries*, 10, 293-323.

Smith, G.R., & Koehn, R.K. (1971). Phenetic and cladistics studies of biochemical and morphological characteristics of *Catostomus*. *Systematic Zoology*, 20, 282- 297.

Tranah, G.J., Agresti, J.J., May, B. (2001). New microsatellite loci for suckers (Catostomidae): primer homology in Catostomus, Chasmistes, and Deltistes. *Molecular Ecology Notes*, 1, 55-60.

Turner, T.F., Dowling, T.E., Osborne, M.J., McPhee, M.V., Broughton, R.E., Gold, J.R. (2009). Microsatellite markers for the endangered razorback sucker, Xyrauchen texanus, are widely applicable to genetic studies of other catostomine fishes. *Conservation Genetics*, 10, 551-553.

Van Oosterhout, C., Hutchinson, W.F., Wills, D.P.M., Shipley, P. (2004). Micro-checker: Software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes*, 4, 535-538.

Weir, B.S., & Cokerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 1358-1370.

Whittaker, J.C., Harbord, R.M., Boxal, N., Mackay, I., Dawson, G., Sibly, R.M. (2003). Likelihood-based evolution of microsatellite mutation rates. *Genetics*, 164, 781-787.

Wright, S. (1978). *Evolution and the genetics of populations – Variability within and among natural populations* (1st ed.). University of Chicago Press, Chicago IL.

CHAPTER 4: GENERAL CONCLUSIONS

The longnose sucker is a freshwater fish found across North America and Siberia and is an important subsistence fish for small communities across Canada. The western part of the range has been studied for phylogeographic origins of longnose suckers. Hybrid identification with close relatives, the white sucker based on morphological characters has been studied across western and central Canada. However, the eastern part of the range has remained unstudied providing difficulties when attempting to assess an effective management strategy for this important subsistence fish. The most eastern part of their range, Labrador is an area of interest as it has remained relatively unaltered by human involvement until the recent construction of the Trans-Labrador Highway (TLH). This new construction may have altered fish dispersal ability and subsequent longevity of the longnose sucker in that range as they may become isolated. Therefore, the objectives of the present study were to understand the refugial origin and route(s) of post-glacial re-colonization of *Catostomus catostomus* into Labrador, present-day patterns of gene flow and population structure in Labrador, and the extent of hybridization with the white sucker.

Refugial origins and genetic structure of longnose suckers in Labrador

Longnose suckers (*Catostomus catostomus*) were thought to have survived the Wisconsinan glaciation period in at least three main refugia: the Beringian, Pacific, and Mississippian (McPhail& Taylor 1999). However, only longnose suckers from the western part of their range were used in their study so it remained unclear which post-

glacial dispersal routes and refugial origins were present in the eastern range. To address this, the evolutionary genetic marker (Nabholz*et al.* 2008) mitochondrial DNA was chosen as it is useful in phylogeographic and phylogenetic studies since it evolves rapidly and is maternally inherited (Zardoya*et al.*1995); mtDNA was therefore also used here to assess genetic structure and post-glacial dispersal of longnose suckers in the remote area of Labrador.

Among 362 fish, the mitochondrial DNA fragment yielded 20 haplotypes within Labrador, five Yukon haplotypes, and four British Columbia haplotypes, with no shared haplotypes among any of these provinces. We identified a main Labrador haplotype which was found in all regions (northern, central, western, and southeastern) and lakes within Labrador as well as numerous closely-related, less frequent haplotypes which were generally located in only one region or lake. Negative values of Tajima's D and Fu's Fs indicated that our populations have not reached equilibrium since the re-colonization event into Labrador. These patterns are consistent with a recent colonization of these longnose suckers to Labrador. Spatial analysis of molecular variance (SAMOVA) and a median spanning network of haplotypes provided evidence for one main glacial refugial source of suckers within Labrador: the Atlantic refugium. However slight evidence also exists to support two additional refugial sources in western and northern Labrador: The Beringian and the Mississippian. It is likely that longnose suckers entered Labrador from the Mississippian and the Beringian during a similar time period via Québec while longnose suckers from the Atlantic likely traveled up along the east coast as ice melted.

In the course of the mtDNA investigation, we identified four fish with white sucker mitochondrial DNA, which suggested that these individuals may be hybrids between the longnose and white sucker or we may have misidentified the fish. Hybridization between these two species was first reported by Nelson (1973) with a second report by Dion and colleagues (1994) confirming interspecific mating between these sucker species. However maternal white suckers had not been documented in either previous study, and therefore we are the first to document this directionality of spawning between the longnose and white sucker if further investigation proves these fish to be hybrids.

Hybridization and population structure

Hybridization of catostomids has been an active area of research as they appear to hybridize freely with the ability to produce viable offspring (Ferris 1984) and in the Western USA one in every hundred suckers was an interspecific hybrid (Hubbs 1955). Extensive hybridization occurred more often when ideal conditions were not present for mating in the system such as a low population size of one sucker species relative to another (Hubbs 1955). Hybridization between the longnose and the white sucker has been documented across North America (Dion *et al.* 1994; Nelson 1973), using meristic characters. Here we used microsatellite data to document hybridization in Labrador as microsatellite markers have high mutation rates leading to high allelic diversity, and are inherited biparentally (Cardall *et al.* 2006; Roy *et al.* 1994). Further, we used

microsatellites to describe the patterns of genetic diversity and population structure of longnose suckers in Labrador and compared these results with the mitochondrial DNA patterns.

A total of six microsatellite loci were used to genotype 302 morphologically-identified longnose and 14 white sucker samples from Labrador. Based on the STRUCTURE and NewHybrids analyses, 1.32% of longnose suckers have hybrid or admixed ancestry, respectively, with all admixed individuals being found in a single lake (Birchy Lake) in southeastern Labrador. This lake may have an uneven ratio of one sucker to the other (lack of mate availability) or the habitat of this lake may be primarily suited for one fish over the other and therefore one species may be outcompeting the other during spawning time. However, overall these values are less than the values determined for other sucker species (Dauble & Buschbom 1981; McDonald *et al.* 2008) where disturbed populations tended to have a highly increased level of hybridization. This could be because these suckers may not be in the same spawning areas or there may be pre-zygotic isolating barriers not observed in other freshwater systems across North America. It will be important to look into a few of these possibilities in future work in order to understand why these two suckers are not hybridizing at a higher rate.

North American freshwater fish are some of the most diverse in the world (Saitoh *et al*. 2006). Partially due to vicariance, as the landscape is extremely variable with a wide range of habitats, this diversity can be attributed primarily to the way fish dispersed during the pleistocene Era (Perry *et al.* 2002). However this diversity along with cohabitation of species tends to increase the incidence of hybridization and may even

facilitate it. As previously reported by Dion and colleagues (1994), white sucker females and longnose sucker males were not observed to spawn together, therefore we were not expecting to observe any white sucker mitochondrial genes within longnose suckers. However, we observed four putative longnose suckers with white sucker mtDNA. After microsatellite analysis we were able to attribute three of these to misidentification, as these individuals were fully white sucker, however one of them we could not. One individual contained mtDNA of white sucker but microsatellite loci of longnose sucker indicating that a paternal longnose spawned with a maternal white sucker. This is the first documented case of this directionality of spawning.

The highest levels of allelic richness and expected heterozygosity were found in northern and western Labrador while the lowest values were found in central and southeastern Labrador which indicated the mixture of refugial origins (one main and two minor) in each region. Most lakes were moderately and significantly different from each other based on F_{ST} values which indicated a finer level of population structure than detectable with mtDNA, suggesting that ongoing gene flow may also be limited and based on small amount of migrants in each region from each region, we confirm limited gene flow to be likely. Based on bar plots in STRUCTURE at K=3, the clusters identified correspond closely to the three network clades identified by mitochondrial DNA. Therefore we can conclude the refugial origins of the Atlantic, Beringian, and Mississippian proposed in Chapter 2.

The impact of the Trans-Labrador Highway on longnose sucker populations

The Trans-Labrador Highway (TLH), of which construction began in 1992 and was completed in 2010, is the primary public road in Labrador, running from Québec to western Labrador and on to the southeast. Specifically, the TLH runs from the Québec-Labrador border to Labrador City/Wabush then to Churchill Falls, connecting to Happy Valley-Goose Bay, and ending in Cartwright junction. What was once a fully remote area was made more accessible by the construction of this highway as it connected previously isolated areas including lakes and streams. Longnose suckers among other fish may have experienced habitat restriction or destruction, or even population isolation, due to the TLH. The highway separates the southeastern populations studied here, as well as both the western populations. The significant values of F_{ST} between these pairs of populations and the levels of hybridization within them indicates that the TLH may be altering the way fish are dispersing in these regions of Labrador. However, the level of hybridization observed is less than levels observed in other catostomids across North America (Dauble & Buschbom 1981; McDonald *et al.* 2008) which implies that the highway may not be acting as a disturbance event (geographic barrier) which tends to increase hybridization in catostomids (Dion *et al.* 1994). However, the time period since the construction of the highway may not be long enough for populations of longnose sucker in each region to experience increased levels of hybridization.

Comparison of marker types and utility of using both microsatellites and mtDNA

The use of only one molecular marker when answering large scale or complex questions can lead to only half of the picture as our data clearly shows. Mitochondrial DNA indicated that there were four putative hybrids in the data set and based on the known ability of longnose suckers and white suckers to hybridize readily, as well as the fact that these four individuals were adults that reduced the chance of misidentification, it seemed plausible that they were true hybrids. The white sucker mtDNA may have been obtained in a historical introgression event as mtDNA is maternally inherited. By using microsatellites we were able to quantify the level of hybridization and uncovered a few inconsistencies. We identified hybrid individuals via microsatellites that were not detected via mitochondrial DNA which indicates the importance of using both marker types when doing this kind of analysis. We also identified three individuals which had been identified as longnose suckers but contained microsatellite loci and mitochondrial DNA identical to the white sucker. This indicates the utility of multilocus analysis for increased confidence in identifying species.

The use of both microsatellites and mtDNA for investigating population structure is advantageous as well. MtDNA is historical and reflects female-mediated patterns of gene flow while microsatellites can overlay contemporary patterns involving both sexes. Here, both marker systems revealed three clusters that could be attributed to three glacial refugial origins. Microsatellites indicated fine scale structure in northern Labrador, separating certain northern lakes into a separate group. The values of F_{ST} among populations within each region, revealed by the mtDNA data, were not significant,

indicating no differentiation among geographically close populations. While this could be interpreted as either large amounts of contemporary gene flow leading to homogenous population structure within regions, it could also mean that mtDNA is not evolving rapidly enough to detect recent patterns of gene flow. Gene flow is an important factor in introducing genetic variation into populations and high levels of gene flow can even decrease the chance of speciation as genes that move around keep different populations genetically similar (Slatkin 1987). The migrants found in different Labrador populations suggest that longnose suckers will remain a single strong species as exchange of genes between populations and regions increases fitness (Sexton *et al.* 2011) and ultimately the perseverance of this long understudied species.

Conclusions

The longnose sucker is genetically understudied, globally and in Labrador, and has been given a new perspective by the use of mitochondrial DNA and microsatellite analysis. Phylogeographic origins, dispersal patterns, and population structure as well as extensive hybridization with their close relative, the white sucker, have been explored and discussed here, along with the potential impact of the newly constructed Trans-Labrador highway on dispersal patterns and levels of hybridization. This research should increase awareness and interest in this long overlooked species. More study is needed throughout the full range of the species to determine the complete extent of hybridization with other

species and to better inform conservation and management of the longnose sucker in Labrador and elsewhere.

REFERENCES

Cardall, B.L., Bjerregaard, L.S., Mock, K.E. (2006). Microsatellite markers for the June sucker (Chasmistes liorus mictus), Utah sucker (Catostomus ardens), and five other catostomid fishes of western North America. *Molecular Ecology Notes*, 7, 457-460.

Dauble, D.D., & Buschbom, R.L. (1981). Estimates of hybridization between 2 species of catostomids in the Columbia River. *Copeia*, 4, 802-810.

Dion, R., Richardson, M., Roy, L., Whoriskey, F.G. (1994). Spawning patterns and interspecific matings of sympatric white (*Catostomus commersoni*) and longnose (*C. catostomus*) suckers from the Gouin reservoir system, Quebec. *Canadian Journal of Zoology*, 72, 195-200.

Ferris, S.D., & Whitt, G.S. (1980). Genetic variability in species with extensive gene duplication: the tetraploid Catostomid fishes. *The American Naturalist*, 115, 650-666.

Hubbs, C.L. (1955). Hybridization between fish species in nature. *Systematic Zoology*, 4, 1-20.

McDonald, D.B., Parchman, T.L., Bower, M.R., Hubert. W.A., Rahel, F.J. (2008). An introduced and a native vertebrate hybridize to form a genetic bridge to a second native species. *Proceedings of the National Academy of Sciences*, 105, 10837-10842.

McPhail, J.D., & Taylor, E.B. (1999). Morphological and genetic variation in northwestern longnose suckers, *Catostomus catostomus*: The salish sucker problem. *American Society of Ichthyologists and Herpetologists*, 1999, 884-893.

Nabholz, B., Mauffrey, J.F., Bazin, E., Galtier, N., Glemin, S. (2008). Determination of mitochondrial diversity in mammals. *Genetics*, 178, 351-361.

Nelson, J.S. (1973). Occurrence of hybrids between longnose sucker (*Catostomus catostomus*) and white sucker (*Catostomus commersoni*) in Upper Kananaskis Reservoir, Alberta. *Journal of the Fisheries Research Board of Canada*, 30, 557-560.

Perry, W.L., Lodge, D.M., Feder, J.L. (2002). Importance of hybridization between indigenous and nonindigenous freshwater species: An overlooked threat to North American biodiversity. *Systematic Biology*, 51, 255-275.

Roy, M.S., Greffen, E., Smith, D., Ostrander, E.A., Wayne, R.K. (1994). Patterns of differentiation and hybridization in North American wolflike canids, revealed by analysis of microsatellite loci. *Molecular Biology and Evolution*, 11, 553-570.

Saitoh, K., Sado, T., Mayden, R.L., Hanzawa, N., Nakamura, K., Nishida, M., Miya, M. (2006). Mitogenomic evolution and interrelationships of the Cypriniformes (Actinopterygii: Ostariophysi): The first evidence toward resolution of higher-level relationships of the world's largest freshwater fish clade based on 59 whole mitogenome sequences. *Journal of Molecular Evolution*, 63, 826-841.

Sexton, J.P., Strauss, S.Y., Rice, K.J. (2011). Gene flow increases fitness at the warm edge of a species' range. *Proceedings of the National Academy of Sciences (PNAS)*, 168, 11704-11709.

Zardoya, R., Garrido-Pertierra, A., Bautista, J.M. (1995). The complete nucleotide sequence of the mitochondrial DNA genome of the rainbow trout, *Oncorhynchus mykiss*, 41, 942-951.

APPENDIX 1

Co-Authorship Statement

The study was designed by Barbara Langille and Dawn Marshall. Samples were collected by Robert Perry, Donald Keefe and Barbara Langille. All laboratory work was performed by Barbara Langille. Data analysis and the first draft of the manuscript were completed by Barbara Langille. Dawn Marshall, Robert Perry, and Donald Keefe provided comments and suggestions and edited the manuscript.

Novel mitochondrial DNA primers for identification of population trends in longnose suckers (*Catostomus catostomus*) and multispecies identification

Barbara L. Langille (BLL), Robert Perry (RP), Donald Keefe (DK), H. Dawn Marshall (HDM)

The understudied longnose sucker has received no attention by the International Union for Conservation of Nature (IUCN) wherein almost 55% of fish in this family have been classified as threatened, endangered, extinct, or experiencing population declines. Understanding population trends of the ecologically diverse longnose sucker is critical for preservation of this unique species, as it hybridizes readily with congenerics. Here we describe three novel pairs of mitochondrial DNA (mtDNA) primers with which we

amplified genes from four species: longnose and white sucker, round whitefish, and slimy sculpin. Nucleotide substitutions ranged from 10 to 111 between longnose sucker and among various species. Hybridization may be occurring as a morphologically identified longnose sucker contained white sucker mtDNA. Species level identification using these primers with a larger scope of landscape studies is possible here as these species belong to three orders, showing the utility of these primers on a variety of taxa.

Introduction

The longnose sucker (*Catostomus catostomus*) is a long-lived, freshwater/brackish-water fish, geographically ranging across North America, and is one of two catostomids found in Siberia (McPhail and Taylor 1999). During the spring, longnose suckers swim from lakes to shallow streams to spawn and tend to return to the same location over successive years (Bailey 1969). This catostomid is found alongside several native and non-native fish, including cutthroat trout, grayling and white sucker (Crait et al, 2006; Nelson 1973).

Despite its prevalence, longnose sucker has yet to be classified by the *International Union for Conservation of Natures* (IUCN) Red List which describes threatened species. The IUCN Red List for the family Catostomidae currently comprises 67 species of which ~31% are threatened and ~24% of least concern species are undergoing population declines or are unknown. These Catostomidae may move into a higher priority category with need for conservation in the near future, demonstrating the

importance of research on this family. Known to be avid hybridizers (Hubbs 1955), catostomids require monitoring as hybridization may threaten the prevalence of parental species (Dowling and Childs 1992).

As a first step towards identifying population trends, we created primers pairs for the mitochondrial cytochrome b (CYTb), NADH dehydrogenase 4 (ND4), and cytochrome oxidase 1 (CO1) gene regions. Developed for the longnose sucker, these primers have cross-species utility with white sucker (*Catostomus comersonni*), round whitefish (*Prosopium cylindraceum*), and slimy sculpin (*Cottus cognatus*).

Methods

To design primers, conserved regions of the white sucker mitochondrial genome (Accession: NC_008647.1) was compared to longnose sucker gene fragments (Accession: U40559.1, EU524470.1, and FJ751808.1), found in the National Centre of Biotechnology Institute (NCBI) nucleotide database (Table 1). Fish were caught in Labrador using gill nets, excluding sculpin which were removed from fish stomachs, and fin clippings were stored in -20°C. Genomic DNA was extracted from 16 fin clips using the Qiagen QIAamp® DNA Mini Kit (Qiagen Inc., Toronto, Canada) according to the manufacturer's tissue protocol. Amplification of gene regions by the polymerase chain reaction (PCR) was carried out in reactions with a final volume of 25uL containing: 10X Qiagen PCR buffer (Qiagen Inc.), 200 µM dNTPs (New England Biolabs Inc., Canada), 400 nM of forward/reverse primers, 1U HotStar *Taq* DNA polymerase (Qiagen Inc.), and 25-200 ng

of DNA template. Thermal cycling consisted of an initial denaturation step of 95℃ for 5 min (2 min for CO1), followed by 30 cycles of 94℃ for 60 s, 45℃-60℃ for 60 s (30 s for CO1) and 72℃ for 90 s, with a final extension of 72℃ for 5 min. PCR products were purified using the QIAquick PCR Purification kit (Qiagen Inc.) and cycle sequencing was subsequently run on the ABI 3730 DNA Analyzer (Applied Biosystems Inc., Foster City, USA). Sequence reads were aligned and edited using Sequencher v.5.1.

Results

The three fragments resulted in 1344 bp of sequence data for the morphologically identified longnose sucker. Using a basic local alignment search tool (BLAST), each sequence was verified to species, with 9 sequences matching longnose sucker (sequence identity 100%; E-value 1e-144), and one matching white sucker (sequence identity 100%; E-value 3e-175). We obtained 308 bp (CTYb) for round whitefish (sequence identity 99%; E-value 1e-144), and 630 bp (CO1) for slimy sculpin (sequence identity 99%; E-value 0.0) which were also verified to species using BLAST. These sequences were characterized by 10 variable sites (0.744% nucleotide variation) among longnose suckers, 120 variable sites (8.93% nucleotide variation) between longnose and white suckers, 78 variable sites (10.92% nucleotide variation) between longnose and round whitefish, and 111 variable sites (18.29% nucleotide variation) between longnose and sculpin (Table 2).

Table 1. Details of three primer sets developed for longnose sucker using white sucker sequence as reference. Abbreviations: LNS= longnose sucker, RWF= round whitefish, WS= white sucker, and SS= slimy sculpin.

Locus name	Forward (5'-3')	Reverse (5'-3')	Species	Ta (°C)	Sequence length bp	Sample size
Cytb	GTAAAACGACGGCCAGTATGATGAAA	CAGGAAACAGCTAATATTTGTCCTCA	LNS	45	364	9
			RWF	45	364	2
			WS	45	364	1
ND4	GATTTTGGCCAGCCAGAACCA	TCAGGACTCAAGGACAAGGGGT	LNS	50	350	9
			WS	50	350	1
CO1	TCAACCAACCACAAAGACATTGGCAC	TAGACTTCTGGGTGGCCAAAGAATCA	LNS	60	630	9
			SS	60	630	4
			WS	60	630	1

Table 2. Summary of sequence data based on comparisons between longnose sucker and longnose sucker, white sucker, round whitefish, and slimy sculpin. Abbreviations: LNS= longnose sucker, RWF= round whitefish, WS= white sucker, and SS= slimy sculpin.

CYTb

LNS	Variable sites	% variable	Number of haplotypes	Transitions/ transversions	Synonymous/ Nonsynonymous	dN/dS
LNS	3	0.974	2	3/0	1/2	0.020
WS	24	7.79	2	22/2	4/18	0.180
RWF	69	22.40	2	44/25	25/54	0.540

ND4

LNS	Variable sites	% variable	Number of haplotypes	Transitions/ transversions	Synonymous/ Nonsynonymous	dN/dS
LNS	4	1.37	2	4/0	1/3	0.031
WS	38	12.93	2	30/8	5/33	0.337

CO1

LNS	Variable sites	% variable	Number of haplotypes	Transitions/ transversions	Synonymous/ Nonsynonymous	dN/dS
LNS	3	0.494	2	2/1	2/1	0.005
WS	53	8.73	2	42/11	46/7	0.034
SS	111	18.29	3	58/53	94/14	0.069

Discussion

Based on white sucker mtDNA we conclude presence of one putative longnose/white sucker hybrid or historical introgression. Hybridization may occur from introductions of one or more species, disturbance of habitat, overlapping in spawning periods, or lack of an appropriate mate (Barthel et al, 2010; Nelson 1965). Ongoing research is being conducted using microsatellites to quantify extent of white sucker introgression (Cardall et al, 2006; Turner et al, 2009) into Labrador longnose populations, based on observations by Nelson (1973). Using morphological characteristics and scale raker counts, Nelson (1973) noted hybrids were intermediates of the parental suckers, however could be more like one parent over the other. A putative hybrid has been uncovered here which would have been misidentified as longnose sucker using visual characteristics alone.

Orders represented in this study, Cypriniformes, Salmoniformes, and Scorpaeniformes, encompass ~4788 species. Partially-digested sculpin and round whitefish were successfully identified, indicating the ability for species level identification in degraded tissue samples and for various taxa. The wide variety of species used here suggests these primers may be useful for a larger host of fish which will be helpful for landscape studies.

Acknowledgements

We thank the Government of Newfoundland and Labrador (Department of Environment and Conservation, Wildlife Division) for funding this project.

References

Bailey M (1969) Age growth and maturity of Longnose sucker (*Catostomus catostomus*) of Western Lake Superior. *Journal of the Fisheries Research Board of Canada*, 26, 1288-1298.

Barthel BL, Dorothy OM, Philipp DP (2010) Molecular genetic confirmation of hybridization between largemouth and smallmouth bass (Micropterus) in the wild. *Copeia*, 2010, 671-675.

Cardall BL, Bjerregaard LS, Mock KE (2006) Microsatellite markers for the June sucker (*Chasmistes liorus mictus*), Utah sucker (*Catostomus ardens*), and five other catostomid fishes of western North America. *Molecular Ecology Notes*, 7, 457-460.

Crait JR, Ben-David M (2006) River otters in Yellowstone National lake depend on a declining cutthroat trout population. *Journal of Mammalogy*, 87, 485-494.

Dowling TE, Childs MR (1992) Impact of hybridization on a threatened trout of the southwestern United States, Conservation Biology, 6, 355-364.

Hubbs CL (1955) Hybridization between fish species in nature. *Systematic Biology*, 4, 1-20.

McPhail JD, Taylor EB (1999) Morphological and genetic variation in northwestern longnose suckers, *Catostomus catostomus*: The salish sucker problem. *American Society of Ichthyologists and Herpetologists*, 1999, 884-893.

Nelson JS (1965) Effect of fish introductions and hydroelectric development on fishes in the Kananaskis River system, Alberta. *J. Fish. Res. Board Can,* 22, 721-753.

Nelson JS (1973) Occurrence of hybrids between longnose sucker (*Catostomus catostomus*) and white sucker (*C. commersoni*) in upper Kananaskis reservoir, Alberta. *Journal of Fisheries Research Board of Canada*, 30, 557-560.

Turner TF, Dowling TE, Osborne MJ, McPhee MV, Broughton RE, Gold JR (2009) Microsatellite markers for the endangered razorback sucker, Xyrauchen texanus, are widely applicable to genetic studies of other catostomine fishes. Conservation Genetics, 10, 551-553.