HIERARCHICAL STRUCTURE AND DIVERSITY IN A DENDRITIC LAKE TROUT (*SALVELINUS NAMAYCUSH*) SYSTEM IN NORTHERN LABRADOR

by

Gregory Richard McCracken

Submitted in partial fulfilment of the requirements
for the degree of Master of Science

at

Dalhousie University
Halifax, Nova Scotia
September 2012

DALHOUSIE UNIVERSITY

Department of Biology

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled "HIERARCHICAL STRUCTURE AND DIVERSITY IN A DENDRITIC LAKE TROUT (*SALVELINUS NAMAYCUSH*) SYSTEM IN NORTHERN LABRADOR" by Gregory Richard McCracken in partial fulfilment of the requirements for the degree of Master of Science.

Dated: September 24, 2012

Supervisor: _____

Readers: _____

_____

DALHOUSIE UNIVERSITY

DATE: September 24, 2012

AUTHOR: Gregory Richard McCracken

TITLE: HIERARCHICAL STRUCTURE AND DIVERSITY IN A DENDRITIC LAKE TROUT (*SALVELINUS NAMAYCUSH*) SYSTEM IN NORTHERN LABRADOR

DEPARTMENT OR SCHOOL: Department of Biology

DEGREE: MSc CONVOCATION: May YEAR: 2013

_____

Signature of Author

To my wife and parents. Your support made this possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

I examined the relationship between landscape attributes and molecular genetic diversity and differentiation among lake trout (*Salvelinus namaycush*) populations inhabiting a hierarchically structured dendritic freshwater system in northern Labrador, the Kogaluk River system. Samples were collected from a total of 10 lakes which differed in size, elevation, level of connectivity, and position within the system. STRUCTURE analysis provided evidence of significant population structure within the system likely attributed to a varying degree of asymmetric gene flow. Gene flow estimates were generally low with some exceptions. Gene flow appears to be influenced by the presence of waterfalls as well as geographic distance. Isolation by distance tests coupled with decomposed pairwise regression analysis suggest there is a significant influence of geographic distance on population differentiation. Mantel testing also showed that population differentiation is significantly correlated with the position of waterfalls. Estimates of effective population size reveal significantly smaller population sizes in headwater than in non-headwater lakes, and this pattern is not attributed to lake size. Effective size estimates also suggest that the populations south and west of the Kogaluk River fjord are significantly smaller than those in the north.

LIST OF ABBREVIATIONS USED

AIC	Akaike's Information Criteria
AIC_c	Corrected Akaike's Information Criteria
A_r	Allelic Richness
cm	Centimetre
dNTP	Deoxyribonucleotide triphosphate
H_e	Expected Heterozygosity
H_O	Observed Heterozygosity
K	Number of Populations
km	Kilometer
$\widehat{m}$	Estimate of Migration Rate
m	Metre
MCMC	Markov Chain Monte Carlo
N	Sample Size
N_b	Effective Number of Breeders
N_e	Effective Population Size
$\widehat{N}_e$	Estimate of Effective Population Size
ng	Nanogram
PCA	Principal Coordinates Analysis
PCR	Polymerase Chain Reaction
SNP	Single-Nucleotide Polymorphism
WF	Waterfall
µL	Microlitre
µM	Micromolar

ACKNOWLEDGEMENTS

I thank my supervisor Dr. Daniel Ruzzante for support and guidance during this project, and for taking a chance on me right out of undergrad without prior formal training. I also thank Dr. Denis Roy who has served as a source of inspiration and new ideas for this project as well as Abby van der Jagt for my initial laboratory training. I also thank Dr. Paul Bentzen and Dr. Mark Johnston for serving on my committee and reviewing my thesis. In addition, I thank Dr. Anthony Gharrett for reviewing my thesis.

The Newfoundland and Labrador Department of Environment and Conservation provided the funding for this project. I thank members of the Department of Environment and Conservation especially Rob Perry and Don Keefe for assistance in the field and helping me understand the unique landscape that is Northern Labrador in addition to better understanding the unique life history of lake trout in this region. I also thank Bill Green and Shawn Avery for fieldwork assistance as well as Lorne Pike of Universal Helicopters who flew me safely place to place.

A big thank you to everyone in the Marine Gene Probe Lab at Dalhousie University, especially Ian Paterson who assisted in a variety of equipment related issues in addition to helping me tame the spiteful "PCR gods," Julie Rivard, Lyndsey Baillie, Meghan McBride, Devon Johnstone, Cecelia Carrea, and Vicky Yaroshewski for sharing ideas.

I would also like to thank my wife Danielle McCracken for her constant support during this project. She has been a constant source of encouragement and motivation. Lastly, I thank my parents Ann and Dwight McCracken who, in addition to my wife,

have been my greatest motivators. Their style of parenting encouraged me to focus on education and allowed me to develop my own perspective on life.

CHAPTER 1: INTRODUCTION

Understanding the spatial distribution of genetic diversity or genetic structure of populations and the factors affecting this distribution is a fundamental goal in evolutionary and conservation biology. Genetic structure has been estimated in countless natural systems and organisms worldwide ranging from humans (Rosenberg *et al.* 2002), to fishes (Roy *et al.* 2012), to insects (Pizarro *et al.* 2008; Seyahooei *et al.* 2011). Most studies typically estimate structure using neutral genetic markers including microsatellites. In addition, many studies are now inferring relationships between environmental factors and genetic structuring. This specific field of study, known as landscape genetics, aims to understand patterns of gene flow and local adaptation in the context of the landscape (Manel *et al.* 2003; Storfer *et al.* 2007), and uses the interaction between the physical landscape and life history traits to account for current levels of genetic structuring within a system. Genotypic information at multiple neutral markers can be used to cluster individuals based on their genetic composition (Falush *et al.* 2003), and can also be combined with environmental and landscape data in general linear models to assess the effects of these variables on patterns of diversity (Foll and Gaggiotti 2006).

It is important that neutral markers are used in a landscape genetics approach as these markers yield unbiased estimates of population structure, gene flow, and genetic variation (Schwartz *et al.* 2010; Manel *et al.* 2003). This however, does not discount the importance of non-neutral markers in conservation genetic studies. Non-neutral markers are very important for studying local adaptation (Reed and Frankham 2001; see Limborg

et al. 2012 for a recent example). In the context of this project however, it is important to note that markers under selection can bias estimates of gene flow and population structure (Schwartz *et al.* 2010) and thus lead to erroneous inferences.

The spatial distribution of populations in a geographically fragmented system can create a complex distribution of genetic diversity within a system, making inferences regarding population structure difficult. Freshwater systems are unlikely to conform to ideal models such as the island model (Wright 1931), or the linear stepping-stone model (Kimura and Weiss 1964), but likely fit the dendritic model, especially in areas consisting of elevational gradients (Morrissey and de Kerckhove 2009). In a dendritic system, two or more headwater source populations converge into a downstream population. Two or more of these downstream populations then themselves converge into a further downstream sink population. These dendritic systems are often associated with changes in elevation which can result in asymmetries in gene flow because of the influence of gravity (Morrissey and de Kerckhove, 2009). Such asymmetries can have significant influence over the distribution of genetic diversity with downstream populations, typically exhibiting higher genetic diversity than headwater populations (Morrissey and de Kerckhove 2009).

The focal species of this study, *Salvelinus namaycush* has been researched in a variety of large lakes (Giroux *et al.* 2009; Northrup *et al.* 2010), with an emphasis on the Great Lakes region due in part to the drastic reduction in lake trout numbers by sea lamprey (Martin and Olver 1980). However, there are relatively few studies on this species in its northern habitat, and a dearth on population structure in northern fragmented habitats. This concentration of effort in a limited segment of the species range

may skew our understanding of the ecology of the species as a whole. I should also note that studies of the species in its southern range are often subject to issues regarding anthropogenic influences such as overfishing, and the authors generally acknowledge this lack of pristine status. These studies have shown that in general, lake trout exhibit a broad range extending from Alaska to Nova Scotia, and into of the parts of the north-eastern United States (Scott and Crossman 1973) where the species is generally considered a top predator. *Salvelinus namaycush* is described as a nomadic species, and individuals have been shown to travel up to hundreds of kilometers in larger lakes such as Lake Superior (Martin and Olver 1980). This tendency for migration has been attributed to a variety of factors, from feeding to oxygen content (Martin and Olver 1980).

My study system, located in Northern Labrador, is a dendritically structured freshwater system closed to oceanic immigrants (Anderson 1985). Because of this system's northern location and lack of human influence, it is considered pristine. Several pairs of lakes in this system are asymmetrically isolated from one another due to the presence of waterfalls. Understanding how landscape variables influence genetic diversity is one objective of this study, and I will be testing the effects of distance, elevation, slope, the number of intermediate lakes, and the presence of waterfalls on genetic diversity within this system.

Although studies of the population structure of lake trout in similar small fragmented habitats are unavailable, studies from landlocked arctic charr suggest an extensive degree of population structure within systems (Wilson *et al.* 2004; Bernatchez *et al.* 2002; Primmer *et al.* 1999). Wilson *et al.* (2004) noted one particular extreme example in which the F_{ST} between a single pair of lakes 50 km apart was estimated to be

0.627. In this same study, global estimates of F_{ST} from different geographical regions ranged from 0.173 to 0.263, while Primmer *et al.* (1999) estimated a global F_{ST} of 0.360. A recent study by Northrup *et al.* (2010) measured the genetic structuring within a range of western populations of lake trout and estimated F_{ST} values ranging from 0.014 between two connected lakes of the same watershed, to 0.616 between two lakes in different watersheds, separated by 780 km and likely belonging to different glacial lineages. A study by Wilson and Hebert (1998) on postglacial dispersal of lake trout suggests that lake trout from my region of study (Northern Labrador), originate from a single glacial refugium, the Atlantic, thus I do not expect the same high values of F_{ST} evident in the Northrup *et al.* (2010) study between sample sites from different refugia.

With the unique life history of lake trout, the existence of physical barriers, as well as the dendritic nature of the system, the concept of effective population size may help disentangle some of the issues related to population structure and gene flow. In a dendritically structured system with possible asymmetries in gene flow, genetic variation is expected to be higher in downstream than headwater populations (Morrissey and de Kerckhove 2009; Junker *et al.* 2012; Caldera and Bolnick 2008), which would therefore reflect lower effective population size estimates in the headwaters. In essence, we can use estimates of effective population size as a measure of diversity between various lakes within a system.

Due to the dendritic spatial distribution of the lakes in this system, the presence of large asymmetric barriers to gene flow, and the likelihood of asymmetric gene flow, I expect higher levels of diversity in downstream populations than in upstream populations. In addition, I expect that the waterfalls present in this system will play an important role

in genetically structuring this system, whereby populations for which connections are mediated by one or more of these waterfalls will likely represent independent populations.

CHAPTER 2: MATERIALS AND METHODS

2.1 System and Sampling Layout

The Kogaluk River system comprises a series of hierarchically interconnected lakes on the barren grounds of Northern Labrador (Figure 2.1.1). The lakes are located north and south of a major fjord lake (Cabot Lake), into which they all drain and which is part of the Kogaluk River. Five major waterfalls are present in the system (Figure 2.1.1,WF1-5), the last one of which lies approximately 9 km upstream of the mouth of the Kogaluk River, blocking immigration from the ocean (see Figure 2.1.1, WF5; Anderson 1985). Fish populations in this system are therefore landlocked. All other labelled waterfalls in Figure 2.1.1 are considered complete barriers to (upstream) migration (Anderson 1985; see Table 2.1.1). Due to the northern geographical location and distance from settled areas, the system is largely free from anthropogenic influence.

Salvelinus namaycush were collected from 11 lakes in the system between 2006 and 2011. Sample size for *S. namaycush* in one of these lakes (Mistastin Lake) was too low for population level analysis, leaving a total of 10 lakes for this study. Nine of these lakes are located on the barren grounds, six north, and the remaining three either west or south of the Kogaluk River and Cabot Lake (Figure 2.1.1). Lake trout were sampled using variable sized gill nets (1.27 cm to 8.89 cm diagonal). Fish were measured (total length) and weighed, and otoliths and adipose fin clips were taken *in situ*. Fin clips were stored in 95% ethanol. In total, 567 lake trout were sampled from the various lakes (Table 2.1.1).

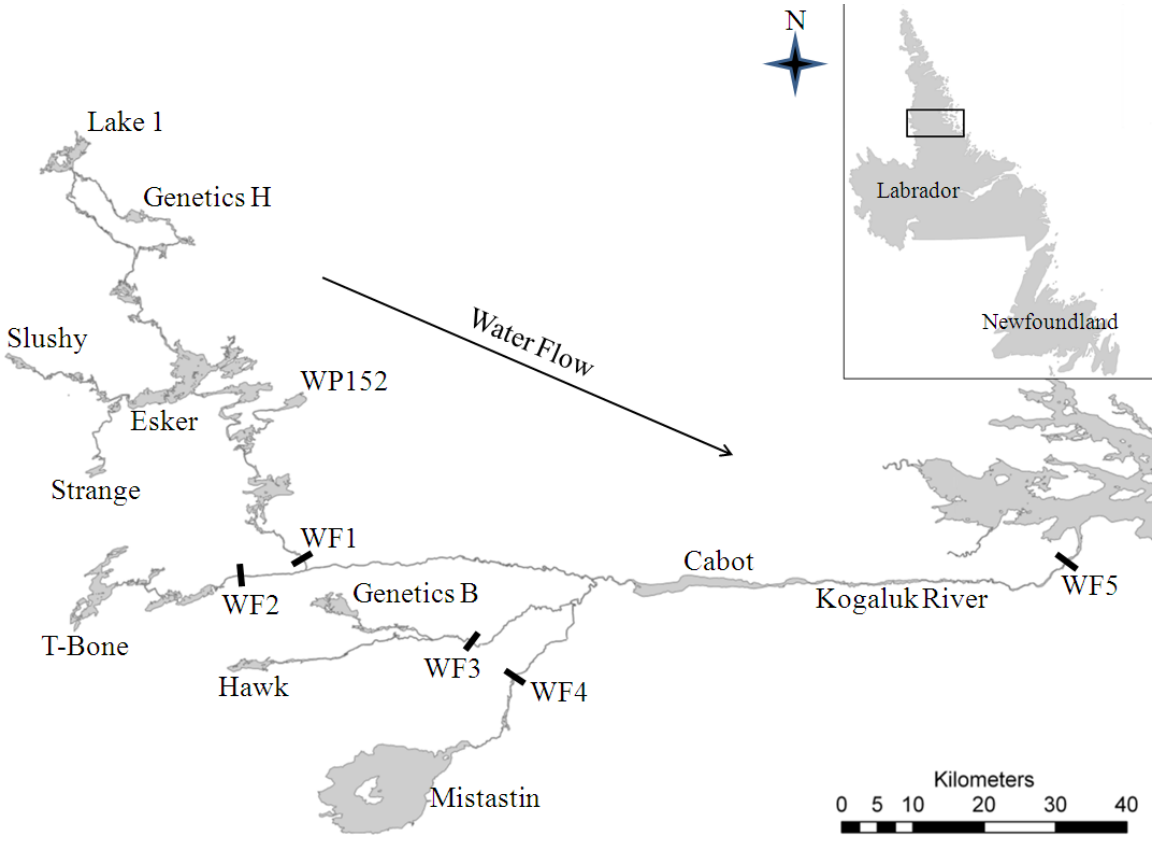


Figure 2.1.1. Represented in this figure are the various sampled lakes and their connecting water bodies. The Kogaluk River drains into the Atlantic Ocean via Voisey Bay. WF1 to WF5 represent the approximate locations of major waterfalls (gene flow barriers) within the system.

Table 2.1.1. Waterfall number on figure 2.1.1, height of waterfall in meters, waterfall angle in degrees, and the type of barrier to migration for each of the 5 main waterfalls in the system. Source: Anderson (1985).

Waterfall number	**Height (m)**	**Falls Angle (°)**	**Barrier to upstream migration**
WF1	15.3	90	Complete
WF2	12.2	90	Complete
WF3	5.4	90	Complete
WF4	5.4	90	Complete
WF5	9.2	60-90	Complete

2.2 DNA Extraction, Amplification, and Genotyping

Adipose fin tissue samples were digested with Proteinase K (Bio Basic Inc., Markham, Ontario) at 55°C for approximately 8 hours. The DNA was extracted from the resulting digest using a Glassmilk protocol (Elphinstone *et al.* 2003) with a Perkin Elmer Multiprobe II plus liquid handling system (Perkin Elmer, Waltham, Massachusetts). Random selections of DNA samples were electrophoresed on 2% agarose gel and compared against a size standard in an effort to ensure sufficient quantity of DNA for subsequent polymerase chain reactions. A suite of 13 loci were chosen on the basis of their polymorphism and ease of scoring, SSA85 (O'Reilly *et al.* 1996)*, SCO215, SCO202* (Dehaan *et al.* 2005)*, OGO1A* (Olsen *et al.* 1998)*, SNAMSU06, SNAMSU12, SNAMSU02* (Rollins *et al.* 2009), *SFO334* (Perry *et al.* 2005)*, OTSG253b, OTSG83b* (Williamson *et al.* 2002)*, SCO102, SCO107* (Sewall Young unpublished)*,* and *OMM1105* (Rexroad *et al.* 2002). Eleven of these loci were arranged into 4 multiplex primer panels, following the procedures of Qiagen master mix (Qiagen Inc., United States); the remaining two loci were run individually (See Appendix A for multiplex primer panel arrangement and annealing temperatures).

Multiplex PCRs were amplified in 5µL reactions and in general contained 2.5µL of Qiagen Master Mix (Qiagen Inc., United States), 0.5µLQiagen Q-solution, 0.5µL RNAse free water, 0.1-0.2µM fluorescently labelled forward primer, 0.1-0.2µM unlabelled reverse primer, and approximately 50 ng of DNA. Single primers were also amplified in 5µL reactions containing 2.35µL of RNAse free water, 0.5µL 10X reaction buffer (Bio Basic Inc., Markham, Ontario), 0.5µL $MgSO_4$ (Bio Basic Inc., Markham, Ontario), 0.1µM fluorescently labelled forward primer, 0.1µM unlabelled reverse primer,

0.25U TSG polymerase (Bio Basic Inc., Markham, Ontario), 200 µM dNTPs (Bio Basic Inc., Markham, Ontario), and approximately 50 ng of DNA. The thermocycler profile for multiplex reactions consisted of 15 min at 95°C, 30 cycles of 94°C for 30 s, primer specific annealing temperature for 1 min, 72°C for 1 min, followed by a final extension at 60°C for 30 min. For single primer reactions, the thermocycler profile consisted of 95°C for 3 min, 30 cycles of 94°C for 30 s, primer specific annealing temperature for 45 s, 72°C for 1 min, and a final extension at 72°C for 5 min.

PCR products for multiplex reactions were diluted between 1:10 and 1:20 with formamide depending on intensity, while single primer PCRs were diluted 1:20 with formamide. The diluted product was then imaged on a set of 8 Licor 4200/4300 DNA analyzers (LICOR, Lincoln, Nebraska). Individual genotypes were collected using SAGA Automated Microsatellite Software 3.3 (LICOR, Lincoln, Nebraska) followed by manual checking to ensure scoring accuracy. Genotypes were then run through MICROCHECKER 2.2.3 (van Oosterhout *et al.* 2004) to assess the presence of null alleles, or scoring inconsistencies.

2.3 Analysis

2.3.1 Test for Selection

Detection of molecular markers under selection is critical to the analysis of population structure. Divergent selection leads to over-inflated F_{ST} estimates (Luikart *et al.* 2003), potentially biasing inferences on gene flow, population structure, and proper assignment of individuals to populations. I thus tested my microsatellite loci for neutrality using the LOSITAN selection workbench (Antao *et al.* 2008). The test implemented in LOSITAN is based on the fact that for each level of expected heterozygosity, there will be an expected distribution of F_{ST} values (Antao *et al.* 2008). If the F_{ST} estimate at a particular marker is above or below the expected range, the marker is assumed to be experiencing positive or balancing selection respectively, and should be excluded from further analyses that assume neutrality (Luikart *et al.* 2003). I conducted my analyses of selection using 300,000 permutations utilizing the stepwise mutation model and with a sample size of 50, which is approximately the average sample size per lake (56.7) in this study.

2.3.2 Population Structure Analysis

Population structure was examined using the Bayesian approach implemented in STRUCTURE 2.3.3 (Hubisz *et al.* 2009). The analysis was conducted hierarchically. I first examined the entire data set and each identified cluster was then independently subjected to further STRUCTURE analysis. This process was continued on individual clusters until no further evidence of population structure was detected. I estimated the most likely number of clusters based on the Evanno methodology (Evanno *et al.* 2005) implemented in STRUCTURE HARVESTER v0.6.92 (Dent *et al.* 2012). Each independent STRUCTURE run was conducted using 10 separate iterations where each iteration was run for 1,000,000 replications with an initial burn-in of 200,000. The results of these 10 separate replications were then combined into a single population output using the program CLUMPP 1.1.2 (Jakobsson and Rosenberg 2007) for the most likely number for K, and visualized using the program DISTRUCT 1.1 (Rosenberg 2004).

2.3.3 Genetic linkage, Hardy-Weinberg equilibrium, Population Genetic Diversity, Analysis of Molecular Variance

Genotypic linkage and conformity to Hardy-Weinberg equilibrium were both tested using Arlequin 3.5.1.2 (Excoffier and Lischer 2010). Genotypic linkage between all pairs of loci (per population) were estimated utilizing 20,000 permutations, while conformity to Hardy-Weinberg equilibrium was tested for each locus and population using 1,000,000 permutations and 100,000 dememorization steps. Results were then subjected to sequential Bonferroni correction (Rice 1989) to maintain an overall type 1 error probability at 0.05. Per population observed (H_o) and expected (H_e) heterozygosities were also estimated with Arlequin 3.5.1.2. Allele frequencies and allelic richness were estimated using F_{STAT} (Goudet 2001). Private alleles were identified with GenAlEx 6.4 (Peakall and Smouse 2006). Genetic differentiation (F_{ST}) was estimated using MSA 4.05 (Dieringer and Schlötterer 2003) using 100,000 individual MCMC replicates. Analysis of molecular variance was conducted using Arlequin 3.5.1.2 utilizing 50175 permutations both on the full dataset, and subsets consisting of only northern or southern lakes. Each AMOVA was run locus by locus to account for variations in the degrees of freedom per locus caused by missing data. By correcting for these missing data, I was able to get a more accurate estimate (Arlequin 3.5 manual, Excoffier and Lischer 2010).

2.3.4 Identification of Migrant Individuals, Effective Population Size and Gene Flow Estimation

Potential immigrants were identified with GeneClass2 (Piry *et al.* 2004), which uses a Bayesian method developed by Ranalla and Mountain (1997) for detecting migrants using a Monte-Carlo re-sampling method (Paetkau *et al.* 2004). Individuals identified as potential migrants were removed from the dataset prior to the estimation of effective population size ($\widehat{N}_e$). $\widehat{N}_e$s were then estimated for all lake populations on the basis of the linkage disequilibrium method implemented in LDNe (Waples and Do 2008). The program implements a bias correction for cases when the sample sizes are smaller than the actual effective population size (Waples 2006). Estimation via LDNe was conducted using the (P_{crit}) critical value (allele frequencies greater than) 0.02, as described by Waples and Do (2010) as all my sample sizes were >25, with 95% confidence intervals generated via jackknifing between pairs of loci. Gene flow was then estimated using BayesAss+ (Wilson and Rannala 2003) which uses a Bayesian framework to infer recent migration rate. BayesAss+ was run for 5,000,000 iterations with an initial burnin of 1,000,000, all other variables used default values.

2.3.5 Isolation by Distance

In an effort to determine which factors are influencing differentiation a Mantel test of isolation by distance test was conducted using the Mantel function available in GenALEx, utilizing 9,999 permutations. Input data for this program were calculated using MSA 4.05 for pairwise F_{ST} estimates and ArcGIS Desktop: Release 10 (ESRI 2011) for pairwise waterway distances. F_{ST} values were linearized [(FST/(1-FST)] prior to testing distance (Rousset 1997). Due to the presence of physical barriers (waterfalls) in this system, simply testing for a relationship between genetic and geographic distance by means of correlation can mask populations which may be uniquely influenced by some sort of environmental barrier (Koizumi *et al.* 2006). As a result, a decomposed pairwise regression analysis was conducted, whereby outlier populations that deviate from the IBD pattern can be determined using the mean and 95% confidence interval values obtained from the residuals of all pairwise comparisons between genetic and geographic distance. Populations considered "putative outliers" were those for which the 95% confidence interval of their residuals did not include 0. Initial putative outlier populations were subsequently removed one by one, and the analysis was repeated. This process continued until no further putative outliers were found. This created several models (combinations of removed "putative outlier" populations). These models were then subjected to analyses using Akaike's information criteria (AIC) which allows for the identification of "true outliers" from the combinations of "putative outliers" based on AIC value (see Koizumi *et al.* 2006 for additional method details).

2.3.6 Estimating Effects of Various Landscape Factors

A number of landscape factors have the potential to be quite influential with regards to population differentiation within this system. In addition to the IBD test described above, I examined the effects of the presence or absence of migratory barriers (waterfalls assumed to be complete barriers to upstream migration). Binary data were used to indicate presence or absence of waterfalls (1: waterfall present; 0: no waterfall). I also estimated the correlation between genetic differentiation and elevational difference between pairs of lakes, as well as average waterway slope (simply calculated as the change in elevation between lakes divided by the waterway distance; Stelkins *et al.* 2012), and the number of intermediate lakes between sampled lakes. Following methods implemented by Kanno *et al.* (2011), I used a series of Mantel tests to examine possible correlations between the matrices comprised of different environmental variables and the pairwise F_{ST} matrix. The tests were conducted using the vegan: Community Ecology Package (Oksanen *et al.* 2012) available for R instead of GenALEx (see: IBD section) because of the relative ease of inputting multiple independent matrices. These Mantel tests were conducted using both the full dataset and a subset - the northern lakes, to account for any influence or bias from environmental factors. Significance levels for all analyses were kept at $\alpha=0.05$.

The effect of landscape characteristics on population structure was tested using GESTE (Foll and Gaggiotti 2006). GESTE estimates the influence of environmental (landscape) factors on genetic diversity by calculating population F_{ST} values, and comparing these values with environmental factors using a general linearized model. The same five environmental factors were used as with the Mantel tests, however they

required transformation to fit program standards – a single value for each factor per population. The mean value of all pairwise comparisons was used for geographic distance, slope, elevation difference, and the number of intermediate lakes. For each sampled lake, I accounted for the effect of waterfalls by calculating the number of sampled lakes which were physically connected without waterfall barriers.

CHAPTER 3: RESULTS

3.1 General Statistics

There was no evidence of null alleles reported from MICROCHECKER, nor evidence of scoring errors as a result of stutter in any of the loci. Tests using LOSITAN showed potential evidence of positive selection for one out of 13 loci (i.e. locus SCO215; Figure 3.1.1. Recent studies suggest that detecting loci under selection by means of an outlier approach, as incorporated into LOSITAN, may be problematic due to the detection of false positives especially when there is evidence of hierarchical population structure (Narum and Hess 2011). As a result, I realize that this marker, SCO215, may not be under selection. However, to eliminate any concern regarding the neutrality of this marker, I decided to exclude it from further analysis. Of the remaining 12 loci, observed (H_O) and expected (H_E) heterozygosities ranged from 0.433 to 0.569 and from 0.41 to 0.57, respectively. Allelic richness (A_R) ranged from 3.50 to 4.76 (Table 3.1.1). There was no evidence of deviation from Hardy-Weinberg equilibrium within any of the 10 putative populations (see Table 3.1.1).

Linkage disequilibrium tests performed with the 12 neutral loci showed that one locus pair (Otsg83b and Sco107) consistently appeared linked in each of the 10 populations (see Table 3.1.1 and Appendix B). Sequential Bonferroni correction using a dataset with 12 loci would allow 3 linked pairs at α=0.05 to be present per population due to chance alone (66 comparisons per population multiplied by the error rate of 0.05). However, the same two loci appeared linked consistently (a result that is unlikely due to chance alone), thus I assumed that these two loci are indeed linked. Such linkage could

bias results including the assessment of population structure (Kaeuffer *et al.* 2007); therefore one of the loci had to be excluded.

I preliminarily tested for the potential effect of this linkage on the estimation of effective population size prior to the removal of first generation migrants. $\widehat{N}_e$s with all 12 neutral loci were substantially lower than after exclusion of either one of these two linked loci. In most cases, eliminating either of these loci resulted in > 100% increase in $\widehat{N}_e$, some significantly more (Table 3.1.2). I thus eliminated locus Otsg83b from further analysis on the basis of its failure rate.

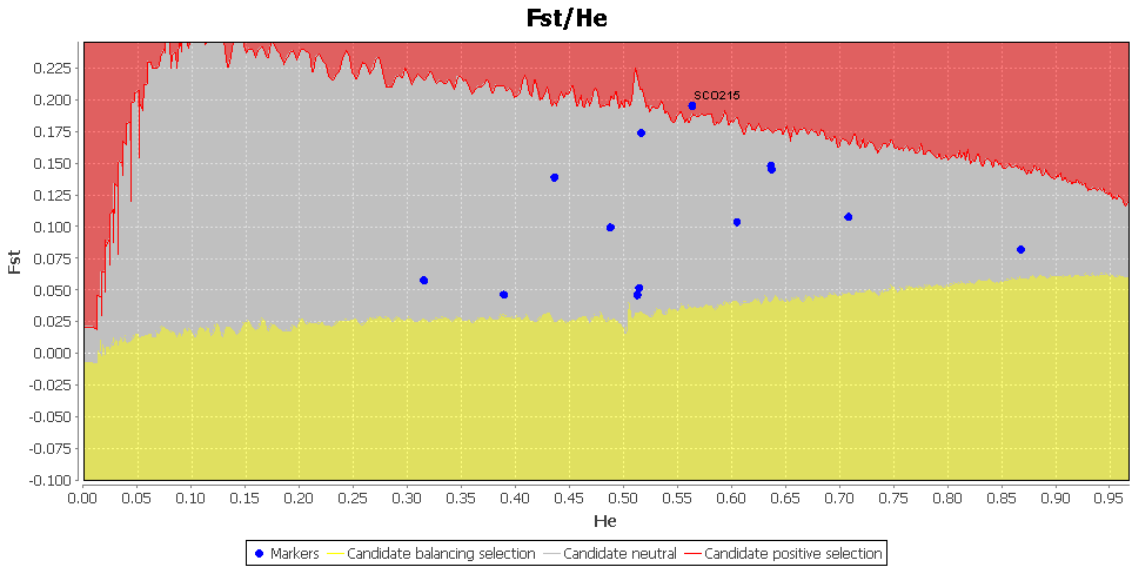


Figure 3.1.1. LOSITAN results after analysis for microsatellite markers under selection. The red shaded area represents potential positive selection, the yellow area represents potential balancing selection, and the gray area indicates the region of selective neutrality. The blue points represent each of the 13 microsatellite markers. Locus SCO215 is within the red shaded area (positive selection – higher than expected F_{ST} for the estimated level of heterozygosity) while all other loci exhibited F_{ST} values consistent with selective neutrality.

Table 3.1.1. Lake, headwater (HD) versus non-headwater (NHD), sample size, Latitude and Longitude, elevation in meters above sea level (MAS), expected and observed heterozygosity (H_e, and H_o, respectively), allelic richness (A_r), presence or absence of population linkage disequilibrium (LD), conformance to Hardy-Weinberg equilibrium (HWE), and maximum measure depth in meters.

Lake Name	Lake Order	N	Latitude	Longitude	Elevation (MAS)	H_e	H_o	A_r	LD	HWE	Depth (m)	Area (Km^2)
Lake 1	Headwater	79	N 56° 40' 31.7"	W 64° 00' 07.5"	525	0.49	0.49	4.39	No	**Yes**	3.9	11.3
Genetics H	Non-headwater	81	N 56° 36' 13.7"	W 63° 52' 09.1"	512	0.5	0.51	4.1	No	**Yes**	6.5	2.81
Slushy	Headwater	50	N 56° 24' 56.2"	W 64° 06' 08.1"	464	0.5	0.53	3.71	No	**Yes**	15.3	2.99
Strange	Headwater	55	N 56° 17' 24.8"	W 63° 56' 53.4"	487	0.52	0.52	4	No	**Yes**	Unavailable	2.09
Esker	Non-headwater	48	N 56° 24' 53.4"	W 63° 40' 15.1"	431	0.48	0.47	4.58	No	**Yes**	Unavailable	49.84
WP152	Headwater	47	N 56° 22' 08.7"	W 63° 29' 30.5"	445	0.49	0.51	3.68	No	**Yes**	16.1	4.14
T-Bone	Headwater	41	N 56° 09' 09.7"	W 63° 56' 21.2"	468	0.52	0.51	4.76	No	**Yes**	Unavailable	19.76
Cabot	Non-headwater	54	N 56° 08' 27.9"	W 62° 37' 52.4"	60	0.57	0.57	4.67	No	**Yes**	Unavailable	25.39
Genetics B	Headwater	50	N 56° 06' 38.4"	W 63° 23' 18.9"	239	0.47	0.53	4.39	No	**Yes**	27	9.71
Hawk	Headwater	62	N 56° 02' 52.1"	W 63° 35' 54.4"	466	0.41	0.43	3.5	No	**Yes**	21	5.74

Table 3.1.2. Estimates of effective population size and the 95% confidence intervals illustrating the effect of including linked loci.

	Full Dataset		Without SCO107		Without OTSG83b	
Lake	**Effective Size**	**95% CI.**	**Effective Size**	**95% CI.**	**Effective Size**	**95% CI.**
Lake 1	93	36.4 - ∞	220	86.1 - ∞	208.1	83.8 - ∞
Genetics H	72.4	24.6 - ∞	203	75.4 - ∞	241.7	76.7 - ∞
Slushy	78.6	23.3 - ∞	230.6	37.9 - ∞	192.5	35.9 - ∞
Esker	544.7	31.3 - ∞	Undefined	141.5 - ∞	Undefined	102.7 - ∞
Strange	48.7	12.2 - ∞	205.2	67.6 - ∞	190.5	64.9 - ∞
WP152	68.8	17.1 - ∞	Undefined	100.0 - ∞	Undefined	102.5 - ∞
T-Bone	28.4	9.5 – 494.4	90.1	36.9 - ∞	65.2	30.7 – 458.1
Cabot	185.8	45.4 - ∞	Undefined	126.0 - ∞	2323	115.1 - ∞
Genetics B	46.2	24.8 – 133.4	47.5	26.3 – 125.9	50.4	26.9–155.5
Hawk	24.7	11.6 – 63.6	39.6	19.2 – 124.0	37.6	18.8–105.5

3.2 Population Structure Analysis, Genetic Diversity, and Gene Flow

An initial STRUCTURE analysis with 11 loci and using the entire dataset indicated K=2 (Figure 3.2.1a). Each of these population components was then examined separately resulting in K=5 and K=4 within each original pool, with *S. namaycush* in two lakes, Esker and WP152 being genetically indistinguishable (See Appendix C for likelihood plots). No further substructure was detected in any of the remaining lakes resulting in a total of nine genetic pools.

Pairwise F_{ST} estimates indicated all populations differed significantly from each other with the single exception of lakes Esker and WP152, supporting the results of STRUCTURE. The inferred relationships between sampled lakes, based on F_{ST} results, are shown in the Principal Coordinates Analysis illustrated in Figure 3.2.3.

Estimates of molecular variance were conducted at all levels of hierarchical population structure. At K = 2, the AMOVA results reveal that 4.28% of the total genetic variation is explained by differences among groups, while a further 8.31% is explained by differences among populations within groups, indicating that the group of lakes within the northern cluster vary genetically from the group of lakes in the southern cluster by approximately 12.59% (Table 3.2.1). The same test conducted solely on the northern cluster (with Lakes Esker and WP152 pooled) reveals that 5.23% of the genetic variation is explained among the five groups, while a further 0.71% is explained by variation among individuals within populations (Table 3.2.2). Finally, within the southern cluster, the individual lakes vary genetically from each other by approximately 14.11% (Table 3.2.3). All AMOVA results are significant ($p < 0.05$; Tables 3.2.1-3.2.3).

Pairwise gene flow estimates varied greatly within this system (Table 3.2.4 $\hat{m}$= 0.001 to $\hat{m}$= 0.288), with the majority of estimates suggesting little to no gene flow between lakes. The highest gene flow estimate is that from Esker Lake to Lake WP152 ($\hat{m}$= 0.288) which is consistent with STRUCTURE identifying these populations as one genetic cluster. This gene flow is unidirectional as the migration rate from WP152 to Esker is substantially less ($\hat{m}$= 0.002). Additional relatively high gene flow estimates are evident in this system and predominantly follow an asymmetric pattern from the northern lakes to Lake Genetics B south of the Kogaluk River Fjord. These rates range from $\hat{m}$= 0.003 to $\hat{m}$= 0.138 into Lake Genetics B. These results also appear consistent with STRUCTURE results, as genetically, individuals from Lake Genetics B appeared more similar to northern lakes than did other southern populations (Figure 3.2.1). However, these results are not consistent with the current geographical waterway corridors and waterfall positions.

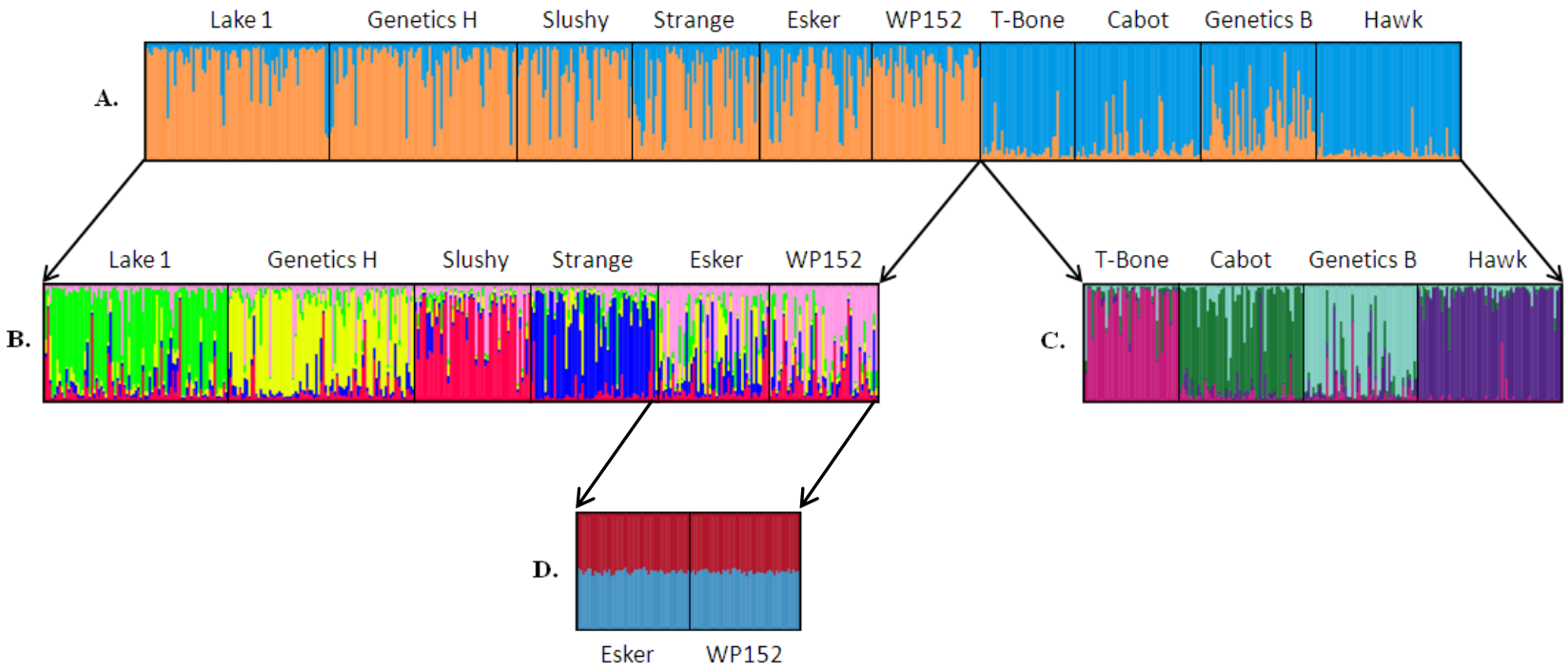


Figure 3.2.1. Hierarchical population structure analysis based on 11 neutral loci. *Salvelinus namaycush* were collected from 10 lakes in the Kogaluk River system of northern Labrador. Lines represent individual admixture coefficients (Q). (A) entire system indicating 2 separate clusters. (B) lakes from the initial grouping in (A) in which all individual lakes except Esker and WP152 exhibit unique population structure. (C) illustrates the differences between the lakes from the initial second grouping of (A). Finally, (D) indicates that lakes Esker and WP152 are genetically indistinguishable.

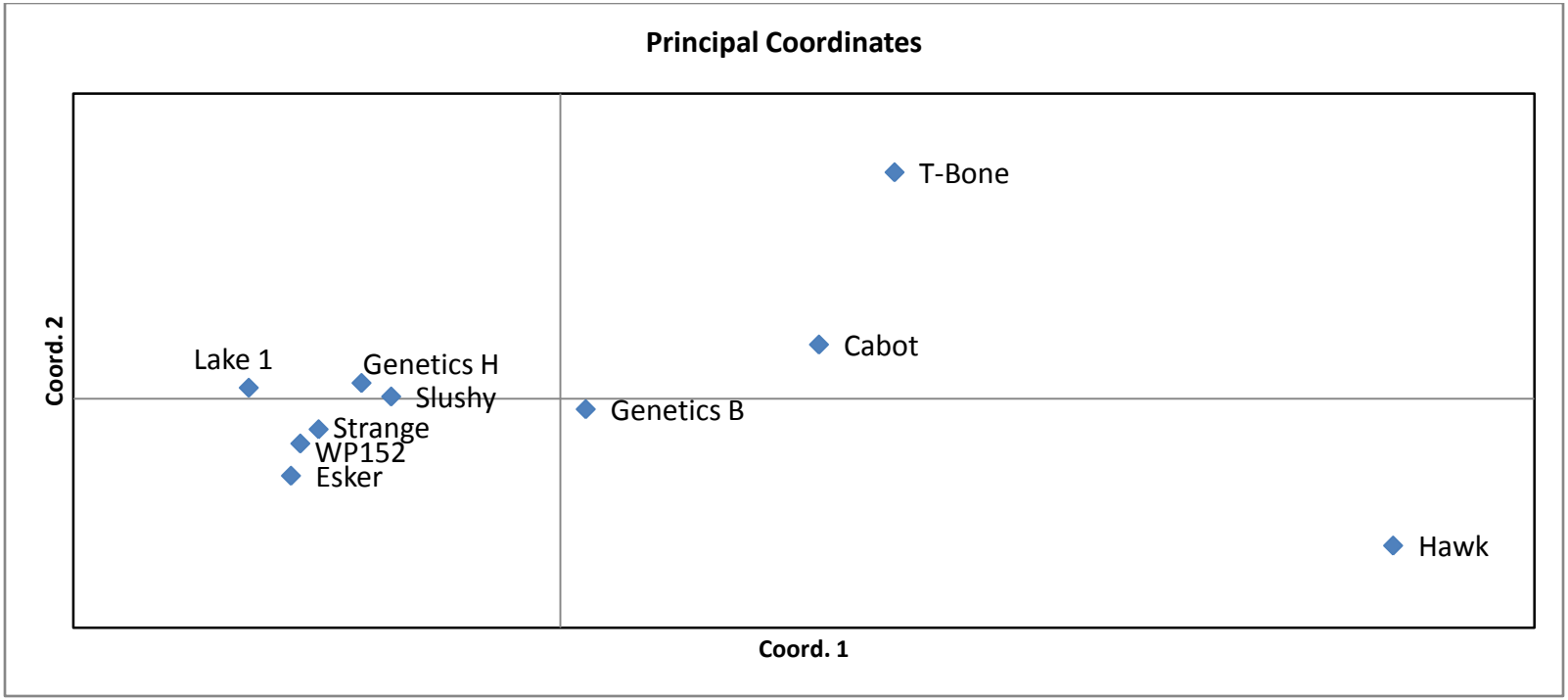


Figure 3.2.2. Principal coordinate analysis (11 neutral loci) conducted using the pairwise F_{st} estimates obtained with MSA. Coordinate 1 (the x axis) accounts for 35.67% of the variation, while coordinate 2 (y axis) accounts for a further 25.01% of the variation.

26

Table 3.2.1. AMOVA results for the full system indicating the amount of genetic variation explained by various groupings. This is an average degree of freedom computed by obtaining the mean value of locus by locus AMOVAs(*).

Source of Variation	d.f.	Sum of Squares	Variance components	Percent variation	P-value
Among groups	1.00	94.40	0.13	4.28	<0.001
Among populations within groups	8.00	236.63	0.26	8.31	<0.001
Within populations	*1046	2822.17	2.72	87.41	<0.001
Total	1055	3153.20	3.11		

Table 3.2.2. AMOVA results for the northern cluster lakes indicating the amount of genetic variation explained by various groupings. This is an average degree of freedom computed by obtaining the mean value of locus by locus AMOVAs (*).

Source of Variation	**d.f.**	**Sum of Squares**	**Variance components**	**Percent variation**	**P-value**
Among groups	4.00	101.42	0.15	5.23	<0.001
Among populations within groups	1.00	4.48	0.02	0.71	0.03
Within populations	*668	1821.58	2.74	94.06	<0.001
Total	673	1927.48	2.92		

Table 3.2.3. AMOVA results for the southern cluster lakes indicating the amount of genetic variation explained by various groupings. This is an average degree of freedom computed by obtaining the mean value of locus by locus AMOVAs(*).

Source of Variation	**d.f.**	**Sum of Squares**	**Variance components**	**Percent variation**	**P-value**
Among populations within groups	3	130.73	0.44	14.11	<0.001
Within Populations	*377	1000.59	2.67	85.89	<0.001
Total	380	1131.32	3.11		

Table 3.2.4. Pairwise gene flow estimated obtained via BayesAss+ 1.3 indicating little to no migration between most lakes with a few notable exceptions (in bold).

From/To	Lake 1	Genetics H	Slushy	Strange	Esker	WP152	T-Bone	Genetics B	Hawk	Cabot
Lake 1	0.988	0.002	0.001	0.008	0.008	0.010	0.003	**0.025**	0.001	0.002
Genetics H	0.001	0.938	0.001	0.004	**0.039**	0.007	0.002	**0.011**	0.001	0.002
Slushy	0.002	0.004	0.989	0.002	0.004	0.007	0.002	0.005	0.001	0.001
Strange	0.001	0.002	0.001	0.974	**0.010**	0.004	0.002	**0.031**	0.001	0.002
Esker	0.003	**0.045**	0.003	0.003	0.921	**0.288**	0.003	**0.138**	0.001	0.002
WP152	0.001	0.001	0.001	0.001	0.002	0.673	0.001	0.003	0.001	0.001
T-Bone	0.001	0.002	0.001	0.001	0.002	0.003	0.981	0.003	0.001	0.001
Genetics B	0.001	0.002	0.001	0.002	0.004	0.003	0.003	0.774	0.001	0.002
Hawk	0.001	0.003	0.001	0.003	0.004	0.003	0.002	0.007	0.993	0.003
Cabot	0.001	0.002	0.001	0.001	0.006	0.002	0.002	0.004	0.001	0.985

3.3 Isolation by Distance

An initial Mantel test suggested that there is evidence of Isolation by Distance when considering all 10 lake populations ($R^2 = 0.374$ and $p \leq 0.001$; Figure 3.3.1). Decomposed pairwise regression analysis revealed however, the presence of several putative outliers – initially Hawk Lake (Figure 3.3.2), followed by T-Bone Lake (Figures 3.3.2 to 3.3.3), and Cabot Lake (Figure 3.3.4) respectively. "True outliers" were identified from the various models of "putative outliers" using the AIC method (Table 3.3.1). My final analysis revealed that all three of Hawk, T-Bone, and Cabot Lakes were outliers based on the AIC_c value. Models which have a $\Delta AIC_c < 2$ are equally likely (Koizumi *et al.* 2006); however, I chose the model without these three lakes due to its relatively high R^2 value.

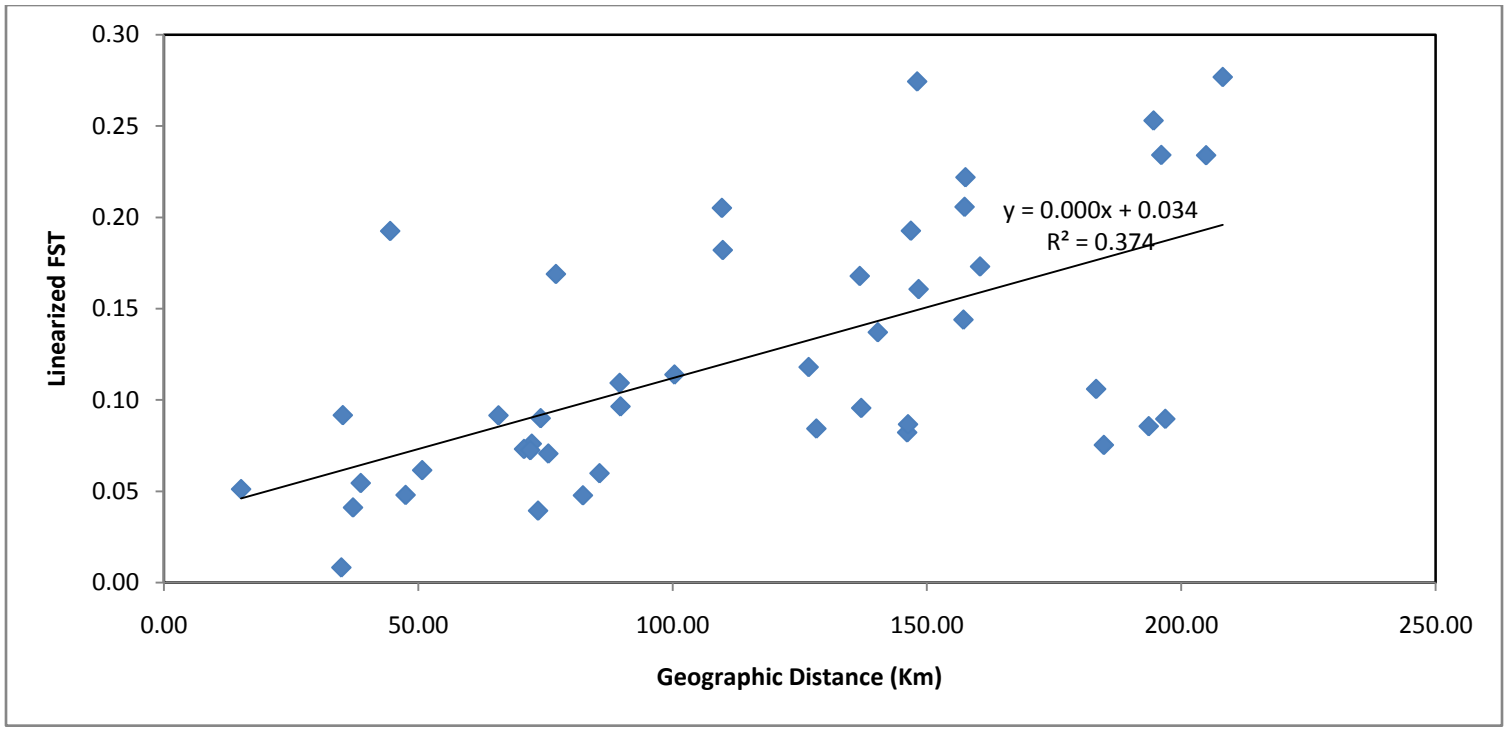


Figure 3.3.1. Correlation between linearized Fst and Geographic distance showing a significant positive correlation (p= 0.001).

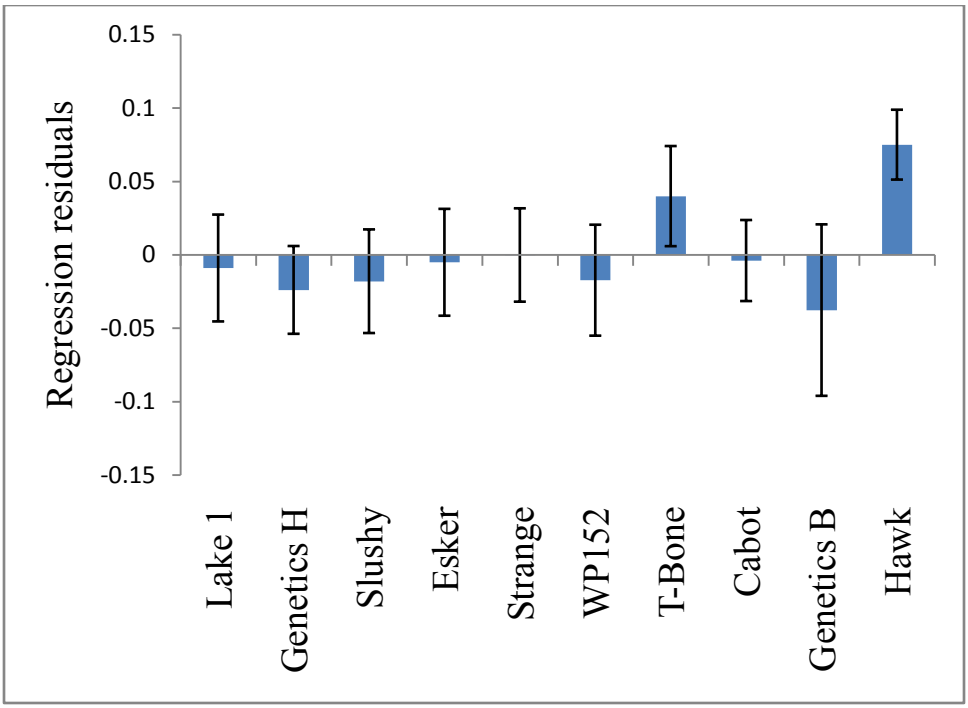


Figure 3.3.2. Analysis of 95% confidence intervals of residuals computed from the regression of all pairwise comparisons between genetic and geographic distance. T-Bone and Hawk lakes are putative outliers as their confidence intervals do not include 0.

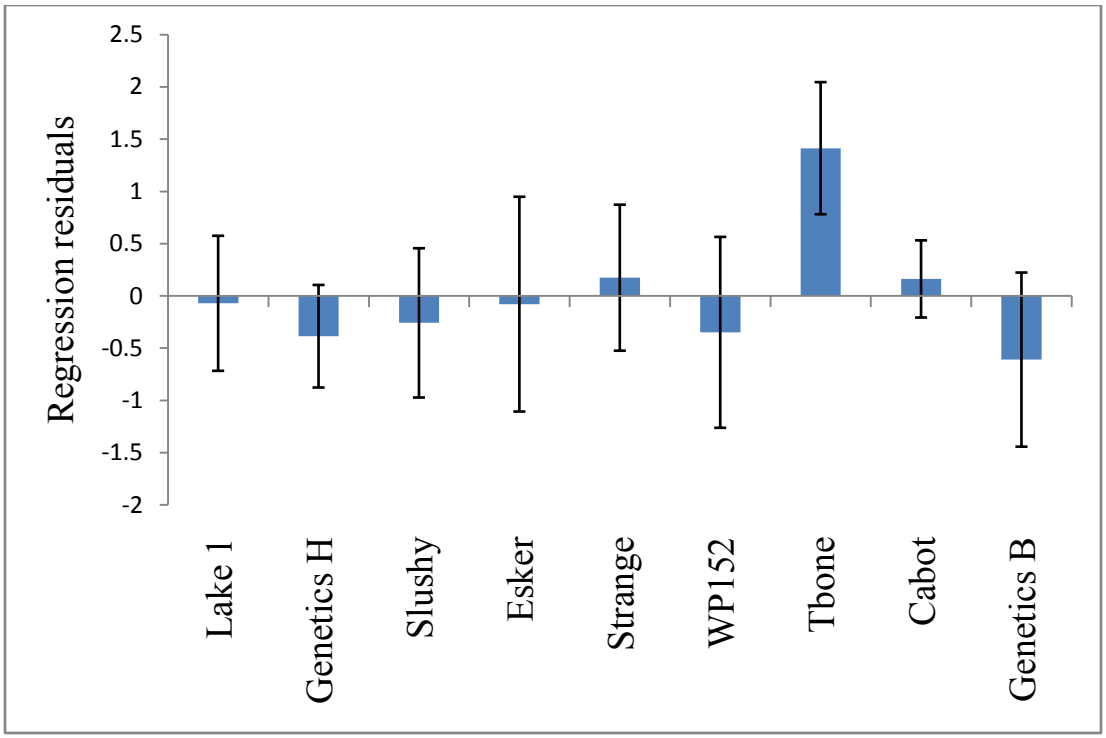


Figure 3.3.3. Analysis of 95% confidence intervals of residuals computed from the regression of pairwise comparisons excluding Hawk Lake. The removal of Hawk Lake from the analysis amplified the dissimilarity between T-Bone and all other Lakes, again pegging it as a putative outlier.

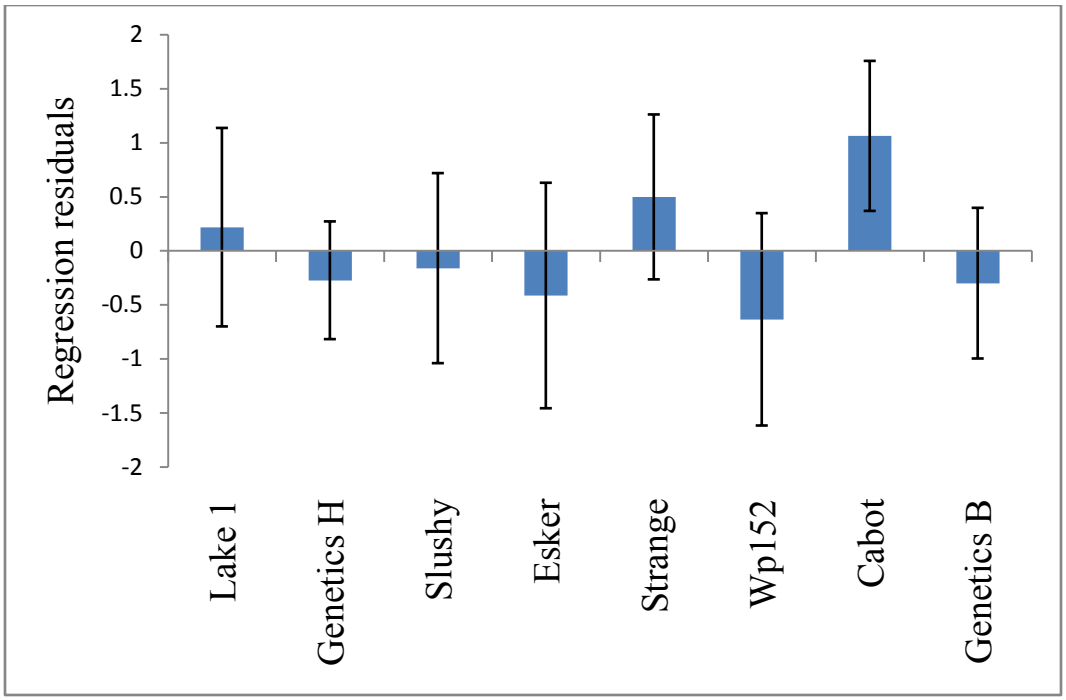


Figure 3.3.4. Analysis of the 95% confidence intervals of residuals computed from the regression of pairwise comparisons excluding both T-Bone and Hawk lakes. The removal of these two lakes amplified the distinctiveness of Cabot Lake which, prior to this analysis, had not been identified as an outlier.

Table 3.3.1. Various models composed of different combinations of putative outlier populations, their R^2 values, and corrected AIC (AIC_c) values.

Lakes Excluded	**R2 value**	**P value**	**AIC_c**	**ΔAIC_c**
Hawk, T-Bone, Cabot	0.388	< 0.003	-66.56	0
Hawk, T-Bone	0.318	< 0.002	-66.09	0.47
Hawk	0.277	< 0.001	-56.85	9.24
None	0.374	< 0.001	-51.93	4.92

3.4 Effects of Various Landscape Factors

In addition to distance, several other landscape factors were tested (Table 3.4.1). The only significant correlation obtained was between linearized F_{ST} and the influence of waterfalls for the dataset as a whole (see Table 3.4.1). No waterfalls exist within the subset of the six lakes located north of the Kogaluk River.

The results from the GESTE analysis suggest that the best model is that of the constant regression term (the null model) with a posterior probability = 0.645 (Table 3.4.2). The posterior probability associated with this model is substantially higher than the next highest alternative, constant regression term and distance, which has a posterior probability = 0.0696. These results indicate that none of the environmental factors are useful in predicting genetic diversity among lake trout populations in this system.

Table 3.4.1. Mantel test results conducted between various landscape factors to test for statistically significant correlation with genetic diversity. Results from both full data set and the subset of northern lakes are provided. The asterisk denotes tests which were statistically significant in the full dataset.

	Full Dataset		**Northern Lakes Only**	
Landscape attributes tested	**r value**	***p* value**	**r value**	***p* value**
Linearized FST and elevation	-0.1621	0.78	0.208	0.247
Linearized FST and geographic distance*	0.6116	0.0012	0.3191	0.119
Linearized FST and presence of waterfalls*	0.579	0.0025	NA	
Linearized FST and average slope	-0.2025	0.856	0.145	0.324
Linearized FST and number of intermediate lakes	0.2044	0.181	0.375	0.089

Table 3.4.2. GESTE output indicating the posterior probability for the first 10 (highest posterior probability) models indicating model 1, the constant, as the best model.

Model	Factors Included	Posterior Probability
1	Constant (null model)	0.645
2	Constant, Distance	0.007
3	Constant, Slope	0.061
4	Constant, Elevation	0.050
5	Constant, Waterfall Blockage	0.043
6	Constant, Number of Intermediate Lakes	0.041
7	Constant, Elevation, Slope	0.029
8	Constant, Waterfall Blockage, Slope	0.007
9	Constant, Waterfall Blockage, Distance	0.006
10	Constant, Waterfall Blockage, Elevation	0.006

3.5 Effective Population Size

Effective population size was estimated via LDNe (Waples and Do 2008) before and after removal of first generation migrants. Lakes containing first generation migrants had lower estimates of effective size, prior to the removal of migrants (Table 3.5.1). $\widehat{N}_e$ from individual lakes ranged from 38.3 to 2323 (Table 3.5.1). Mean estimates for Lakes Esker and WP152 were undefined, however, these populations represent a single genetic cluster and when individuals from both lakes were pooled, $\widehat{N}_e$ = 441.5. There was no evidence of a significant relationship between genetic cluster total lake area and mean $\widehat{N}_e$ ($p > 0.50$, Figure 3.5.1). There are 7 headwater and 3 non-headwater lakes in the system (Table 3.1.1). $\widehat{N}_e$s were significantly lower for *S. namaycush* populations inhabiting headwater than for those inhabiting non-headwater lakes (*t*-test $p < 0.05$).

Table 3.5.1. Effective population size estimates obtained via LDNe for the dataset prior to, and after the identification and removal of first generation migrants. Also included is the number of immigrants which were identified using GeneClass2.

Including First Generation Migrants			*Excluding Migrants*		
Sampled Lake	$\widehat{N}_e$	**95% CI**	$\widehat{N}_e$	**95% CI**	**# of Immigrants**
Lake 1	208	83.8 - ∞	208	83.8 - ∞	0
Genetics H	242	76.7 - ∞	977	102.9 - ∞	2
Slushy	192.5	35.9 - ∞	192.5	35.9 - ∞	0
Esker	Undefined	102.7 - ∞	Undefined	159.8 - ∞	2
WP152	Undefined	102.5 - ∞	Undefined	102.5 - ∞	0
Pooled Esker and WP152	572	110.7 - ∞	442	121 - ∞	1
Strange	191	64.9 - ∞	160	58.9 - ∞	1
T-Bone	65	30.7 – 458.1	66	32.2 – 378.0	1
Cabot	2323	115.1 - ∞	2323	115.1- ∞	0
Genetics B	50	26.9 – 155.5	50	26.9 – 155.5	0
Hawk	38	18.8 – 105.5	38	19.2 – 109.1	1

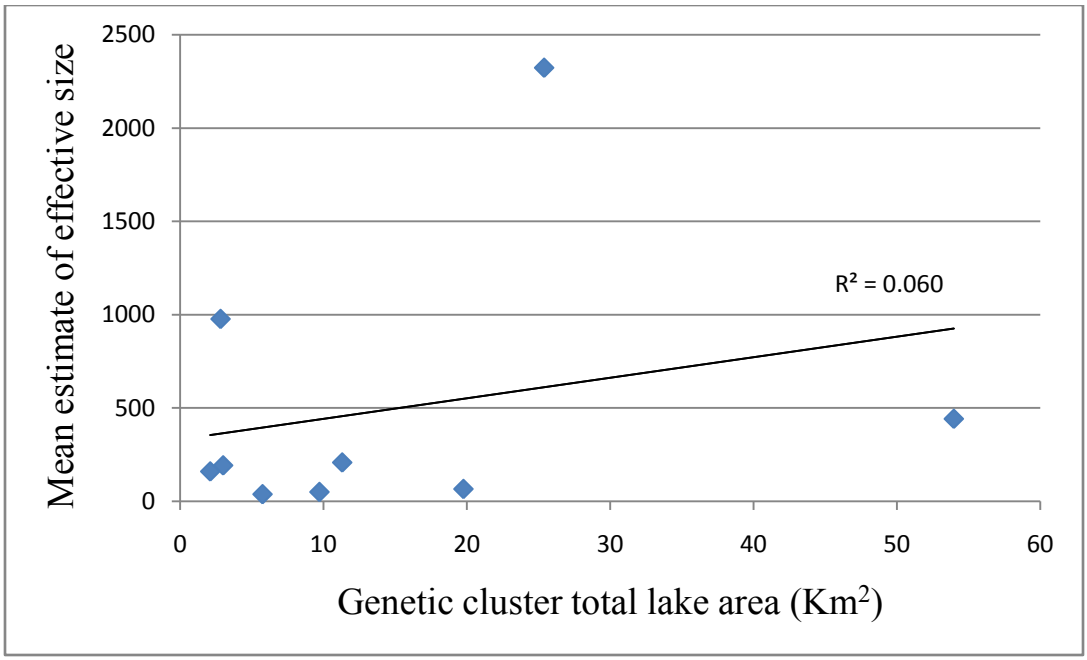


Figure 3.5.1. The relationship between genetic cluster total lake area and mean estimates of effective population size for each of the 9 unique clusters identified with STRUCTURE. The relationship is non-significant ($p > 0.50$) indicating that the estimates of effective population size are not dependent on area.

CHAPTER 4: DISCUSSION

4.1 Summary of Results

My results indicate that lake trout in the Kogaluk River System in Northern Labrador exhibit a hierarchical population structure in which lakes north and south of the Kogaluk River form two distinct clusters. Examination of F_{ST} values via PCA revealed significant differences within these initial clusters with the highest levels of differentiation occurring among the southern lakes, a result further corroborated by additional STRUCTURE analyses. These southern lakes are linked to each other and to the Kogaluk River by rivers containing waterfalls, unlike the northern lakes which lack connections with waterfalls. A system-wide analysis suggested a pattern of IBD was evident, but a more detailed examination using a decomposed pairwise regression approach identified three lakes as containing "outlier" lake trout populations (Hawk, T-Bone, and Cabot Lakes) suggesting that these lakes were more genetically divergent than can be explained by distance alone. These outlier lakes have their connectivity limited to each other and to the rest of the system by waterfalls, suggesting a landscape effect on the distribution of genetic diversity among lake trout in this system. Lake Trout in these three lakes exhibited a relatively high level of divergence, likely the result of genetic drift acting on relatively isolated populations of small effective sizes (except Cabot Lake). No additional landscape variables appeared correlated with genetic divergence. Preliminary $\widehat{N}_e$ ranged widely and there is no significant relationship with lake area. However, these $\widehat{N}_e$s are relatively higher in non-headwater populations. Below, I discuss these issues in detail in the context of lake trout life history variation and landscape effects.

4.2 Population Structure and Gene Flow

Previous studies on lake trout have suggested that levels of genetic diversity evident in a region could be representative of secondary contact between glacial lineages (Grewe and Hebert 1988; Ihssen *et al.* 1988; Wilson and Hebert 1996). I argue that the differences in population structure evident in my study system are not a function of secondary contact between multiple glacial lineages. This is due in part to the difference in scale between my own study and that of these previous studies, and because phylogeographic studies based on mtDNA polymorphism have suggested that lake trout from this particular region of Labrador come from a single glacial lineage originating from an Atlantic refugium (Wilson and Hebert 1996; Wilson and Hebert 1998). I suggest that the high level of genetic diversity seen in my system is a result of lake trout life history, the landscape, and their interaction.

At the highest hierarchical level, all lakes draining into the Kogaluk River from the north differ significantly from those draining into it from the west and south (Figures 2.1.1, 3.2.1). Waterfalls were shown to be highly correlated with genetic structure in this system (discussed below) and likely play a part in structuring the genetic makeup at this high hierarchical level. In particular, one large waterfall (Figure 2.1.1, WF1) appears to be the dividing point between the initial two clusters. All sampled lakes north of this waterfall exist as one cluster, as do all lakes to the south, including Cabot Lake which is part of the Kogaluk River. At the finest hierarchical scale, there appear to be nine unique clusters, and there is only one occurrence of two lakes identified as forming a single cluster (Lakes Esker and WP152). Two waterfalls are present within this southern region

(see Figure 2.1.1 for waterfall positions; Table 2.1.1 for waterfall heights and angles) preventing upstream migration and thus creating asymmetries in gene flow.

The influence of waterfalls on asymmetries in gene flow as well as divergence between populations has been reported for other freshwater species (see Gomez-Uchida 2009; Crispo *et al.* 2006; Kanno *et al.* 2011; Whiteley *et al.* 2010; Naurem *et al.* 2006; and Tatarenkov *et al.* 2010; Junker *et al.* 2012). These waterfalls help explain some of the genetic structuring at a finer spatial scale within the southern cluster; however, the lakes in the northern cluster are not separated by waterfalls, thus the high degree of genetic structuring is likely a result of other factors. In particular, this species exhibits homing behaviour whereby adults return to the same spawning areas every year (Esteve *et al.* 2008; Marsden *et al.* 1995) and doing so contributes to genetic isolation (Ihssen *et al.* 1988). My F_{ST} estimates support this idea of genetic isolation as my pairwise estimates ranged from 0.008 to 0.217 with an average $F_{ST} = 0.123$. These estimates are comparable to estimates obtained for landlocked arctic charr (*Salvelinus alpinus*) in a similar study by Bernatchez *et al.* (2002) in which F_{ST} estimates ranged from approximately 0.05 to 0.20.

Supporting patterns recognizable in the pairwise F_{ST} estimates, migration rates suggest little movement between lakes, and in most cases these estimates are essentially zero (Table 3.2.4). The relationship between Lakes Esker and WP152 is noteworthy. These are the only two lakes with genetically indistinguishable lake trout. This clustering is likely due to the high (however asymmetric) gene flow from Lakes Esker to WP152 ($\hat{m}$= 0.288), probably resulting from the relative proximity between, and large waterway connecting these two lakes. The high rate of migration from Esker Lake to Lake Genetics

B however is not consistent with present geographical limitations. I suggest that this relatively high estimate of gene flow may be reflective of random genetic drift.

I should note that BayesAss+ estimates recent migration - migration within the previous few generations (Wilson and Rannala 2003), thus inferences on connectivity in the distant past based on BayesAss+ estimates are likely unwarranted. An additional hypothesis regarding such high gene flow is the possibility of anthropogenic transplantation. This is unlikely due to the isolated nature of the system.

Field observations from the air revealed that streams connecting lakes were relatively shallow; precise depth measurements were not taken but the streams exhibited depths generally ≤ 1 m in early July. Water temperature in the lakes, during this same time period, ranged from 4.8 to 10.1°C with an average temperature of 6.7°C (data not shown). Lake trout are known to prefer waters ranging in temperature from 6 - 13°C (Martin and Olver 1980) with an optimal range between 8 and 12°C (Magnuson *et al.* 1990) though there is much debate over these exact temperatures (see MacKenzie-Grieve and Post 2006). Temperature has been suggested to be a very important factor influencing the movement of lake trout (Martin and Olver 1980). Martin (1954) noted the influence of a thermal barrier on *S. namaycush* in Lake Louisa, Ontario in which lake trout were limited to feeding on plankton during the summer months because the prey fish species, minnows, were found only in shallow, warmer waters. Therefore, I suggest that although the sampled lakes were shallow themselves, some having observed maximum depths <10m (Table 3.1.1), they were significantly deeper than the connecting streams and likely maintained a lower average temperature. This differential in stream and lake temperatures in addition to the general shallow nature of the streams may negatively influence the

tendency to migrate and may be aiding in the genetic structuring of this system. Temperature has been shown to be associated with genetic structure in numerous aquatic species including Atlantic salmon (Dionne *et al.* 2008), rainbow/steelhead trout (Narum *et al.* 2008), sea urchins (Banks *et al.* 2007), and short-beaked common dolphin (Amaral *et al.* 2012) and may account at least in part, for the high degree of genetic structure seen between pairs of lakes not influenced by waterfalls in this system.

An additional factor likely acting against lake trout leaving the lake and entering the streams is the acquisition of food. It has been speculated that a lake trout movement may be dependent on the pursuit of prey (Martin and Olver 1980; Schmalz *et al.* 2002). Due to the shallow nature of these streams, they may simply not accommodate sufficient quantities of prey, or like lake trout, the prey species avoid the streams due to undesirable temperatures or insufficient space. Thus, the lake trout in this system may tend to reside in the lakes where the prey species including round whitefish (*Prosopium cylindraceum*), lake chub (*Couesius plumbeus*), Arctic charr (*Salvelinus alpinus*), brook trout (*Salvelinus fontinalis*), and occasionally burbot (*Lota lota*) are also present.

4.3 Influence of Landscape Factors

Structure, F_{ST}, Mantel tests, and migration rate estimates suggest an influence of waterfalls on genetic structure. Initially, geographic distance was found to have an influence on genetic divergence on the system as a whole, however, finer scale decomposed pairwise regression analysis suggested that all southern populations, except Lake Genetics B, are outliers exhibiting higher levels of genetic divergence than can be explained by distance alone (Koizumi *et al.* 2006). This is correlated with the geographical positions of these lakes as they all drain into the Kogaluk River (except for Cabot Lake which is a part of the Kogaluk River itself) and are isolated from the northern lakes due to the dendritic nature of the Kogaluk System and the presence of numerous waterfalls between the northern and southern lakes. Outlier populations in natural systems are not uncommon. Populations exhibiting higher or lower genetic diversity than can be explained by distance alone are not unique to this system, as they have been shown in several previous studies (e.g., Koizumi *et al.* 2006; Junge *et al.* 2011; Cunningham *et al.* 2009). Surprisingly, the *S. namaycush* population from one of the southern lakes, Genetics B, was not identified as an outlier consistent with my results from previous analyses, suggesting *S. namaycush* in this lake exhibit some degree of genetic similarity with *S. namaycush* from lakes north of the Kogaluk System.

Interestingly, no other landscape variables tested including average slope, elevation, and number of intermediate lakes appeared to have significant influence on genetic structure (see Table 3.4.1). GESTE results suggest that no landscape factors are particularly useful for predicting genetic diversity (Table 3.4.2). This was interesting especially with regards to the differences in slope. Although a number of previous studies

have suggested a correlation between slope and genetic diversity in other systems and species (Caldera and Bolnick 2008; Cook *et al.* 2011; and Kanno *et al.* 2011 for examples) my study is not the first to show no effect of slope (see Stelkins *et al.* 2012 for a recent example). I argue that the absence of a correlation between genetic distance and slope is likely due to one of two possibilities. First, the slope between any two pairs of lakes is not very steep; the highest average slope is only 5.27 m per kilometer (between Hawk and Cabot Lakes - see Appendix C for average slope measurements). The slope measurements are so gradual in this system; they likely have very little impact on the migration of local species except when there is the presence of a waterfall. The second possibility is that fish are simply not moving between lakes, thus environmental factors like slope of the connecting waterways would be expected to play little to no role in shaping the genetic structure of the system.

4.4 Effective Population Size

Estimates of effective population size varied by two orders of magnitude from 38 in Hawk Lake to over 2300 in Cabot Lake. First generation migrants were evident in Genetics H, Hawk, T-Bone, Strange, and pooled Esker/WP152 Lakes. After their exclusion, mean $\widehat{N}_e$ increased greatly in Lake Genetics H (300% increase) and decreased moderately for Strange Lake (16% decrease) and pooled Esker/WP152 Lakes (22.7% decrease). There was no change in the mean estimate for Hawk Lake and only a slight change in the mean estimate for T-Bone Lake (1.5% increase), suggesting that although migrations can introduce linkage disequilibrium potentially lowering the estimates of N_e, the degree to which these migrants affect linkage disequilibrium depends on how genetically distinct the migrants are from the local population, coupled with the number of migrants. One caveat regarding this analysis is that examination of the confidence intervals of these estimates indicates a non-significant change post migrant removal in all cases, due to the presence of overlapping confidence intervals. This suggests that inclusion of first generation migrants had no statistically significant impact on $\widehat{N}_e$.

To test whether these results were a function of immigration (increased linkage disequilibrium) or simply a result of reduced sample size, I followed the techniques used in a recent study brown trout (*Salmo trutta*) by Serbezov *et al.* (2012) who, using various estimators of N_e including LDNe, reported similar changes in mean $\widehat{N}_e$ and associated 95% confidence intervals once migrants were removed. To address the issue of determining what was affecting the mean estimates and 95% confidence intervals, they randomly removed the same number of individuals as the indicated number of putative migrants and re-ran their analysis. If they achieved the same degree of change by

removing random individuals as opposed to immigrants, the initially observed changes would have had to have been a result of reduced sample size and not immigration.

To examine this issue in my study, I focused on Lake Genetics H which had the greatest change in mean $\widehat{N}_e$ after the removal of migrants (Table 3.5.1) and found that removing the same number of random samples as those identified as migrants, resulted in a much smaller mean $\widehat{N}_e$ ($\widehat{N}_e$ = 310 – see Appendix E) than in Table 3.5.1. This indicates that the initial change observed as a result of removing putative migrants, is a result of removing some linkage disequilibrium brought on by immigration, and was not simply a result of reduced sample size. Thus, although these differences post immigrant removal were not statistically significant, I have shown that they are related to immigration and cannot be ignored. Therefore, I suggest that I have evidence for a pattern which is consistent with Waples and England (2011), who suggest that immigrants can either increase $\widehat{N}_e$ provided that the immigrants are from a genetically similar pool or population, or could decrease $\widehat{N}_e$ estimates if immigrants are genetically different. I argue that the differences we observed were not statistically significant due to the relatively small number of immigrants in each population (2 or fewer – see Table 3.5.1), and had the migration rate been higher, the results would likely have been significant.

My estimates were limited to single sample methods as they were estimated based on linkage disequilibrium using LDNe. When estimating N_e in a system such as this, using single sample methods, additional source of linkage disequilibrium need to be eliminated prior to estimation. Migration from other populations can lead to the subsequent interbreeding with the local population (creating first generation migrants) creating admixture which results in increased linkage disequilibrium (Waples 2006),

thereby reducing N_e estimates. It is important to note that this method estimates the effective number of breeders in the previous generation if the estimate is based on individuals of the same cohort (Waples 2006). If age is unknown and individuals differ in age, as is likely the case in my study, LDNe estimates a quantity that lies between N_b and N_e. Therefore, the $\widehat{N}_e$ I obtained with LDNe were estimates of the number of breeders which produced my sample (Waples 2006) which is presumably made up of individuals from various different generations. For these reasons, I cannot obtain an accurate N_e estimation for my populations. I present these estimates of N_b per population sample following Phillipsen *et al.* (2010) and suggest that these values should be deemed useful for the sake of comparison with other studies using the LDNe software and the linkage disequilibrium method. Although these are not estimates of N_e, I can infer a more general pattern which suggests that lower estimates of N_b will correlate with lower estimates of N_e (Waples 2005; Phillipsen *et al.* 2010).

As I expected, all headwater lakes (except WP152) have significantly smaller effective sizes than the non-headwater lakes (T-test, $p<0.05$). Surprisingly however, although the $\widehat{N}_e$ are significantly different, the mean observed heterozygosity values between types of lakes did not differ (*t*-test, $p>0.10$). Morrissey and de Kerckhove (2009) suggest that in a dendritic system such as this, one would expect higher levels of genetic diversity in downstream populations compared to the headwater populations. This would lead to higher $\widehat{N}_e$ in these downstream populations. However, the lack of significance between mean values of observed heterozygosity between the lake types, coupled with my estimation of very low rates of migration, suggests the differences in effective

population sizes may not be a function of gene flow, but likely of successive founding events following a pattern of colonization from downstream sources.

4.5 Conclusion

Using a landscape genetics approach, I am able to show the influence of landscape variables on genetic differentiation within this fragmented system. Waterfalls and geographic distance appear to contribute significantly, more than the other factors I tested, and likely contribute to the hierarchical population structure and lack of gene flow evident within this system, though further examination of untested variables would be beneficial to support these conclusions. Surprisingly however, my data suggested that both geographic distance and waterfall position could not be used as reliable predictors of genetic differentiation, even though there was evidence of a significant correlation between them. I propose that in addition to the landscape variables, lake trout life history, in particular its preference for colder and deeper waters may be at least partially responsible for the genetic structuring evident within this system.

References

Amaral, AR; Beheregaray LB; Bilgmann, K; Boutov, D; Freitas, L; Robertson, KM; Sequeira, M; Stockin, KA; Coelho, MM; Moller, LM (2012) Seascape genetics of a globally distributed, highly mobile marine mammal: the short-beaked common dolphin (genus *Delphinus*). *PLoS ONE* 7(2): e31482.

Anderson,TC (1985) The rivers of Labrador. *Canadian Special Publication of Fisheries and Aquatic Sciences* 81: 1-389

Antao, T; Lopes, A; Lopes, RJ; Beja-Pereira, A; Luikart, G (2008) LOSITAN: A workbench to detect molecular adaptation based on a F_{ST}-outlier method. *BMC Bioinformatics* 9: 323.

Banks, SC; Piggott, MP; Williamson, JE; Beheregaray, LB (2007) Microsatellite DNA markers for analysis of population structure in the sea urchin *Centrostephanus rodgersii*. *Molecular Ecology* 7: 321-323.

Bernatchez, L; Rhydderch, JG; Kircheis, FW (2002) Microsatellite gene diversity analysis in landlocked Arctic Charr from Maine. *Transactions of the American Fisheries Society* 131: 1106-1118.

Caldera, EJ; Bolnick, D (2008) Effects of colonization history and landscape structure on genetic variation within and among threespine stickleback (*Gasterosteu aculeatus*) populations in a single watershed. *Evolutionary Ecology Research* 10, 575-598.

Cook, BD; Kennard, J; Real, K; Pusey, BJ; Hughes, JM (2011) Landscape genetic analysis of the tropical freshwater fish *Mogurnda mogurnda* (Eleotridae) in a monsoonal river basin: importance of hydrographic factors and population history. *Freshwater Biology* 56, 812-827.

Crispo, E; Bentzen, P; Reznick, DN; Kinnison, MT; Hendry, A (2006) The relative influence of natural selection and geography on gene flow in guppies. *Molecular Ecology* 15: 49-62.

Cunningham, KM; Canino, MF; Spies, IB; Hauser, L (2009) Genetic isolation by distance and localized fjord population structure in Pacific cod (*Gadus macrocephalus*): limited effective dispersal in the northeastern Pacific Ocean. *Canadian Journal of Fisheries and Aquatic Sciences* 66: 153-166.

Dehaan, PW; Ardren, W (2005) Characterization of 20 highly variable tetranucleotide microsatellite loci for bull trout (*Salvelinus confluentus*) and cross-amplification in other *Salvelinus* species. *Molecular Ecology Notes* 5, 582-585.

Dent, EA, von Holdt, BM (2012) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources* 4: 359-361.

Dieringer, D; Schlotterer, C (2003) MICROSATELLITE ANALYSER (MSA): a platform independent analysis tool for large microsatellite data sets. *Molecular Ecology Notes* 3: 167-169.

Dionne, M; Caron, F; Dodson, JJ; Bernatchez, L (2008) Landscape genetics and hierarchical genetic structure in Atlantic Salmon: the interaction of gene flow and local adaptation. *Molecular Ecology* 17: 2382-2396.

Elphinstone, MS; Hinten, GN; Anderson, MJ; Nock, CJ (2003). An inexpensive and high-throughput procedure to extract and purify total genomic DNA for population studies. *Molecular Ecology Notes* 3: 317-320.

Esteve, M; McLennan, DA; Gunn, JM (2008) Lake Trout (*Salvelinus namaycush*) spawning behaviour: the evolution of a new female strategy. *Environmental Biology of Fishes* 83: 69-76.

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

Excoffier, L; Lischer, HEL (2010) Arlequin suite ver 3.5: A new series of programs to perform population genetic analysis under Linux and Windows. *Molecular Ecology Resources* 10: 564-567.

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

Foll, M; Gaggiotti, OE (2006) Identifying the environmental factors that determine the genetic structure of Populations. *Genetics* 174: 875-891.

Giroux, TMJ; Chivers, DP; Fitzsimmons, MJ; Chilton, NB (2009) Genetic diversity in a remnant population of lake trout (*Salvelinus namaycush*). *Canadian Journal of Zoology* 87: 642-647.

Gomez-Uchida, D; Knight, TW; Ruzzante, DE (2009) Interaction of landscape and life history attributes on genetic diversity, neutral divergence and gene flow in a pristine community of salmonids. *Molecular Ecology*18: 4857-4869.

Goudet, J (2001) FSTAT, a program to estimate and test gene diversities and fixation indicies (version 2.9.3) Available from http://www.unil.ch/izea/softwares/fstat.html.Updated from Goudet, J. (1995).FSTAT v-1.2.A computer program to calculate F-statistics. *Heredity* 86: 485-486.

Grewe, PM; Hebert, PDN (1988) Mitochondrial DNA diversity among brood stocks of the Lake Trout, *Salvelinus namaycush*. *Canadian Journal of Fisheries and Aquatic Sciences* 45: 2114-2112.

Hubisz, MJ; Falush, D; Stephens, M; Pritchard, JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* 9: 1322-1332.

Ihssen, PE; Casselman, JM; Martin, GW, Phillips, RB (1988) Biochemical genetic differentiation of lake trout (*Salvelinus namaycush*) stocks of the Great Lakes region. *Canadian Journal of Fisheries and Aquatic Sciences* 45: 1018-1029.

Jakobsson, M; Rosenberg, NA (2007) CLUMPP: a cluster matching and permutation program for dealing with label switching and multimodality analysis of population structure. *Bioinformatics* 23: 1801-1806.

Junge, C; Vollestad, LA; Barson, NJ; Haugen, TO; Otero, J; Saetre, GP; Leder, EH; Primmer, CR (2011) Strong gene flow and lack of stable population structure in the face of rapid adaptation to local temperature in a spring-spawning salmonid, the European grayling (*thymallus thymallus*). *Heredity* 106: 460-471.

Junker, J; Peter, A; Wagner, CE; Mwaiko, S; Germann, B; Seehausen, O; Keller, I (2012) River fragmentation increases localized population genetic structure and enhances asymmetry of dispersal in bullhead (*Cottus gobio*). *Conservation Genetics* 13: 545-556.

Kanno, Y; Vokoun, J; Letcher, BH (2011) Fine-Scale population structure and riverscape genetics of brook trout (*Salvelinus fontinalis*) distributed continuously along headwater channel networks. *Molecular Ecology* 20: 3711-3729.

Kaeuffer, R; Reale, D; Coltman, DW; Pontier, D (2007) Detecting population structure using STRUCTURE software: effect of background linkage disequilibrium. *Heredity* 99: 374-380.

Kimura, M; Weiss, GH (1964) The stepping stone model of population structure and the decrease of genetic correlation with distance. *Genetics* 49: 561-576.

Koizumi, I; Yamamoto, S; Maekawa, K (2006) Decomposed pairwise regression analysis of genetic and geographic distances reveals a metapopulation structure of stream-dwelling Dolly Varden charr. *Molecular Ecology* 15: 3175-3189.

Limborg, MT; Helyar, SJ; De Bruyn, M; Taylor, MI; Nielsen, EE; Ogden, R; Carvalho, GR; Consortium, FPT; Bekkevold, D (2012) Environmental selection on transciptome-derived SNPs in a high gene flow marine fish, the Atlantic herring (*Clupea harengus*). *Molecular Ecology* 21: 3686-3703.

Luikart, G; England, PR; Tallmon, D; Jordan, S; Taberlet, P (2003) The power and promise of population genomics: from genotyping to genome typing. *Nature Reviews Genetics* 4: 981-994.

MacKenzie-Grieve, JL; Post, JR (2006) Thermal habitat use by Lake Trout in two contrasting Yukon Territory lakes. *Transactions of the American Fisheries Society* 135: 727-738.

Magnuson, JJ; Meisner, JD; Hill, DK (1990) Potential changes in the thermal habitat of Great Lakes fish after global climate warming. *Transactions of the American Fisheries Society* 119: 254-264.

Manel, Stephanie; Schwartz, MK; Luikart, G; Taberlet, P (2003) Landscape genetics: combining landscape ecology and population genetics. *TRENDS in Ecology and Evolution* 18: 189-197.

Marsden, JE; Casselman, JM; Edsall, TA; Fitzsimons, JD; Horns, WH; Manny, BA; McAughey, SC; Sly, PG; Swanson, BL (1995) Lake Trout spawning habitat – a synthesis of current knowledge. *Journal of Great Lakes Research* 21: 487-497.

Martin, NV (1954) Catch and winter food of lake trout in certain Algonquin Park lakes. *Journal of the Fisheries Research Board of Canada* 11: 5-10.

Martin, NV; Olver, CH (1980) The lake charr, *Salvelinus namaycush*. Pages 205-277 in Balon EK, editor. Charrs: salmonid fishes of the genus Salvelinus. Higham, Massachusetts: Kluwer Boston.

Morrissey, M; de Kerckhove, DT (2009) The Maintenance of genetic variation due to asymmetric gene flow in dendritic metapopulations. *The American Naturalist* 174: 875-889.

Naurem, SR; Boe, S; Moran, P; Powell, M (2006) Small-scale genetic structure and variation in steelhead in the Grande Ronde River, Oregan, USA. *Transaction of the American Fisheries Society* 135: 979-986.

Naurem, SR; Hess, JE (2011) Comparison of F_{ST} outlier test for SNP loci under selection. *Molecular Ecology Resources* 11: 184-194.

Narum, SR; Zendt, JS; Graves, D; Sharp, WR (2008) Influence of landscape on resident and anadromous life history types of *Oncorhynchus mykiss*. *Canadian Journal of Fisheries and Aquatic Sciences* 65: 1013-1023.

Northrup, S; Connor, M; Taylor, EB (2010) Population structure of lake trout (Salvelinusnamaycush) in a large glacial-fed lake inferred from microsatellite DNA and morphological analysis. *Canadian Journal of Fisheries and Aquatic Sciences* 67: 1171-1186.

Oksanen, J; Blanchet, G; Kindt, R; Legendre, P; Minchin, PR; O'Hara, RB; Simpson, GL; Solymos, P; Henry, MH; Wagner, S; Wagner, H (2012) Vegan: Community Ecology Package. R package version 2.0-3. http://CRAN.R-project.org/package=vegan

Olsen, JB; Bentzen, P; Seeb, J (1998) Characterization of seven microsatellite loci derived from pink salmon. *Molecular Ecology* 7: 1083-1090.

O'Reilly, PT; Hamilton, LC; McConnell, SK; Wright, JM (1996) Rapid analysis of genetic variation in Atlantic salmon (*Salmo Salar*) by PCR multiplexing of dinucleotide and tetranucleotide microsatellites. *Canadian Journal of Fisheries and Aquatic Sciences* 53: 2292-2298.

Paetkau, D; Slade, R; Burden, M; Estoup, A (2004) Genetic assignment methods for the direct, real-time estimation of migration rate: a simulation-based exploration of accuracy and power. *Molecular Ecology* 13: 55-65.

Peakall, R; Smouse, PE (2006) GENALEX 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes* 6: 288-295.

Perry, GML; King, TL; St. Cyr, J; Valcourt, M; Bernatchez, L (2005) Isolation and cross-familial amplification of 41 microsatellites for the brook charr (*Salvelinus fontanalis*). *Molecular Ecology Notes* 5: 346-351.

Phillipsen, IC; Bowerman, J; Blouin, M (2010) Effective number of breeding adults in Oregon spotted frogs (*Rana pretiosa*): genetic estimates at two life stages. *Conservation Genetics* 11: 737-745.

Piry, S; Alapetite, A; Cornuet, JM; Paetkau, D; Baudouin, L; Estoup, A (2004) GeneClass2: A software for genetic assignment and first-generation migrant detection. *Heredity* 95: 536-539.

Pizarro, JC; Gilligan, LM; Stevens, L (2008) Microsatellites reveal a high population structure in *Triatoma infestans* from Chuquisaca, Bolivia.*PLoS Neglected Tropical Diseases* 2: e202.

Primmer, CR; Aho, T; Piironen, J; Estoup, A; Cornuet, J-M; Ranata, E (1999) Microsatellite analysis of hatchery stocks and natural populations of Arctic charr, *Salvelinus alpinus*, from the Nordic region: implications for conservation. *Hereditas* 130: 277-289.

Rannala, B; Mountain, JL (1997) Detecting immigration by using multilocus genotypes. *Proceedings of the National Academy of Sciences USA* 94: 9197-9201.

Reed, DH; Frankham, R (2001) How closely correlated are molecular and quantitative measures of genetic variation? A meta-analysis. *Evolution* 55: 1095-1103.

Rexroad III, CE; Coleman, RL; Hershberger, WK; Killefer, J (2002) Rapid communication: Thirty-eight polymorphic microsatellite markers for mapping in rainbow trout. *Journal of Animal Science* 80: 541-542.

Rice, WR (1989) Analyzing tables of statistical tests. *Evolution* 43: 223-225.

Rollins, MF; Vu, NH; Spies, IB; Kalinowski, ST (2009) Twelve microsatellite loci for lake trout (*Salvelinus namaycush*). *Molecular Ecology Resources* 9: 871-873.

Rosenberg, NA; Pritchard, JK; Weber, JL; Cann, HM; Kidd, KK; Zhivotovsky, LA; Feldman, MW (2002) Genetic Structure of Human Populations. *Science* 20: 2381-2385.

Rosenberg, NA (2004) DISTRUCT: a program for the graphical display of population structure. *Molecular Ecology Notes* 4: 137-138.

Rousset, F (1997) Genetic Differentiation and Estimation of Gene Flow from *F*-Statistics Under Isolation by Distance. *Genetics* 145: 1219-1228.

Roy, D; Hurlbut, T; Ruzzante, DE (2012) Biocomplexity in a demersal exploited fish, white hake (*Urophycis tenuis*): Depth related structure and inadequacy of current management approaches. *Canadian Journal of Fisheries and Aquatic Sciences* 69: 414-429.

Schmalz, PJ; Hansen, MJ; Holey, ME, McKee, PC; Toneys, ML (2002) Lake trout movements in northwestern Lake Michigan. *North American Journal of Fisheries Management* 22: 737-749.

Schwartz, M; Luikart, G; Cushman, S (2010) Landscape genomics: a brief perspective. Pages 165-174 in Huettman, F; Cushman, S: *Spatial Complexity, Informatics, and Wildlife Conservation*. Springer-Verlag, Berlin.

Scott, WB; Crossman, EJ, (1973) Freshwater fishes of Canada. *Bulletin of the Fisheries Research Board of Canada* 184: 1-966.

Serbezov, D; Jorde, PE; Bernatchez, L; Olsen, EM; Vollestad, LA (2012) Short-Term Genetic Changes: Evaluating Effective Population Size Estimates in a Comprehensively Described Brown Trout (*Salmo trutta*) Population. *Genetics* 191: 579-592.

Seyahooei, MA; van Alphen, JJM; Kraaijeveld (2011) Genetic structure of *Leptopilina boulardi* populations from different climatic zones of Iran. *BMC Ecology* 11: 4.

Stelkens, RB; Jaffuel, G; Escher, M; Wedekind, C (2012) Genetic and phenotypic population divergence on a microgeographic scale in brown trout. *Molecular Ecology* 12: 2896-2915.

Storfer, A; Murphy, M; Evans, Goldberg, CS; Robinson, S; Spear, SE; Dezzani, R; Delmelle, E; Vierlin, L; Waits, LP (2007) Putting the "landscape" in landscape genetics. *Heredity* 98: 128-142.

Tatarenkov, A; Healey, CIM; Avise, JC (2010) Microgeographic population structure of green swordtail fish: genetic differentiation despite abundant migration. *Molecular Ecology* 19:257-268.

Van Oosterhout, C; Hutchinson, WF; Wills, DPM; Shipley, P (2004) MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4: 535-538.

Waples, RS (2005) Genetic estimates of contemporary effective population size: to what time periods do the estimates apply? *Molecular Ecology* 14: 3335-3352.

Waples, RS (2006) A bias correction for estimates of effective population size based on linkage disequilibrium at unlinked gene loci. *Conservation Genetics* 7: 167-184.

Waples, RS; Do, C (2008) LDNE: a program for estimating effective population size from data on linkage disequilibrium. *Molecular Ecology Resources* 8: 753-756.

Waples, RS; Do, C (2010) Linkage disequilibrium estimates of contemporary Ne using highly variable genetic markers: A largely untapped resources for applied conservation and evolution. *Evolutionary Applications* 3: 244-262.

Waples, RS; England, PR (2011) Estimating Contemporary Effective Population Size on the Basis of Linkage Disequilibrium in the Face of Migration. *Genetics* 189: 633-644.

Whiteley, AR; Hastings, K; Wenburg, JK; Frissell, CA; Martin, JC; Allendorf, FW (2010) Genetic variation and effective population size in isolated populations of coastal cutthroat trout. *Conservation Genetics* 11: 1929-1943.

Williamson, KS; Cordes, JF; May, B (2002) Characterization of microsatellite loci in Chinook salmon (*Onorhynchus tshawytscha*) and cross-species amplification in other salmonids. *Molecular Ecology Notes* 2: 17-19.

Wilson, AJ; Gislason, D; Skulason, S; Snorranson, DD; Adams, CE; Alexander, G; Danzmann, RG; Ferguson, MM (2004) Population genetic structure of Arctic Charr, *Salvelinus alpines* from northwest Europe on large and small spatial scales. *Molecular Ecology* 13: 1129-1142.

Wilson, CC, Hebert, PDN (1996) Phylogeographical origins of lake trout (*Salvelinus namaycush*) in eastern North America. *Canadian Journal of Fisheries and Aquatic Sciences* 53: 2764-2775.

Wilson, CC, Hebert, PDN (1998) Phylogeography and postglacial dispersal of lake trout (*Salvelinus namaycush*) in North America. *Canadian Journal of Fisheries and Aquatic Sciences* 55: 1010-1024.

Wilson, GA; Rannala, B (2003) Bayesian inference of recent migration rates using multilocus genotypes. *Genetics* 163: 1177-1191.

Wright, S (1931) Evolution in Mendelian populations. *Genetics* 16: 97-159.

APPENDIX A

Appendix A. Multiplex primer panel composition indicating loci dye label, assigned panel number, and source authors.

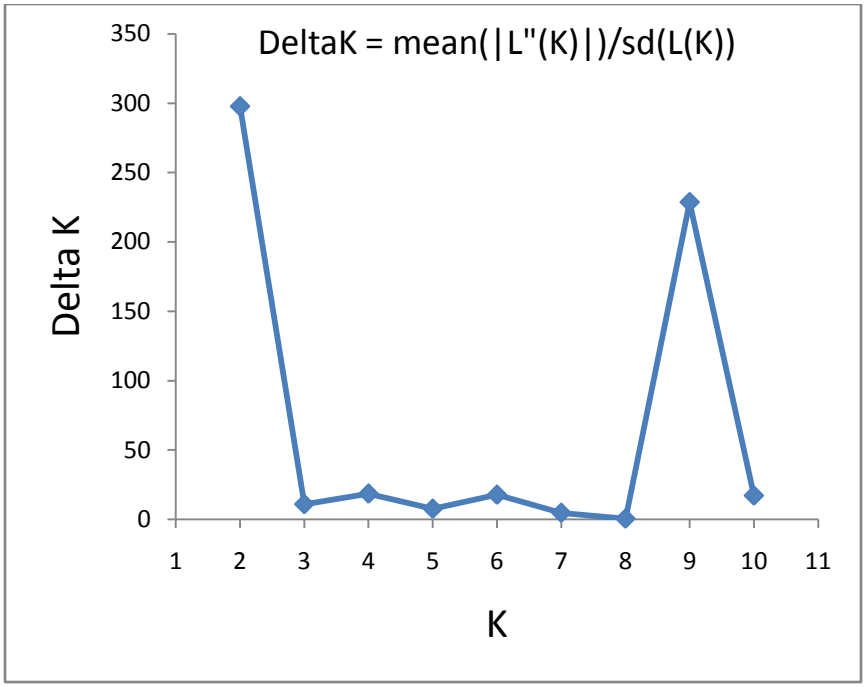
Locus	Primer Panel	T_a	Label	Source
SSA85	1	60	800	O'Reilly *et al.* 1996
SCO215	1	60	700	Dehaan*et al.* 2005
SCO202	1	60	800	Dehaan*et al.* 2005
OGO1A	2	60	700	Olsen *et al.* 1998
SNAMSU06	2	60	800	Rollins *et al.* 2009
SNAMSU12	2	60	800	Rollins *et al.* 2009
SFO334	3	57	700	Perry *et al.* 2005
OTSG253b	3	57	700	Williamson *et al.* 2002
SNAMSU02	4	57	800	Rollins *et al.* 2009
SCO102	4	57	800	Sewall Young, personal communication
SCO107	4	57	700	Sewall Young, personal communication
OTSG83b	*	52	700	Williamson *et al.* 2002
OMM1105	*	62	700	Rexroad *et al.* 2002

*These primers were run individually, and not assigned to a multiplex panel as they worked better by themselves.

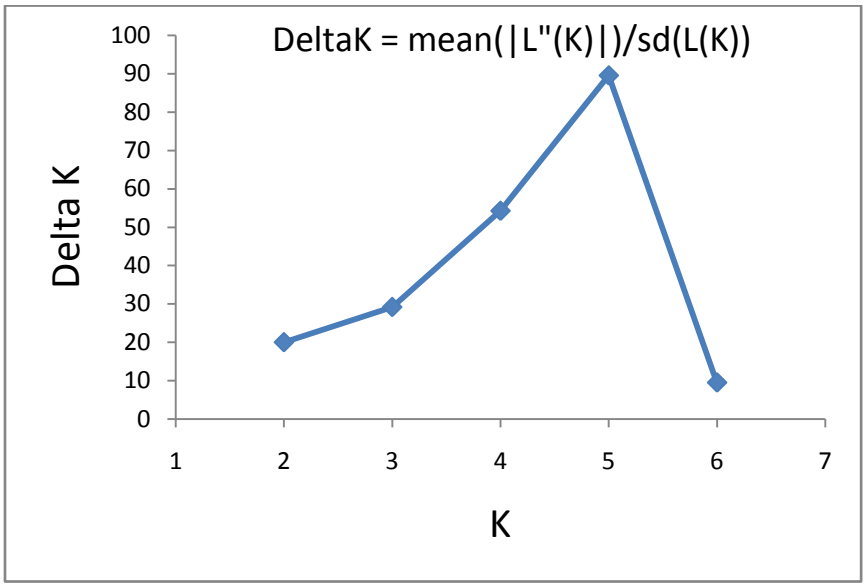
Appendix B. Shown here are the five pairs of loci which have the highest probability of being linked per lake, along with the associated p values. These significance estimates are prior to sequential Bonferroni correction. With an alpha level of $\alpha=0.05$ we expect 3 of the 66 possible pairs of loci to be linked due to chance alone. No lake had more than 3 linked pairs, which can be attributed to chance alone. However, for every lake, loci 1 (SCO107) and 8 (OTSG83b) appeared linked. This is a statistical improbability which caused me to assume that these loci were potentially linked. As a result, I omitted one locus from further analysis (OTSG83b) because it had a slightly higher failure rate.

Lake 1	**P =**	**Genetics H**	**P =**	**Slushy**	**P =**	**Esker**	**P =**	**Strange**	**P =**
Pair(1,8)	0.000	Pair(1, 8)	0.000	Pair(1, 8)	0.000	Pair(1, 8)	0.000	Pair(1, 8)	0.000
Pair(0,2)	0.002	Pair(3, 4)	0.000	Pair(0, 11	0.000	Pair(4, 5)	0.017	Pair(1, 11	0.007
Pair(0,9)	0.004	Pair(0, 2)	0.048	Pair(2, 7)	0.000	Pair(3, 4)	0.040	Pair(0, 10	0.022
Pair(1,3)	0.006	Pair(1, 2)	0.050	Pair(4, 7)	0.004	Pair(0, 10	0.160	Pair(1, 7)	0.039
Pair(1,5)	0.018	Pair(0, 8)	0.060	Pair(1, 2)	0.015	Pair(1, 11	0.160	Pair(3, 7)	0.043
WP152	**P =**	**T-Bone**	**P =**	**Cabot**	**P =**	**Genetics B**	**P =**	**Hawk**	**P =**
Pair(1,8)	0.000	Pair(1, 8)	0.000	Pair(1, 8)	0.000	Pair(1, 8)	0.000	Pair(3, 4)	0.000
Pair(9,10	0.008	Pair(4, 10	0.017	Pair(3, 4)	0.001	Pair(4, 11	0.002	Pair(1, 8)	0.000
Pair(3,10	0.015	Pair(1, 2)	0.019	Pair(5, 7)	0.001	Pair(4, 6)	0.002	Pair(9, 10	0.002
Pair(5,7)	0.034	Pair(4, 6)	0.053	Pair(0, 2)	0.018	Pair(0, 4)	0.025	Pair(7, 11	0.011
Pair(2,8)	0.035	Pair(3, 10	0.089	Pair(10, 1	0.035	Pair(2, 4)	0.049	Pair(0, 1)	0.018

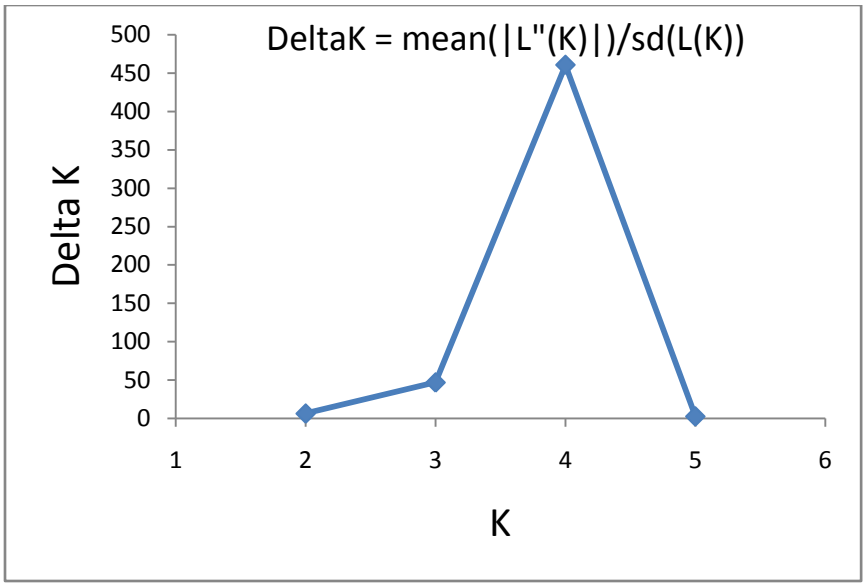
APPENDIX C



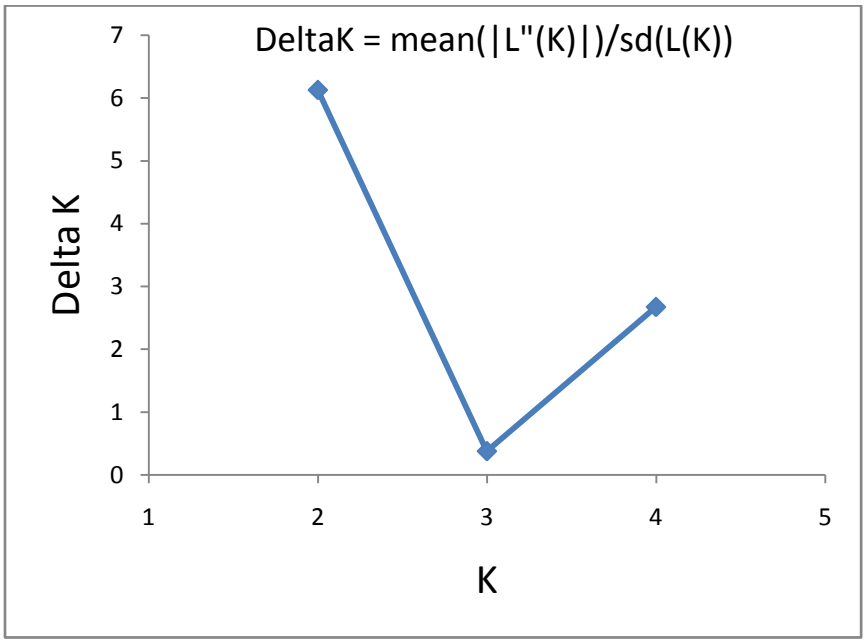
Delta K likelihood plot for all 10 initial populations output from Structure Harvester suggesting the most likely K value at this initial hierarchical level is 2.



Delta K likelihood plot for the 6 lakes north of the Kogaluk River fjord output from Structure Harvester suggesting the most likely K value at this hierarchical level is 5.



Delta K likelihood plot for the 4 lakes west and south of the Kogaluk River fjord output from Structure Harvester suggesting the most likely K value at this hierarchical level is 4.



Delta K likelihood plot for the lakes Esker and WP152 output from Structure Harvester suggesting the most likely K value at this hierarchical level is 2.

Appendix D. Average slope measurements for all pairs of lakes.

	Lake 1	Genetics H	Slushy	Esker	Strange	Wp152	T-Bone	Cabot	Genetics B	Hawk
Lake 1	0.000	-0.858	-0.807	-1.852	-0.513	-0.934	-0.355	-3.314	-1.453	-0.283
Genetics H	0.858	0.000	-0.664	-1.706	-0.353	-0.814	-0.280	-3.298	-1.410	-0.225
Slushy	0.807	0.664	0.000	-0.853	0.654	-0.258	0.027	-3.150	-1.218	0.010
Esker	1.852	1.706	0.853	0.000	1.507	0.402	0.337	-4.142	-1.314	0.222
Strange	0.513	0.353	-0.654	-1.507	0.000	-0.583	-0.129	-3.370	-1.353	-0.108
WP152	0.934	0.814	0.258	-0.402	0.583	0.000	0.209	-4.290	-1.408	0.133
T-Bone	0.355	0.280	0.027	0.337	-0.129	0.209	0.000	-4.066	-1.674	-0.014
Cabot	3.314	3.298	3.150	-4.142	3.370	4.290	4.066	0.000	2.722	5.270
Genetics B	1.453	1.410	1.218	-1.314	1.353	1.408	1.674	-2.722	0.000	5.106
Hawk	0.283	0.225	-0.010	-0.222	0.108	-0.133	0.014	-5.270	-5.106	0.000

APPENDIX E

Appendix E. LDNe estimate for lake Genetics H including the two putative immigrants, but excluding two random samples resulting in a very different estimation of N_e than when the putative immigrants were removed (see Table 3.5.1). Bolded column contains values appropriately comparable with those of table 3.5.1.

Lowest Allele Frequency Used (P_{crit})	**0.05**	**0.02**	**0.01**
Harmonic Mean Sample Size	67.2	**66.9**	67
Independent Comparisons	370	**472**	527
Estimated N_e	232	**310**	571
95% Confidence Interval (JackKnife on loci)	67.1 - ∞	**85.5 -** ∞	102.8 - ∞