# PHYLOGEOGRAPHY AND POST-GLACIAL DISPERSAL OF BROOK TROUT (SALVELINUS FONTINALIS) IN LABRADOR, CANADA, INFERRED FROM MICROSATELLITE AND MITOCHONDRIAL DNA VARIATION

BRETTNEY LOUISE PILGRIM

# PHYLOGEOGRAPHY AND POST-GLACIAL DISPERSAL OF BROOK TROUT (*SALVELINUS FONTINALIS*) IN LABRADOR, CANADA, INFERRED FROM MICROSATELLITE AND MITOCHONDRIAL DNA VARIATION

by

© Brettney Louise Pilgrim

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

April 2011

St. John's, Newfoundland and Labrador

**ABSTRACT**

Understanding how contemporary and historical factors have shaped genetic structure of populations is important for making management decisions, in order to characterise and preserve diversity. Genetic structure of brook trout (*Salvelinus fontinalis*) in Labrador and eastern Canada was investigated utilizing microsatellite and mitochondrial DNA (mtDNA) variation to collect baseline data so future impacts of the Trans-Labrador Highway may be monitored, and to investigate post-glacial recolonization. Genetic variation was moderate ($H_E$ = 0.622) to low (overall pairwise sequence diversity = 0.00094) based on microsatellite and mtDNA variation, respectively. Investigation of mtDNA diversity provided evidence that the mitochondrial genome of brook trout (GenBank AF15480) may have recombined with Artic charr mtDNA. Microsatellite variation was able to detect structure, which revealed limited ongoing gene flow at the level of the watershed. Finally, both microsatellite and mtDNA variation revealed dual routes of colonization in Labrador. Overall, my results emphasize the strong influence of both contemporary and historical events on shaping population genetic structure of brook trout in Labrador.

**ACKNOWLEDGEMENTS**

First off, I would like to sincerely thank my supervisor, Dr. H. Dawn Marshall, for giving me the opportunity to pursue my graduate studies in her lab. I am so grateful for the all of the guidance and encouragement she provided me through the process of completing my thesis; a task I could not complete without her constant support and advice. I am so grateful for all of the time she took to go through countless analyses and drafts with me, and her easy-going, patient and kind nature made her more than a pleasure to work for.

I would also like to thank Robert Perry and Donald Keefe of the Wildlife Division (Newfoundland and Labrador Department of Environment and Conservation) who were responsible for coordinating all of the sampling for this project. They were also kind enough to take me into the field with them, and were a lot of fun to work with and learn from. This project wouldn't be possible without the fieldwork and the financial support of the Newfoundland and Labrador Department of Environment and Conservation, as well as financial support from NSERC.

I am also grateful to Elizabeth Perry and Ed Yaskowiak of the Genomics and Proteomics (GaP) Facility (CREAIT Network, Memorial University of Newfoundland). They provided me with a considerable amount of help in learning all of the molecular techniques used in this study, and their friendly and helpful nature always made me feel very welcome in their lab.

I would also like to thank my committee, Dr. Craig Purchase and Dr. Steven M. Carr; I am sincerely grateful for the time and effort they put into reading my thesis, which they helped to improve with their thoughtful revisions and comments.

I am very lucky to have had the support of terrific lab mates and friends in the Marshall Lab, from those who helped directly and indirectly with my project including Justin Barron, Nadine Dunn, Barbara Langille, Jo-Anna Clark, Allison Grant, Greg Dale, and many more than can be mentioned here. They all made my time in this lab more than enjoyable. I am also thankful for the encouragement of my close friends who made my life during this time fun; without their support none of this would be possible. They were there when I needed a break to turn my brain off, and helped me make a lot fond memories of my time here in St. John's.

Finally, I owe so much to my family who were so patient, understanding and encouraging throughout my graduate studies and beyond. There was never a minute during the last two years that I didn't know I could count on them. Without their unconditional love and support I would never have been able to accomplish all of the goals I have set for myself.

**TABLE OF CONTENTS**

**LIST OF TABLES**

**LIST OF FIGURES**

## LIST OF ABBREVIATIONS AND SYMBOLS

| | |
|---|---|
| A | Adenine |
| $A$ | Allelic richness |
| AMOVA | Analysis of molecular variance |
| bp | Base pairs |
| C | Cytosine |
| °C | Degrees celsius |
| CO1 | Cytochrome oxidase one |
| CREAIT | Core Research Equipment and Instrument Training Network |
| $D_A$ | Nei *et al.*'s (1983) distance |
| DFO | Department of Fisheries and Oceans |
| DNA | Deoxyribonucleic acid |
| dNTP | Deoxyribonucleotide triphosphate |
| ESU | Evolutionary significant unit |
| $F_{CT}$ | Proportion of total variation among populations |
| $F_{IS}$ | Inbreeding coefficient |
| $F_{SC}$ | Proportion of total variation among subpopulations |

| | |
|---|---|
| $F_{ST}$ | Proportion of total variation within subpopulations |
| G | Guanine |
| GaP | Genomics and Proteomics |
| Gln | Glutamine |
| $H_E$ | Expected heterozygosity |
| $H_O$ | Observed heterozygosity |
| Ile | Isoleucine |
| $K$ | The number of genetic clusters as inferred by the program STRUCTURE |
| Leu | Leucine |
| ln Pr($X$\|$K$) | log likelihood of the data |
| MDSP | Multidimensional scaling plot |
| Met | Methionine |
| mm | Millimetre |
| mM | Millimole |
| mtDNA | Mitochondrial deoxyribonucleic acid |
| n | Number of repeats |

| | |
|---|---|
| $N_A$ | Number of alleles |
| NADH | Nicotinamide adenine dinucleotide |
| ND1 | Nicotinamide adenine dinucleotide dehydrogenase one |
| ng | Nanograms |
| NSERC | Natural Sciences and Engineering Research Council of Canada |
| $P$ | Probability |
| PCR | Polymerase chain reaction |
| RFLP | Restriction fragment length polymorphism |
| $R_{ST}$ | Between-population component of variance taking into account allele size |
| s | Seconds |
| SSRs | Simple sequence repeats |
| T | Thymine |
| *Taq* | *Thermus aquaticus* |
| tRNA | Transfer ribonucleic acid |
| U | Units |
| Val | Valine |

| | |
|---|---|
| μL | Microlitre |
| $\rho R_{ST}$ | Observed $R_{ST}$ between-population component of variance, taking into account allele size |

# INTRODUCTION

## *Phylogeography*

Phylogeography, the study of the spatial distribution of genetic variation of a species, is a field that contributes significantly to understanding the contemporary and historical factors that shape biodiversity. Many different disciplines contribute to this field, including biogeography, population genetics, molecular genetics, demography, and historical geography. Within a species, genetic variation will often occur within and between populations resulting in various spatial and temporal distributions of genetic structure. For example, populations across a specific geographic range may have either a mosaic or a homogeneous pattern of genotype distribution. These patterns of genetic structure can be influenced by one of many biological processes including mutation, migration, natural selection, random genetic drift, non-random mating and recombination. Determining the genetic structure of a species and which processes contribute to its current state can reveal important information about a species such as the dispersal of populations through a region, speciation, adaptation to past and present environmental changes, divergence of populations, and extinction; all processes that have great importance when considering the conservation of the genetic variation of a species.

*Conservation genetics*

One of the main goals in conservation biology is to protect the natural habitat and preserve a species' ability to evolve in that habitat (Moritz & Faith 1998; Frankham *et al.* 2002; Avise 2004). Among the most effective way to sustain a species' ability to evolve is to maintain a population's genetic diversity. Lower levels of genetic variation, especially in small population sizes, can negatively influence a species chance to adapt to habitat changes. Habitat changes can include events such as fragmentation, loss of habitat, pollution, climate change, over-exploitation, or the introduction of a non-native species. All of these events can result in changes to the genetic structure of a species: reduced gene flow can result in small isolated populations; a reduction in population size may lead to increased levels of inbreeding; or hybridized populations may replace "pure" ones. Conservation genetics aims to utilize the principles of population genetics to investigate contemporary structure; either to investigate if any of the above stated changes are ongoing in a population, or to monitor the effects of habitat change in the future. Current genetic structure can be quantified in the following ways: levels of genetic diversity within and between populations, mating patterns, kinship testing, levels of gene flow, levels of hybridization, and the phylogenetic relationships among populations (Avise 2004). Recent advances in molecular techniques, such as the ability to non-invasively screen and analyse DNA from a large number of samples, along with the ability to identify factors that should be considered when developing management plans (such as levels of gene flow), have contributed to the effectiveness of genetics as a tool for conservation research.

*Mitochondrial and microsatellite DNA as molecular markers in phylogeography and conservation genetics*

Mitochondrial DNA (mtDNA) is perhaps the most widely-employed and important molecule applied to the field of evolutionary biology (Avise *et al.* 1987; Moritz *et al.* 1987; Wilson *et al.* 1985). The uniparental, haploid nature of mtDNA results in a reduced effective population size compared to nuclear DNA, thus rendering it more sensitive to genetic drift (Birky *et al.* 1989). MtDNA variation is therefore highly influenced by founder events, making this an extremely useful marker when investigating evolutionarily divergent lineages (Avise 1989; Woodruff 1989; Dizon *et al.* 1992). A basic requirement for conserving biodiversity is to recognize and preserve populations that have a distinct long-term evolutionary history. In conservation biology, evolutionarily distinct populations are also referred to as evolutionary significant units (ESU; Ryder 1986); a concept developed in order to help prioritize specific populations for management. Since phylogeographic studies that utilize mtDNA can help identify these historically divergent populations, and how they are structured, they contribute significantly to defining both ESUs and provide insights into the evolutionary history of the population under investigation. However, it is important to use mtDNA markers in combination with information from nuclear genes as different dispersal patterns may be revealed due to male-biased dispersal or the higher mutation rate at microsatellite loci.

Microsatellites, or simple sequence repeats (SSRs), are selectively neutral nuclear DNA markers that are ubiquitous throughout the genome. They consist of tandem repeat units of 1 – 6 bp in length (e.g. a dinucleotide of $[CA]_n$ or a trinucleotide of $[GAT]_n$) that

have a high mutation rate due to DNA polymerase slippage during replication. As a result, microsatellites are highly variable markers that have the power to reveal subtler or more recently-evolved patterns of population structure than mtDNA, for example differentiation between closely-related populations of the same species. Although mtDNA evolves 5 – 10 times faster than nuclear DNA in general, this rate is still an order of magnitude lower than microsatellites, hence mtDNA variation is shaped more by historical events while contemporary events are better elucidated by investigating microsatellite variation. Since microsatellites are selectively neutral, it is important to note that these markers are useful for distinguishing levels of gene flow rather than reveal any information about selectively important adaptations that may be present between populations. Regardless, phylogeographical applications of microsatellite variation have been utilized to quantify measures important for conservation such as ongoing gene flow between populations, structure in closely-related populations, effective population size, past demographic bottlenecks, relatedness and levels of inbreeding.

*Phylogeography of freshwater fish in formerly glaciated North America*

Historical events have had a great impact on current genetic structure of freshwater fish, a group of organisms that is perhaps one of the most studied groups of vertebrates due to interest in fisheries management. From a conservation perspective, it is important to document the genetic structure of populations at both the local and regional scale in order to make appropriate management decisions, or to assist in identifying

ESUs. Phylogenetic studies of freshwater fish for fisheries management are useful not only for a conservation purpose, but also reveal important information about how historical processes have had a profound effect on shaping population structure. Determining the genetic structure of these populations provides the information needed to elucidate the effect of past processes such as gene flow and effective population sizes (Avise *et al.* 1984, 1988; Wilson *et al.* 1985), processes that have a relatively profound effect on the structure of freshwater fish populations due to the island-like structure of their habitats (Ward *et al.* 1994). The predominant influence of historical processes in shaping variation, especially that of mtDNA, has been continually revealed by phylogeographic studies.

One of the most significantly influential historical events that have contributed to the modern genetic structure of freshwater fish are the Pleistocene glaciations. During this period, viable habitat was continually buried and uncovered by massive glaciers which resulted in a "founder-flush cycle"; populations had to repeatedly abandon their native habitat, disperse to more southern regions, then recolonize their native habitat as glacier ice receded. The impact of these events was felt most significantly in the formerly glaciated portion of North America due to the size of the glacier sheets covering the northern region of the continent. Ice coverage reached a maximum 23 000 – 18 000 years ago during the Wisconsinan glaciation, the most recent of the North American glaciation events. Between 15 000 – 8 000 years ago, the last major deglaciation occurred (Dyke & Prest 1987) during which new available habitats for freshwater fish were formed, proglacial lakes and new rivers, from glacier ice melt (Pielou 1991). These facilitated

dispersal of freshwater fish considerably through formerly glaciated North America; the range size of present day freshwater fish species in the northern region of North America is relatively larger (McAllister *et al.* 1986) than those in from nonglaciated areas since these favourable dispersal conditions were in place (Hocutt & Wiley 1986; Dyke & Prest 1987). Although conditions for dispersal were favourable, a number of factors contributed to decreased levels of genetic variation in populations inhabiting the newly deglaciated northern region of North America. A small number of founders (small evolutionary effective population size) that were likely genetically similar due to restriction to glacial refugia are responsible for this decline in genetic diversity. For example, northern salmonids such as Atlantic salmon (*Salmo salar*) and arctic charr (*Salvelinus alpinus*) have relatively lower levels of mtDNA diversity (Bermingham *et al.* 1991; Tessier *et al.* 1995; Wilson *et al.* 1996) as compared to Pacific salmon (*Oncorhynchus spp.*) (Wilson *et al.* 1985, 1987; Thomas *et al.* 1986). The same pattern is found in lake whitefish (*Coregonus clupeaformis* in North America, *C. lavaretus* in Europe); European populations have much higher variation than those in North America due to a lower impact of the relatively smaller Eurasian ice sheets (Bernatchez *et al.* 1989; Bernatchez & Dodson 1991).

Two fish species, lake whitefish and lake trout (*Salvelinus namaycush*), serve as excellent models to examine the phylogeographic patterns and colonization histories of freshwater fish in formerly glaciated North America based on mtDNA variation. Bernatchez and Dodson (1991) conducted a widespread phylogeographical study of lake whitefish across their native range, one which corresponds closely with the Pleistocene

glacial limits. A restriction fragment length polymorphism (RFLP) analysis was carried out on 41 populations and revealed four clades, each of which correlates with a glacial refugium. Populations in Alaska, northern Canada, and some western Canadian populations were colonized by founders from a Beringian and northern Eurasian refugium. Whitefish originating from a Mississippian refugium had the greatest contribution, recolonizing some of western and all of central Canada. Southern Québec was colonized by whitefish from an Atlantic refugium, while eastern Canada was colonized by fish from an Acadian refugium. The pattern of mtDNA variation revealed very little secondary contact between populations, therefore the geographical correspondence to origins from specific glacial refugia was very clear. Patterns of mtDNA of lake trout revealed that these fish recolonized formerly glaciated North American in a slightly different manner (Wilson & Hebert 1996). Fish from the Beringian refugium, as opposed to the Mississippian for white fish, had the greatest contribution, colonizing regions including Alaska, the Canadian arctic, western and central Canada. Central Canada also had some contribution from a Mississippian refugium, while everything east of and including Québec was colonized from the Atlantic refugium. In contrast to whitefish, lake trout populations appear to have experienced extensive secondary contact among populations that were founded from the Beringian and Mississippian refugia (Wilson & Hebert 1996). Studies on these two species alone reveal the variation in colonization of freshwater fish through formerly glaciated North America, and the contribution and extent of dispersal of fish from different glacial refugia.

*Brook trout biology and brook trout in Labrador*

Brook trout are a primarily freshwater salmonid fish (Scott & Crossman 1973) most often inhabiting clean, cool streams, rivers and lakes (Behnke 1972). Although the structure of these habitats varies significantly in depth and water velocity, brook trout have successfully evolved in a variety of different habitats (Power 1980). Many examples of this variability, both phenotypic and genetic, exist in nature: populations living in sympatry exhibit morphological variation; variation exists in life history traits such that anadromous and freshwater residents occupy the same habitat (e.g. Power 1980; Jones *et al.* 1997; Boula *et al.* 2002,); genetic and morphological variation exists between littoral and pelagic forms (Dynes *et al.* 1999). Brook trout residing in different habitats also differ in life history traits such as growth rates and time of spawning. Species are able to successfully adapt to variable habitats if individuals are distributed randomly with respect to genotype as long as fitness is significantly influenced by the occupied habitat (Bradshaw 1965). It is likely that juvenile brook trout are distributed randomly with respect to genotype since siblings emerge at different times, and weak first-feeding fish face the challenge of swimming against various water current strengths (Field-Dodgson, 1988; Snucins, Curry & Gunn, 1992).

The brook trout's native range extends through eastern North America as far north as Labrador, Canada, and as far south as Georgia, USA. Since they are highly valued for their aesthetic and sport fish qualities and hence economically valuable, brook trout have been introduced and established in temperate regions on almost every continent (sometimes having adverse ecological effects). Currently brook trout are the most

important game fish in Labrador, Canada (DFO 2007), the most northeastern portion of their native range. Labrador composes the eastern-most part of Canadian Shield and consists of two climate types, polar tundra in the north and subarctic in the south. It is found in the Atlantic region of Canada, and is the mainland region of the province of Newfoundland and Labrador. Currently brook trout are under increased anthropogenic threat due to the recent opening of the Trans-Labrador Highway. This threat consists of two parts; first, habitat fragmentation that the highway may impose could interfere with migration, reducing gene flow, and subdividing current populations into smaller isolated populations. Secondly, there is the potential of increased fishing pressure as lakes become more accessible. As part of a ten-year monitoring program by the Government of Newfoundland and Labrador's Department of Environment and Conservation, the genetic structure of brook trout in Labrador will be documented to provide baseline information about current population structure required to monitor possible future impacts of the Trans-Labrador Highway. From a conservation and management perspective the contribution of contemporary and historical processes to the observed geographic pattern of genetic structure is important for identifying evolutionarily distinct lineages such that diversity may be preserved at the greatest possible level. Specifically, it is important to document both nuclear (microsatellite) and mtDNA variation to reveal both the contemporary and historical events that have shaped modern genetic structure.

*Brook trout genetics*

The influence of contemporary processes on current population genetic structure as revealed by microsatellite variation have been well documented across much of the brook trout's native range, including regions such as Maine (Castric *et al.* 2001), Québec (Angers *et al.* 1995, 1999; Angers & Bernatchez, 1998; Fraser *et al.* 2004), New Brunswick (Rogers *et al.* 2004), and Newfoundland (Adams *et al.* 2003). In each region, different contemporary factors were the main influence on shaping genetic structure. In Miramichi River, New Brunswick, and Mistassini Lake, Québec, ecological variables shaped patterns of genetic structure. Specifically in Miramichi River, alternative life history strategies (such as anadromy) and habitat selection influenced the pattern of genetic variation (Rogers *et al.* 2004), while dispersal patterns, specifically male-biased dispersal among tributaries, and female-biased dispersal from tributaries to outflow populations were the main factors in Mistassini Lake (Fraser *et al.* 2004). The structure of the habitat also plays a key role in shaping variation of brook trout populations (and other freshwater fishes). A number of studies across the brook trout's native range found that the connectivity of lakes within a watershed to be the main factor influencing the pattern of genetic diversity on a microgeographic scale (Angers *et al.* 1995, 1999; Angers & Bernatchez 1998; Castric *et al.* 2001; Adams *et al.*, 2003).

The pattern of mtDNA variation of brook trout has also been studied on a large-scale spatial basis across its native range. As mentioned previously, mtDNA is a useful marker for phylogeographical studies as it has a uniparental mode of inheritance and has an evolutionary rate that reflects historical differentiation and founding events. The

pattern of mtDNA variation through the brook trout's native range is strongly influenced by glacial events, as it is for other freshwater fish throughout formerly glaciated North America. Danzmann *et al.* (1998) carried out a large-scale phylogeographic study including 155 brook trout populations from eastern North America using restriction fragment length polymorphisms (RFLP). Large phylogenetic differences existed between populations found in previously glaciated northern regions and those found in more southern regions. Specifically, brook trout in eastern Canada were characterised to be genetically homogeneous, with such low genetic variation relative to the rest of the range that only 13 haplotypes out of a total of 61 characterized in the study were present. The majority of eastern Canada was colonized by hapolotype '1', which was postulated to have been present in both the Mississippian and Atlantic refugia. However, the presence of 10 private haplotypes (haplotypes found in a single region) combined with the lack of haplotype '2' fish provided evidence that there may have been some contribution of fish to this region from a more northeasterly refugium (Danzmann *et al.* 1998), also referred to as the Acadian refugium (Schmidt 1986). Brook trout populations in Labrador, the most northern portion of the brook trout's range, were not included in this study, therefore refugial origins of this region remain unknown.

Certain brook trout populations demonstrate the unique feature of an mtDNA of Arctic charr origin as a result of natural introgression. This was first documented by Bernatchez *et al.* (1995) in a brook trout population in Lake Alain, found in eastern Québec. Following this discovery, the watershed was further surveyed, revealing introgressed Arctic charr mtDNA in brook trout throughout the Rocheuse River

subdrainage of the Portneuf basin. Interestingly, introgression was restricted to this subdrainage alone, in allopatric brook trout populations where Arctic char did not reside. As a result, it was deduced that the timing of introgression was historical, illustrating a strong example of how historical processes have a significant influence on shaping modern genetic structure. This event may have happened as a result of historically larger founding populations of brook trout. Evidence for this lies in the fact that brook trout in the region have higher diversity (McGlade 1981) than Arctic char (Kornfield *et al.* 1981, Anderson *et al.* 1983), and therefore potentially a larger evolutionary effective population size. Hybridization is more likely to occur when one of the parental populations is much smaller (in this case the Arctic char) (Avise *et al.* 1988; Hubbs 1955) and hence liable to out-cross. Freshwater fish also have a number of traits that make them more susceptible to introgression such as competition for suitable habitat for spawning, external fertilization, and the susceptibility for recently-evolved species to have overlapping ranges (Hubbs 1995). This is especially true for brook trout in Québec where fish are approaching the northern end of their range. In addition, it is possible that brook trout with introgressed Arctic charr mtDNA could have a selective advantage. These brook trout would have respiratory enzymes encoded by their own nuclear DNA and Arctic charr mtDNA, which has evolved in a cold environment – potential advantage for brook trout found at the northern end of their range where conditions are much cooler (Glemet, Blier & Bernatchez 1998).

*Summary of goals*

As part of a ten-year monitoring program by the Government of Newfoundland and Labrador's Department of Environment and Conservation, one of the primary goals of this thesis is to measuring the genetic structure of brook trout populations in Labrador. This was carried out using two molecular markers, microsatellites and mtDNA. First, in order to collect baseline data important for monitoring the impact of the recent opening of the Trans-Labrador Highway, contemporary factors such as ongoing gene flow were investigated utilizing microsatellites. Additionally, microsatellites have also been successfully used to elucidate colonization patterns (Angers & Bernatchez 1998; Taylor and McPhail 2000; Lu *et al.* 2001; Koskinen *et al.* 2002), so I also seek to make inferences about post-glacial colonization based on these markers. Second, patterns of intraspecific haplotype and nucleotide diversity based on mtDNA variation will also be utilized to investigate colonization routes through Labrador to supplement patterns found using microsatellites.

**Microsatellite variation and genetic structure of brook trout (*Salvelinus fontinalis*) populations in Labrador and neighbouring Atlantic Canada: evidence for limited ongoing gene flow and dual routes of post-Wisconsinan colonization**

## ABSTRACT

In conservation genetics and management, it is important to understand the contribution of historical and contemporary processes to observed geographic pattern of genetic structure in order to characterise and preserve diversity. As part of a ten-year monitoring program by the Government of Newfoundland and Labrador, Canada, the population genetic structure of brook trout (*Salvelinus fontinalis*) in Labrador and the broader eastern Canadian region was investigated to monitor future impacts of the recently opened Trans-Labrador Highway. Patterns of genetic variation at six microsatellite loci were obtained from 1119 fish representing 32 populations from 8 local regions. Genetic diversity of Labrador brook trout (average $H_E = 0.622$) is typical of populations across their northeastern range. The high average pairwise $F_{ST}$ and AMOVA analysis indicate limited ongoing gene flow occurring among watersheds. STRUCUTURE analysis of two to four genetic groups and the multidimensional scaling plot indicated dual routes of post-Wisconsinan recolonization. Both analyses identified two genetic groups that separate northern and central Labrador from all remaining populations, consistent with two different post-glacial founding groups. These results

represent the first genetic data from the nuclear genome of brook trout in Labrador and emphasize the usefulness of microsatellite data for revealing the extent to which genetic structure is shaped by historical and contemporary processes.

## INTRODUCTION

The extent and structure of genetic variation across a species' range are informative about both historical processes, such as the pattern of recolonization after previous glacial maxima (Avise 2000), and contemporary ones, including ongoing dispersal and gene flow among populations. From a conservation and management perspective, it is important to understand the contribution of each type of process to the observed geographic pattern of genetic structure in order to identify evolutionarily distinct lineages and maximally preserve diversity. Effective management requires the maintenance of genetic diversity to sustain a population's ability to evolve (Frankham *et al.* 2002) and persevere through environmental changes (Moritz & Faith 1998; Avise 2004), such as climate change, biological invasions, or one of many anthropogenic threats like pollution, habitat fragmentation or over exploitation.

Brook trout (*Salvelinus fontinalis* Mitchill 1814) are a salmonid fish that resides primarily in freshwater (Scott & Crossman 1973), inhabiting the clear, cool waters of streams, rivers and lakes (Behnke 1972). Coastal populations with ocean access are often anadromous (Power 1980), which contributes to the ecological complexity of brook trout.

Due to their economic value as game fish, brook trout have been widely introduced to temperate regions outside their native range and now exist as numerous naturalized populations. Endemic to eastern North America (McCrimmon & Campbell 1969), the native range of the species includes much of eastern Canada, and extends southward to the northern tip of Georgia, USA. Labrador, the mainland component of the province of Newfoundland and Labrador in the Atlantic region of Canada, constitutes the northeastern-most portion of the native range. Labrador comprises the eastern-most part of Canadian Shield and consists of two climate types, polar tundra in the north and subarctic in the south. In Labrador, brook trout are the most important game fish (DFO 2007), and are under increased anthropogenic threat due to the recent opening of the Trans-Labrador Highway. In order to monitor future impacts of this highway on brook trout populations, assessment of the current population genetic structure throughout Labrador is crucial.

A number of studies have established the role of contemporary factors in shaping the genetic structure of brook trout populations across portions of the species' native range, including Maine (Castric *et al.* 2001), Québec (Angers *et al.* 1995, 1999; Angers & Bernatchez 1998; Fraser *et al.* 2004), New Brunswick (Rogers *et al.* 2004), and insular Newfoundland and Labrador (Adams *et al.* 2003). Ecological variables were found to affect the patterning of genetic structure in certain cases. For example, in Miramichi River, New Brunswick, population structure was mainly dictated by habitat selection and the life history strategy of anadromy (Rogers *et al.* 2004), while the structure of brook trout populations in Mistassini Lake, Québec, was attributed to dispersal patterns,

specifically male-biased dispersal among tributaries and female-biased dispersal from tributaries to outflow populations (Fraser *et al.* 2004). Another major factor that shapes the population genetic structure of freshwater fish, such as brook trout, is the structure of the habitat. In particular, the connectivity of lakes within watersheds has been found to influence the pattern of genetic diversity of brook trout on a microgeographic scale (Angers *et al.* 1995, 1999; Angers & Bernatchez 1998; Castric *et al.* 2001, Adams *et al.* 2003).

A species' genetic structure is also strongly influenced by historical events, like the Wisconsinan glaciation which reached its maximum 18,000 years ago. Relative to recently re-established northern populations, populations that persisted in isolation in southern refugia have had time to accumulate variation (Hewitt 1996, 1999; Bennett 1997) while simultaneously diverging from populations in other refugia (Hewitt 1996, 2004). This signature is reflected in the distribution and genetic structure of populations today - northern regions typically have fewer species (Pielou 1991) characterized by lower genetic diversity, a pattern that is referred to by Hewitt (1996) as 'southern richness, northern purity'. The pattern of population genetic structure therefore reveals the direction of colonization and the contribution of distinct refugial origins to the present-day species' distribution.

The pattern of colonization of brook trout throughout their native range was assessed by Danzmann *et al.* (1998) on the basis of mtDNA diversity. Brook trout in eastern Canada were shown to possess very little genetic variation relative to the rest of the range, containing only 13 haplotypes of a possible 61. The majority of this range was

colonized by haplotype '1', which was postulated to have been present in both the Mississippian and Atlantic refugia. However, the presence of 10 private haplotypes (haplotypes found in a single region) combined with the lack of haplotype '2' fish suggests that fish from this region may have originated from a more northeasterly refugium (Danzmann *et al.* 1998), also referred to as the Acadian refugium (Schmidt 1986).

As part of a ten-year monitoring program by the Government of Newfoundland and Labrador's Department of Environment and Conservation, one of the goals of this chapter was to measure the genetic structure of brook trout populations in Labrador as revealed by nuclear microsatellite markers. Such analyses provide baseline information about current population structure needed to monitor future impacts of the Trans-Labrador Highway. In absence of a clear elucidation by mtDNA variation of the ancestral origins of northeastern brook trout, I also sought to infer possible routes of post-glacial colonization of brook trout across the Atlantic region, and in particular, to investigate the number and source of refugial populations that colonized Labrador.

## METHODS AND MATERIALS

### *Sample collection*

Brook trout from 31 lakes within 18 watersheds in six regions across eastern Canada (northern, central and southeastern Labrador, Newfoundland, Nova Scotia and New Brunswick; Figure 1.1) were collected by provincial wildlife biologists between

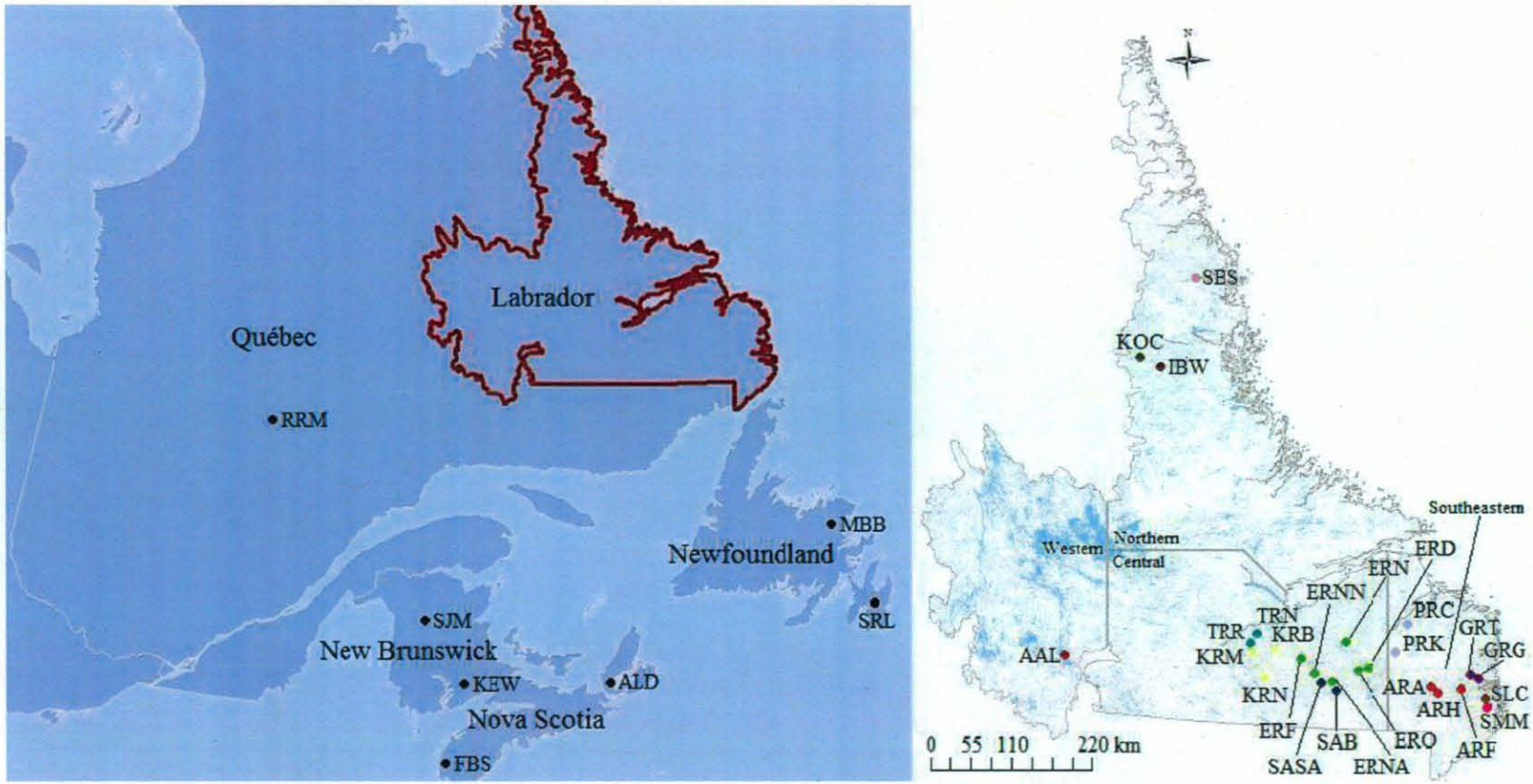


Figure 1.1. Lakes sampled for brook trout across eastern Canada. Population codes are given in Table 1.2.

2003–2008 (1072 samples). Brook trout were sampled via gill netting. Three to four nets were set in random locations through a lake at a time, collected after a period of five to six hours, then re-set in new locations overnight, and collected the next day. This sampling scheme took place in each lake over a period of approximately seven days. Fin clips were stored at -20°C. Brook trout DNA extractions from fish in Mistassini Lake, Québec (11 samples) and Atikonak Lake, western Labrador (36 samples) were provided by Dr. Louis Bernatchez (Université Laval) and Dr. Steven M. Carr (Memorial University of Newfoundland and Labrador), respectively.

*Microsatellite analysis*

Genomic DNA was extracted from a portion (~25 $mm^2$) of the caudal fin clip with a Qiagen QIAmp DNA Mini Kit (Qiagen Inc., Mississauga, Ontario, Canada) following the tissue protocol. Nine microsatellite loci were initially tested: *Sfo*12, *Sfo*18 (Angers *et al.* 1995), *MST*85 (Presa and Guyomard 1996), *Sfo*D91, *Sfo*B52, *Sfo*C129, *Sfo*D100, *Sfo*C86, *Sfo*D75 (T.L. King, US Geological Survey, unpublished). Subsequently the six loci with the greatest amplification success rate (*Sfo*18, *Sfo*B52, *Sfo*C129, *Sfo*D100, *Sfo*C86, and *Sfo*D75) were selected for routine analysis. PCRs were carried out in an Applied Biosystems GeneAmp PCR System 9700 Thermal Cycler (Applied Biosystems Inc., Foster City, CA, USA) and contained 1X PCR Master Mix (Promega Corp., USA), 2 µL 10 mM of each primer and 2 µL genomic DNA (2– 570 ng). Four fluorescent dyes were used to label forward primers as follows: PET – *Sfo*12; VIC – *Sfo*18, *MST*85 and

*Sfo*C129; 6-FAM – *Sfo*D91, *Sfo*B52 and *Sfo*C86; NED – *Sfo*D75. One of two PCR profiles was utilized. For primer pairs *Sfo12, Sfo18* and *MST85*, there was an initial denaturation step at 95°C for 4 min, then 35 cycles of 95°C for 1 minute, 5°C for 45 s, and 72°C for 45 s, followed by a final elongation step at 72°C for 7 min. For the remaining loci, there was an initial denaturation step at 95°C for 4 min, followed by 35 cycles of 94°C for 45 s, 56°C for 45 s and 72°C for 1.5 min, and an elongation step of 72°C for 5 min. Prior to electrophoresis, samples were diluted (1:4 dilution), and 1 µL of sample was added to 8.8 µL of formamide and 0.2 µL of the LIZ-500 size standard. Loci labelled with different dyes and that did not have size range overlaps were run concurrently as follows: i) *Sfo*18; ii) *Sfo*B52, *Sfo*C129, *Sfo*D100; iii) *Sfo*C86, *Sfo*D75. Samples were electrophoresed in an Applied Biosystems 3730 DNA Analyzer using GeneScan Software (Applied Biosystems Inc.).

*Data analysis*

Genetic variation at each locus was measured as the observed ($H_O$) and expected ($H_E$) heterozygosity and the number of alleles ($N_A$) in GENEPOP 1.2 (Raymond & Rousset 1995). The inbreeding coefficient ($F_{IS}$) at each locus was estimated with the Weir and Cockerham (1984) algorithm as implemented in FSTAT 2.9.3 (Goudet 1995) (120 randomizations). Linkage disequilibrium among all pairs of loci was evaluated with the log likelihood ratio test (G-test) of GENEPOP using the default Markov chain parameters.

Diversity measures ($N_A$, $H_O$, and $H_E$) were also calculated in GENEPOP for all brook trout, brook trout in each population (defined as all brook trout from the same lake), as well populations within each region. Allelic richness ($A$) for each population, and $F_{IS}$ (using Weir & Cockerham's (1984) algorithm, over 3840 randomizations) for each population at each locus was calculated in FSTAT.

In order to determine whether mutation has influenced differentiation in addition to drift, the allele size randomization test was utilized (following Hardy *et al.* 2003) as implemented in SPAGeDi 1.3 (Hardy & Vekemans 2002). Random permutation of alleles (2000 iterations) provides a 95% confidence interval of the simulated distribution of $R_{ST}$ (between-population component of variance taking into account allele size) values, (or ρRST). If $R_{ST} > F_{ST}$, $R_{ST}$ will fall outside of the $\rho R_{ST}$ 95% confidence interval, which would indicate that differentiation has been influenced by mutation. If $R_{ST}$ falls within this range, $R_{ST}$ does not significantly differ from $F_{ST}$, indicating that drift mainly contributes to differentiation. Since $R_{ST} = F_{ST}$, pairwise $F_{ST}$ values, as estimated in ARLEQUIN 3.1 (Excoffier *et al.* 2005), were used in subsequent analyses. Due to greater variance, $R_{ST}$ is less reliable at detecting population differentiation than $F_{ST}$ (Sefc *et al.* 2007).

To assess the partitioning of genetic diversity among populations, an analysis of molecular variance (AMOVA) was carried out in ARLEQUIN 3.1 (Excoffier *et al.* 2005), with three hierarchical levels: (i) lakes within watersheds, (ii) lakes within regions, and (iii) watersheds within regions. A multi-dimensional scaling plot (MDSP) based on pairwise $F_{ST}$ values was visualized in XLSTAT 2010.1.03 (Addinsoft SARL, USA) to

investigate any correlation between genetic relationships and geographical locations among populations. In addition, Nei *et al.*'s (1983) distance $D_A$; calculated with MICROSATELLITE ANALYSER 4.05 [Dieringer & Schlötterer 2003]), was used to assess distance-based relationships among populations. The neighbour-joining dendrogram was generated using the NEIGHBOUR and CONSENSE modules in PHYLIP 3.69 (Felsenstein 1989). The tree was visualized in TREEVIEW 1.4 (Page 1996).

STRUCTURE 2.1 (Pritchard *et al.* 2000) was used to identify the number of genetically distinct clusters ($K$) across all sampled lakes and to assign individuals to clusters based upon multi-locus genotypes. For each value of $K$, ten iterations were run to assess convergence of the likelihood with a burn-in period of 100 000, followed by 500 000 iterations for values of $K = 1$ through 33. Each simulation was performed with an ancestry model incorporating admixture, a model of correlated allele frequencies, without prior population information (Falush *et al.* 2003).

## RESULTS

### *Per-locus diversity and linkage disequilibrium*

A total of 1119 brook trout from 33 lakes within 20 watersheds across eight regions of eastern Canada were genotyped at six microsatellite loci. An average of 13.5

alleles was found across all six loci, with a range of eight alleles at *SfoC129* to 19 alleles at *Sfo18*. Average observed heterozygosity was 0.655, with a range from 0.526 (*SfoC129*) to 0.809 (*SfoD100*). Deviations from Hardy-Weinberg equilibrium were observed at four loci: $F_{IS}$ was significantly greater than zero for three loci (*Sfo18*, *SfoB52*, *SfoD75*) and significantly less than zero for one locus (*SfoC86)* after correction for multiple comparisons (Table 1.1; $\alpha = 0.00833$) as analysed in FSTAT. Significant linkage disequilibrium ($P < 0.05$) after Bonferroni correction was found for only 4 of 495 pairwise comparisons of loci by the log likelihood ratio test (G-test) of GENEPOP.

*Diversity by lake and region*

Within-lake genetic diversity was highest in fish from Salmon River in Nova Scotia ($H_E = 0.682$), and lowest in fish from Dead Dog Pond in central Labrador ($H_E = 0.346$). Among regions, genetic diversity was highest in Québec (0.659) and lowest across central Labrador (0.613) (Table 1.2). Mean allelic richness ($A$) was 3.5 overall, ranging from 2.5 in Dead Dog Pond (central Labrador) to 4.3 in Mary's Harbour Big Pond (southeastern Labrador).

$F_{IS}$ was significantly negative in 11 of 33 lakes at $P < 0.01$ (Table 1.3) after correction for multiple tests ($\alpha = 0.0003$). No significant $F_{IS}$ values remained at $P < 0.001$ (Table 1.3) after Bonferroni correction (corrected $\alpha = 0.00003$). Despite a number of $F_{IS}$ values within lakes being significantly negative, no per locus measures of $F_{IS}$

Table 1.1. Allelic variation at each microsatellite locus. Number of alleles ($N_A$), observed ($H_O$) and expected ($H_E$) heterozygosity and $F_{IS}$ (Wright's inbreeding coefficient). Significant heterozygote excess/deficit at $P < 0.05$ after Bonferroni correction ($\alpha = 0.00833$) denoted with an asterisk.

| **Locus** | $N_A$ | $H_O$ | $H_E$ | $F_{IS}$ |
|---|---|---|---|---|
| Sfo18 | 19 | 0.599 | 0.759 | 0.211* |
| SfoB52 | 15 | 0.576 | 0.644 | 0.105* |
| SfoC129 | 8 | 0.526 | 0.535 | 0.016 |
| SfoD100 | 12 | 0.809 | 0.781 | -0.036 |
| SfoC86 | 10 | 0.801 | 0.571 | -0.403* |
| SfoD75 | 17 | 0.619 | 0.701 | 0.118* |
| Overall/average | 14 | 0.655 | 0.665 | 0.015 |

Table 1.2. Locations with representative code, sample sizes (N), observed ($H_O$) and expected ($H_E$) heterozygosity, and allelic richness ($A$, corrected to n = 7) of each lake and across each region.

| Region | Population Code | Watershed | Lake | N | $H_E$ | $H_O$ | $A$ |
|---|---|---|---|---|---|---|---|
| **Northern Labrador** | | | | **36** | **0.621** | **0.637** | **3.8** |
| | SBS | Saputit Brook | Saputit Lake | 7 | 0.625 | 0.762 | 3.7 |
| | KOC | Kogaluk River | Cabot Lake | 16 | 0.544 | 0.616 | 2.8 |
| | IBW | Iladlivik Brook | Walkabout Lake | 13 | 0.601 | 0.596 | 3.4 |
| **Western Labrador** | | | | **36** | **0.630** | **0.579** | **4.1** |
| | AAL | Atikonak Lake | Atikonak River | 36 | 0.630 | 0.579 | 4.1 |
| **Central Labrador** | | | | **519** | **0.613** | **0.609** | **4.1** |
| | TRN | Traverspine River | No Boat Pond | 45 | 0.451 | 0.417 | 3.2 |
| | TRR | Traverspine River | The Right Lake | 17 | 0.525 | 0.546 | 3.5 |
| | KRM | Kenamu River | Mercier Lake | 15 | 0.540 | 0.450 | 3.2 |
| | KRB | Kenamu River | Brennan Lake | 58 | 0.647 | 0.750 | 3.7 |
| | KRN | Kenamu River | Nikki's Pond | 50 | 0.516 | 0.588 | 3.0 |
| | ERF | Eagle River | Fred's Lake | 53 | 0.637 | 0.742 | 3.8 |
| | ERNN | Eagle River | No Name Lake | 27 | 0.570 | 0.606 | 3.5 |
| | ERNA | Eagle River | Nap Pond | 33 | 0.582 | 0.669 | 3.4 |
| | ERO | Eagle River | Osprey Lake | 37 | 0.571 | 0.662 | 3.1 |
| | ERD | Eagle River | Dead Dog Pond | 11 | 0.346 | 0.313 | 2.5 |
| | ERN | Eagle River | Nippard's Lake | 60 | 0.541 | 0.626 | 3.0 |
| | SASA | St. Augustine River | St. Augustine | 54 | 0.543 | 0.580 | 3.0 |
| | SAB | St. Augustine River | Bog Lake | 59 | 0.519 | 0.567 | 3.0 |
| **Southeastern Labrador** | | | | **400** | **0.633** | **0.701** | **4.9** |
| | PRK | Paradise River | Keith's Lake | 41 | 0.531 | 0.579 | 3.3 |
| | PRC | Paradise River | Crooked Lake | 55 | 0.538 | 0.496 | 3.7 |
| | ARA | Alexis River | Alexis Pond | 45 | 0.630 | 0.818 | 3.6 |
| | ARH | Alexis River | Handkerchief Pond | 29 | 0.524 | 0.660 | 3.2 |
| | ARF | Alexis River | Feeder Pond | 24 | 0.564 | 0.648 | 3.4 |
| | GRT | Gilbert River | Tilt Pond | 46 | 0.580 | 0.722 | 3.6 |
| | GRG | Gilbert River | Gilbert Lake | 60 | 0.588 | 0.721 | 3.7 |
| | SLC | St. Lewis River | Curl's Pond | 61 | 0.651 | 0.859 | 4.0 |
| | SMM | St. Mary's River | Mary's Harbour Big Pond | 39 | 0.656 | 0.741 | 4.3 |
| **Newfoundland** | | | | **54** | **0.629** | **0.613** | **4.5** |
| | SRL | Salmonier River | Little Gull Pond | 29 | 0.550 | 0.595 | 3.2 |
| | MBB | Middle Brook | Butt's Pond | 25 | 0.606 | 0.639 | 3.9 |
| **Nova Scotia** | | | | **50** | **0.662** | **0.723** | **5.2** |
| | ALD | Alder Brook | River Denys | 30 | 0.593 | 0.684 | 4.1 |
| | FBS | Farnham Brook | Salmon River | 20 | 0.682 | 0.782 | 4.2 |
| **New Brunswick** | | | | **49** | **0.643** | **0.751** | **5.0** |
| | SJM | Saint John River | Moose Lake | 25 | 0.599 | 0.690 | 3.9 |
| | KEW | Kennebecasis River | Walton Lake | 24 | 0.618 | 0.813 | 3.6 |
| **Québec** | | | | **11** | **0.659** | **0.742** | **4.1** |
| | RRM | Rupert River | Mistassini Lake | 11 | 0.659 | 0.742 | 4.1 |
| Overall/average | | | | 1119 | 0.573 | 0.646 | 3.5 |

Table 1.3. Per locus measures of $F_{IS}$ of brook trout for each lake. Significant heterozygote excess/deficit at $P < 0.01$ after Bonferroni correction ($\alpha = 0.0003$) denoted by bolded values.

| | Locus: | | | | | | |
|---|---|---|---|---|---|---|---|
| Population | Sfo18 | SfoB52 | SfoC129 | SfoD100 | SfoC86 | SfoD75 | All |
| SBS | -0.091 | -0.111 | -0.2 | -0.556 | -0.304 | -0.231 | -0.243 |
| KOC | -0.159 | -0.375 | 0.063 | 0 | -0.244 | -0.136 | -0.138 |
| IBW | 0.136 | 0.117 | 0.286 | -0.609 | -0.343 | 0.298 | 0.008 |
| AAL | 0.292 | 0.032 | -0.105 | -0.104 | 0.112 | 0.17 | 0.084 |
| TRN | 0.147 | 0.087 | 0.124 | 0.03 | -0.197 | 0.313 | 0.075 |
| TRR | -0.086 | 0.223 | -0.185 | -0.118 | -0.143 | -0.045 | -0.042 |
| KRM | -0.1 | 0.798 | 0.03 | -0.1 | -0.13 | 0.48 | 0.172 |
| KRB | 0.094 | -0.041 | -0.249 | -0.157 | -0.83 | 0.034 | **-0.161** |
| KRN | -0.061 | 0.132 | -0.107 | -0.368 | -0.301 | -0.061 | **-0.141** |
| ERF | 0.25 | 0.002 | -0.556 | -0.105 | -0.827 | -0.074 | -0.167 |
| ERNN | 0.498 | -0.247 | -0.368 | -0.126 | -0.34 | 0.058 | -0.064 |
| ERNA | 0.116 | -0.114 | -0.078 | -0.124 | -0.541 | -0.166 | -0.152 |
| ERO | 0.099 | -0.048 | -0.281 | -0.081 | -0.692 | -0.093 | -0.161 |
| ERD | 0.815 | 0.226 | 0 | -0.263 | -0.346 | -0.053 | 0.101 |
| ERN | 0.134 | -0.042 | -0.159 | -0.113 | -0.663 | -0.225 | **-0.159** |
| SASA | 0.152 | 0.01 | 0.136 | -0.134 | -0.509 | -0.136 | -0.069 |
| SAB | 0.043 | 0.071 | -0.016 | -0.077 | -0.564 | 0.003 | -0.094 |
| PRK | 0.194 | -0.024 | -0.201 | -0.124 | -0.365 | -0.169 | -0.092 |
| PRC | 0.433 | 0.109 | -0.058 | -0.026 | -0.312 | 0.036 | 0.078 |
| ARA | -0.238 | -0.109 | -0.575 | -0.099 | -0.638 | -0.265 | **-0.303** |
| ARH | -0.113 | -0.096 | -0.575 | -0.293 | -0.449 | 0.02 | **-0.265** |
| ARF | 0.155 | -0.112 | -0.162 | -0.038 | -0.852 | -0.052 | -0.153 |
| GRT | 0.024 | -0.1 | -0.353 | -0.176 | -0.607 | -0.258 | **-0.247** |
| GRG | -0.054 | -0.13 | -0.152 | -0.197 | -0.586 | -0.215 | **-0.227** |
| SLC | -0.082 | -0.115 | -0.707 | -0.279 | -0.651 | -0.216 | **-0.322** |
| SMM | -0.273 | -0.146 | -0.067 | -0.035 | -0.439 | 0.121 | **-0.132** |
| SRL | -0.068 | 0.03 | -0.087 | -0.314 | -0.12 | 0.05 | -0.083 |
| MBB | -0.034 | 0.092 | -0.11 | -0.077 | -0.249 | 0.027 | -0.055 |
| ALD | -0.185 | -0.025 | 0.111 | -0.214 | -0.45 | -0.14 | -0.156 |
| FBS | -0.018 | -0.188 | -0.015 | -0.005 | -0.638 | -0.152 | -0.15 |
| SJM | 0.089 | 0.17 | -0.276 | -0.2 | -0.648 | 0.008 | -0.155 |
| KEW | -0.048 | -0.169 | -0.363 | -0.206 | -0.92 | -0.392 | **-0.324** |
| RRM | -0.194 | 0.114 | -0.127 | -0.467 | -0.071 | -0.087 | -0.133 |

deviated significantly from zero at $P < 0.001$ after correction for multiple tests ($\alpha = 0.000005$).

*Drift contributes more substantially to differentiation than mutation*

According to the per-locus and global measures of $\rho R_{ST}$ and $R_{ST}$ presented in Table 1.4, all observed $R_{ST}$ values were within the 95% confidence interval of $\rho R_{ST}$, and $R_{ST}$ (0.150) was not significantly different from $F_{ST}$ (0.154) ($P = 0.954$). This indicates that drift has contributed more to differentiation than has mutation, and therefore $F_{ST}$ rather than $R_{ST}$ is used in subsequent analyses investigating population genetic structure.

*Population differentiation*

Pairwise $F_{ST}$ values were calculated to determine if there was any sub-population differentiation. Mean $F_{ST}$ measured among pairs of lakes was 0.149 and all but three comparisons (Nap Pond-Osprey Lake, St. Augustine-Bog Lake, Butt's Pond-Nap Pond) were significantly positive ($P < 0.05$). The lowest significant $F_{ST}$ observed was 0.0054 between Bog Lake and St. Augustine River (both lakes within St. Augustine watershed, central Labrador), and the highest was 0.403 between Mistassini Lake (Québec) and Dead Dog Pond (southeastern Labrador).

The pattern revealed by the multi-dimensional scaling plot (MDSP) of pairwise $F_{ST}$ values corresponds strongly to the geography of the regions. Overall, there seems to

Table 1.4. Summary of allele size permutation test as implemented in SPAGeDi showing estimates of $R_{ST}$, simulated $R_{ST}$ ($\rho R_{ST}$) values and their 95% confidence interval, $F_{ST}$, and at $P$ values following 2000 allele permutations.

| Locus | $R_{ST}$ | $\rho R_{ST}$ | 95% CI | $F_{ST}$ | $P$ |
|---|---|---|---|---|---|
| *Sfo*18 | 0.164 | 0.132 | 0.051 - 0.221 | 0.161 | 0.525 |
| *Sfo*B52 | 0.132 | 0.113 | 0.063 - 0.163 | 0.128 | 0.521 |
| *Sfo*C129 | 0.176 | 0.170 | 0.131 - 0.201 | 0.191 | 0.864 |
| *Sfo*D100 | 0.134 | 0.093 | 0.045 - 0.151 | 0.106 | 0.194 |
| *Sfo*C86 | 0.052 | 0.080 | 0.033 - 0.152 | 0.082 | 0.623 |
| *Sfo*D75 | 0.147 | 0.154 | 0.075 - 0.211 | 0.186 | 0.791 |
| Global | 0.150 | 0.148 | 0.101 - 0.192 | 0.154 | 0.954 |

be two groups of lakes; one that includes lakes in central and northern Labrador, and another that includes lakes in all other regions (Figure 1.2). Lakes within the same watershed throughout Labrador generally cluster together; for example, Gilbert Lake and Tilt Pond, both within the Gilbert River watershed, are adjacent to each other on the scaling plot, as are all lakes within the Eagle River watershed. Watersheds also cluster into groups based on region. Exceptions to this pattern of regional clustering include Butt's Pond, found in eastern Newfoundland, clusters with central Labrador, while Nikki's Pond, found in central Labrador, clusters with populations from Newfoundland, Nova Scotia and New Brunswick. The Mistassini Lake (Québec) and No Boat Pond (Traverspine River) populations were clearly separated from all other regions.

*Hierarchical AMOVA and population structure*

Three hierarchical AMOVA analyses detected significant geographic structuring within and among populations and regions (Table 1.5). The first defined groups by lakes, then watersheds. Most of the variation was detected within lakes (85.70%), followed by between watersheds (8.38%), then among lakes within watersheds (5.92%). All other comparisons had qualitatively similar results. The second analysis grouped fish by lake, then region. Most of the variation was partitioned into lakes (84.34%), followed by among lakes within regions (9.03%), with the least amount of variation found between regions (6.62%). Finally, fish were grouped by watershed, then into regions. Most of the variation was detected within watersheds (86.70%), followed by between watersheds

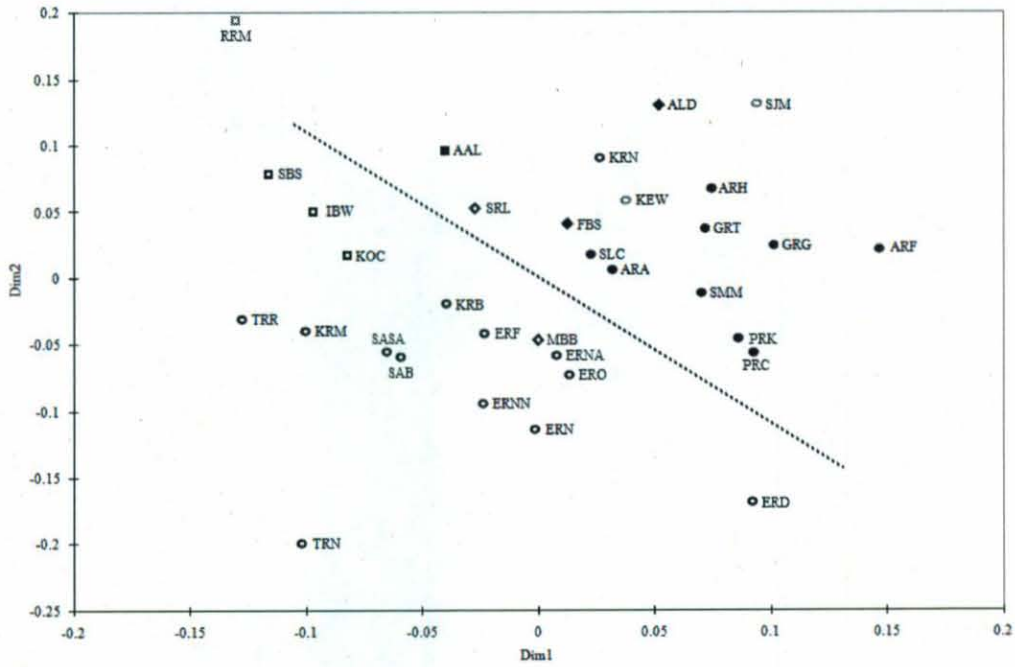


Figure 1.2. A multi-dimensional scaling plot based on pair-wise $F_{ST}$ values among all populations. ◘ = Northern Labrador ◙ = Québec ◈ = Newfoundland ● = Southeast Labrador ◎ = New Brunswick ◆ = Nova Scotia ■ = Western Labrador ○ = Central Labrador. Population codes are given in Table 1.2

Table 1.5. Hierarchical analysis of molecular variance in brook trout populations across eastern Canada performed by grouping fish into populations (either by lake or watershed), then subsequently into either watersheds or regions. Percentage of total variance (%) and $F$-statistics ($F_{CT}$, $F_{SC}$, $F_{ST}$) for each hierarchical level represented.

| | Among regions/ watersheds | | Among populations within watersheds/regions | | Within populations | |
|---|---|---|---|---|---|---|
| Comparison | % | $F_{CT}$ | % | $F_{SC}$ | % | $F_{ST}$ |
| i. Lakes grouped into watersheds | 8.38 | 0.084 | 5.92 | 0.065 | 85.7 | 0.143 |
| ii. Lakes grouped regionally | 6.62 | 0.066 | 9.03 | 0.097 | 84.34 | 0.157 |
| iii. Watersheds grouped regionally | 6.28 | 0.063 | 7.02 | 0.075 | 86.7 | 0.133 |

within regions (7.02%), and the least amount of variation was partitioned among regions (6.28%). Overall, the majority of variation is within lakes, and the next most explanatory level is watershed. There is weak but significant differentiation at each spatial scale.

Genetic structure was also determined using Nei's genetic distance ($D_A$) (Nei *et al.* 1983), which has been suggested to be the most appropriate to use when drift is the main force contributing to differentiation (Paetkau *et al.* 1997). The neighbour-joining dendrogram based on this distance illustrates that differentiation among populations relates spatially to region (Figure 1.3); populations from the Atlantic provinces and Québec, southeastern Labrador, central Labrador and northern Labrador each constitute their own cluster, with a few exceptions. One Newfoundland population, Butt's Pond, clusters with central Labrador populations, and Nikki's Pond (from the Kenamu River in central Labrador) is more closely related to northern Labrador populations. Southeastern Labrador is composed of two separate clusters, one of which comprises populations from Paradise River only. Similarly, central Labrador constitutes two clusters.

A plot of the log likelihood (ln Pr($X|K$)) values obtained from the STRUCTURE analysis for 1 – 15 distinct genetic groups ($K$) is presented in Figure 1.4. The likelihood generally increases from $K = 1$ to $K = 9$ and declines from $K = 10$ to $K = 15$, plateauing from ~ $K = 5$ to $K = 11$. At $K = 2$ to $K = 3$ a decline of the slope of the curve is apparent while the highest ln Pr($X|K$) value is associated with $K = 9$. Bar plots of estimated individual cluster membership coefficients for $K = 2$, 3, 4 and 9 are presented in Figure 1.5 and pie charts of these memberships on a watershed basis are mapped in Figure 1.6.

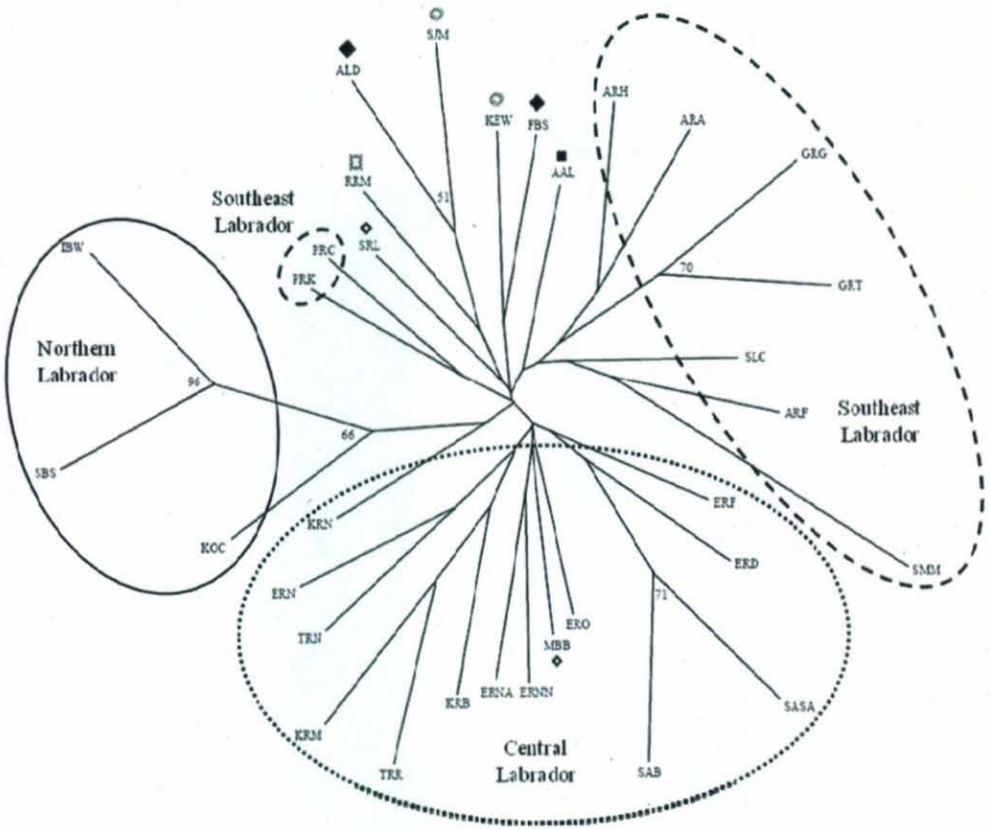


Figure 1.3. Unrooted neighbour-joining dendrogram based on Nei *et al.*'s ($D_A$) distance (1983) for the 33 brook trout populations sampled across eastern Canada. Values at the nodes represent bootstrap resampling percentages greater than 50% (based on 2000 replicates). Population codes are given in Table 1.2. ⧈ = Québec ◆ = Newfoundland ⬯ = New Brunswick ◆ = Nova Scotia ■ = Western Labrador

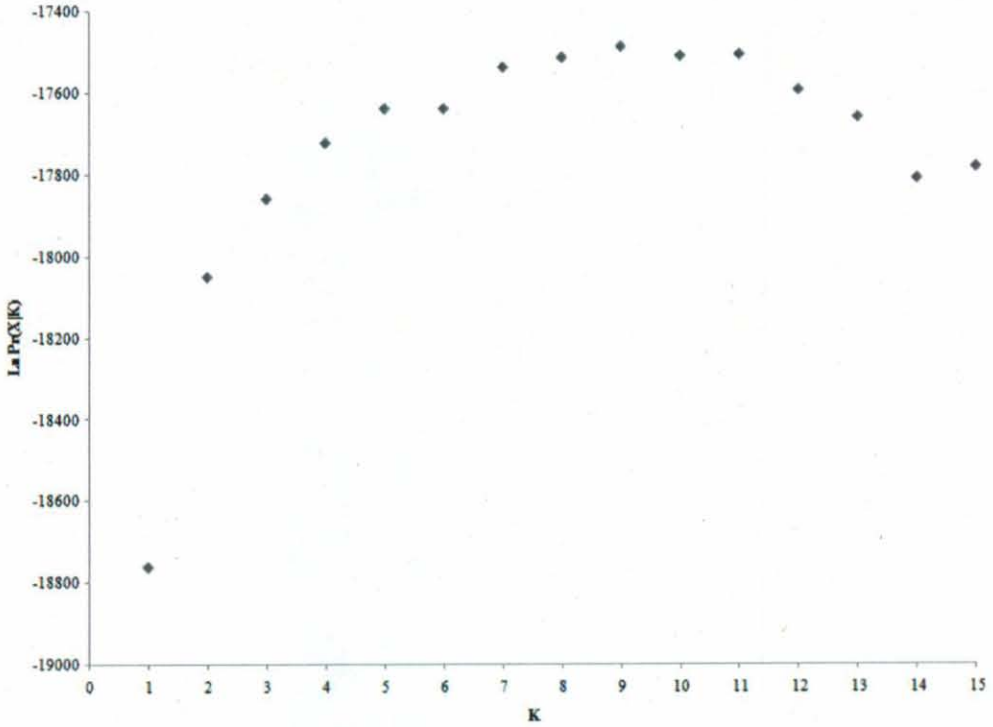


Figure 1.4. A plot of the log likelihood of the data (ln Pr(X|K)) values obtained from the STRUCTURE analysis for 1 – 15 distinct genetic groups (*K*).

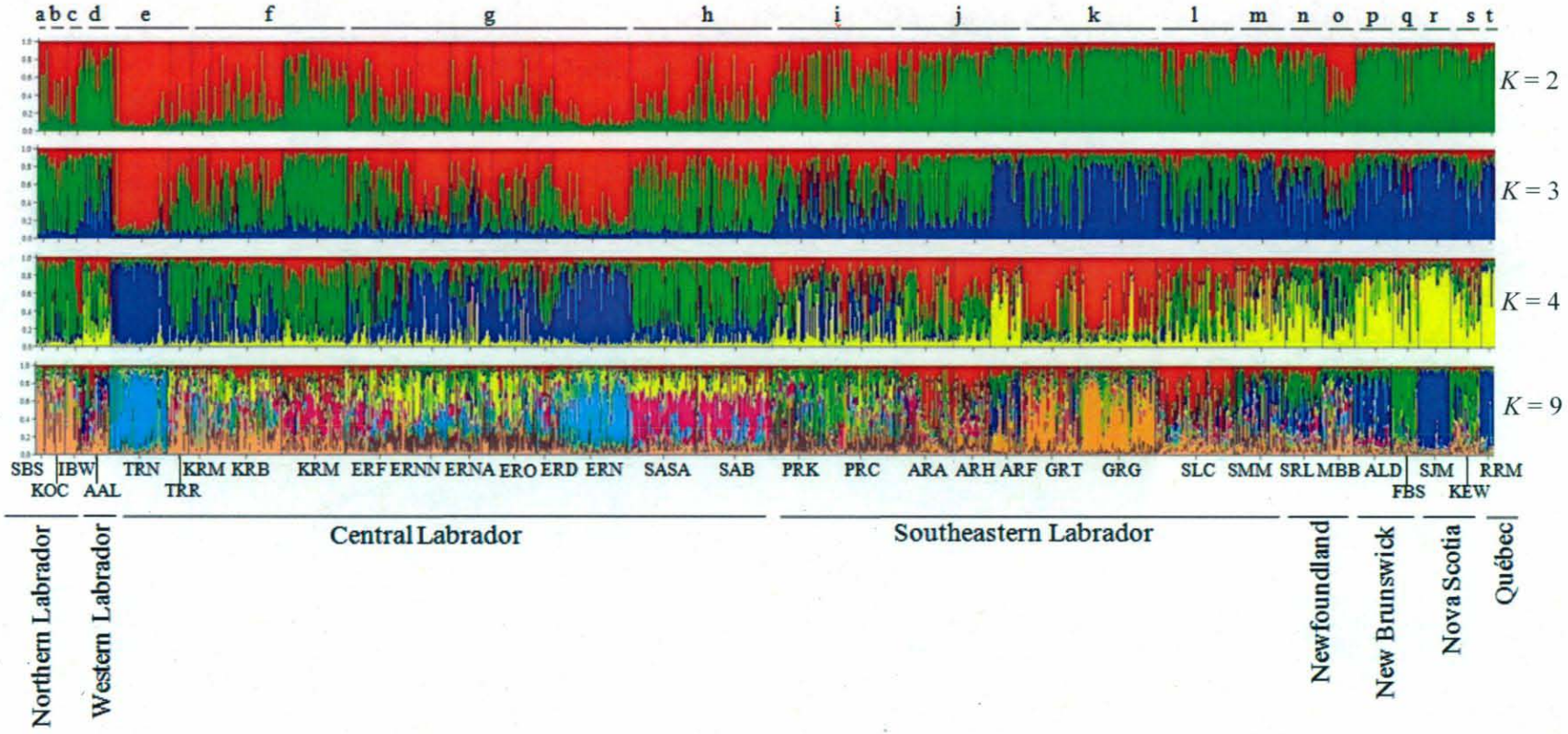


Figure 1.5. Bar plot representing population differentiation according to 2, 3, 4 and 9 genetic groups (*K*) (top to bottom). Each bar represents an individual fish. Population codes are given in Table 1.2. Lowercase letters correspond to watersheds as follows: a. Saputit Brook; b. Kogaluk River; c. Iladlivik Brook; d. Atikonak Lake; e. Traverspine River; f. Kenamu River; g. Eagle River; h. St. Augustine River; i. Paradise River; j. Alexis River; k. Gilbert River; l. St. Lewis River; m. St. Mary's River; n. Salmonier River; o. Middle Brook; p. Alder Brook; q. Farnham Brook; r. Saint John River; s. Kennebecasis River; t. Rupert River.

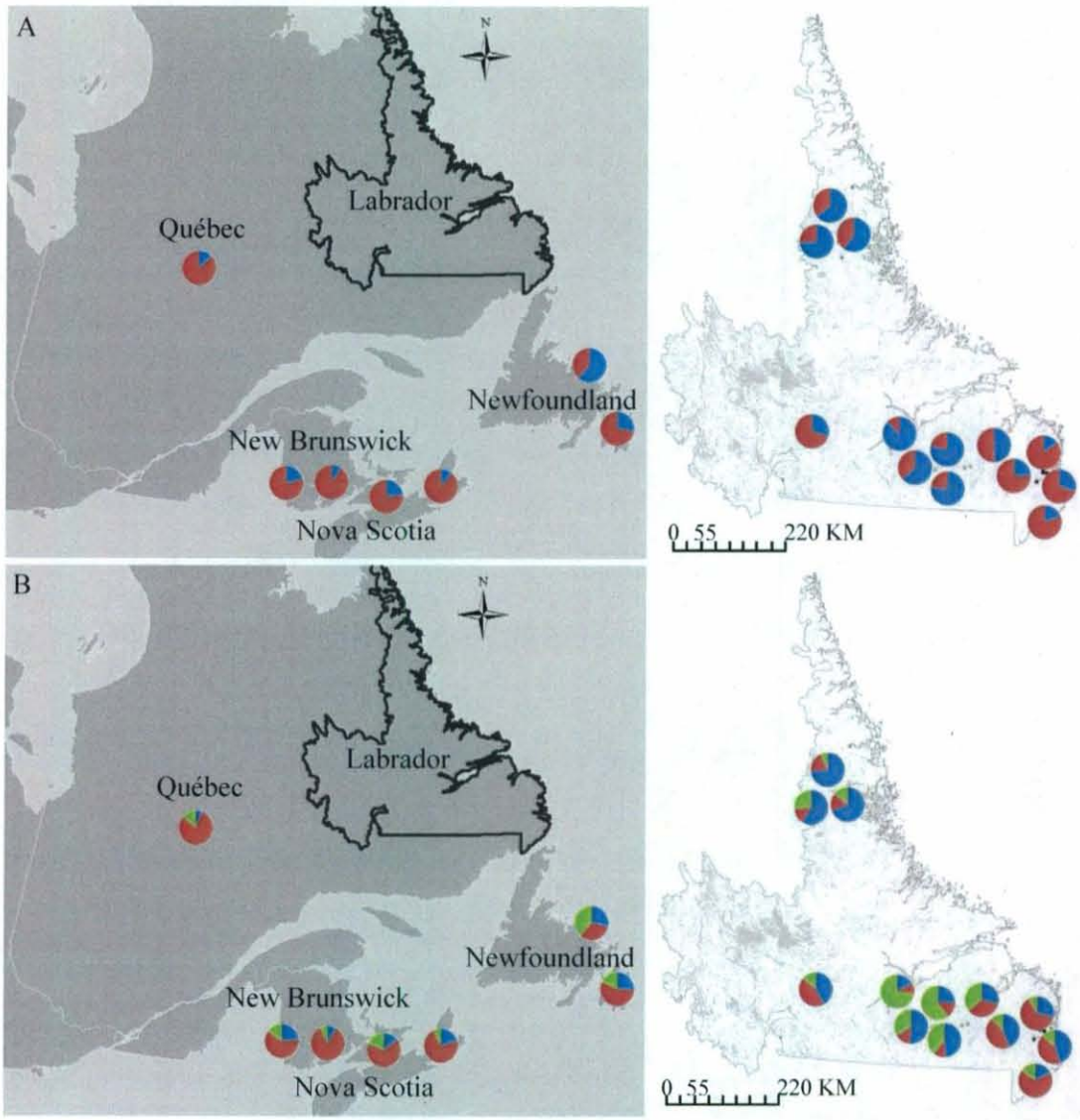


Figure 1.6. Geographical distribution of genetic clusters inferred from the STRUCTURE analysis based on $K = 2$ (A) and 3 (B) for each sampled watershed.

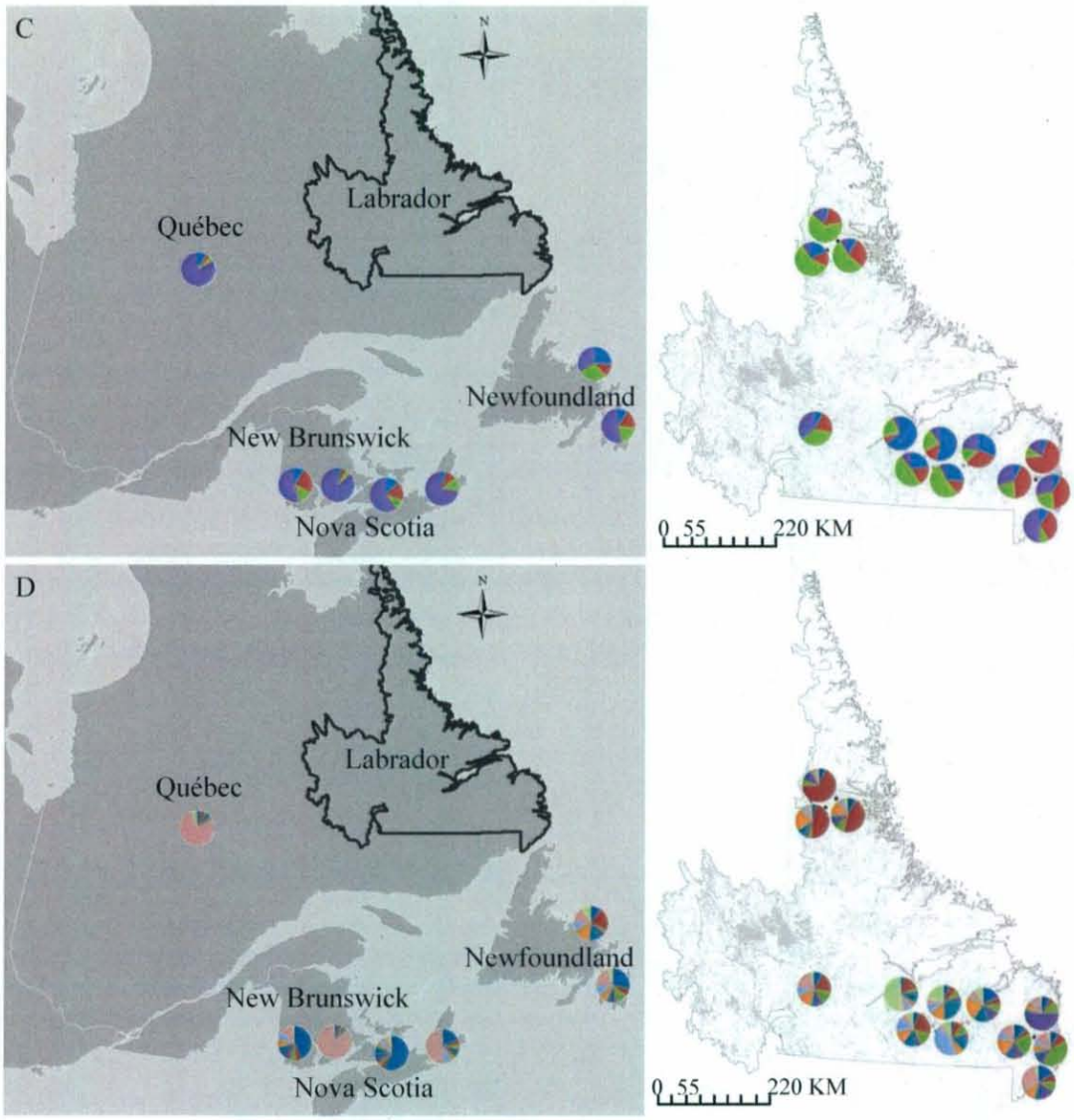


Figure 1.6 (continued). Geographical distribution of genetic clusters inferred from the STRUCTURE analysis based on $K = 2$ 4 (C) and 9 (D) for each sampled watershed.

Genetic structure among populations is apparent at all levels of cluster assignment.$K = 2$ is a biologically relevant number of clusters putatively corresponding to two post-glacial colonization routes and is consistent with the division into two genetic clusters also observed in the multi-dimensional scaling plot (Figure 1.2). STRUCTURE analyses of $K = 2$, 3, and 4 were utilized in an exploratory fashion to investigate the deeper genetic structure associated with post-glacial founding lineages and subsequent spatial genetic differentiation. When $K = 2$ (Figures 1.5A and 1.6A) northern and central Labrador are genetically distinct from the rest of the regions. Finer-scale assignment of individuals into three clusters (Figures 1.5B and 1.6B) separates northern Labrador from central Labrador, and further differentiates central Labrador. Also at $K = 3$, western and southeastern Labrador become differentiated from Québec, New Brunswick, Nova Scotia and Newfoundland. Forcing individuals into four genetic groups (Figures 1.5C and 1.6C) reveals further differentiation between western and southeastern Labrador, as well as within central Labrador (the cluster composition of No Boat Pond in Traverspine River watershed and Eagle River watershed are dominated by a single cluster, different from the rest of the region). Also, western Labrador and coastal southeastern Labrador populations become more like those in Newfoundland, Nova Scotia, New Brunswick and Québec.

The highest ln Pr($X|K$) at $K = 9$ is also biologically relevant in that it corresponds roughly to the number of regions examined (ln Pr ($X|K = 8$) is only slightly lower than ln Pr ($X|K = 9$) and is central in the plateau region) suggesting some level of genetic differentiation among regions. This is apparent in the bar plot of individual estimated cluster membership coefficients presented in Figure 1.5 and the associated mapped pie

chart (Figure 1.6D). Within northern Labrador, the same cluster assignment dominates each lake/watershed. Each genetic cluster contributes to the western Labrador lake/watershed in relatively equal proportions; a similar pattern is observed in the Newfoundland watersheds. A pattern of differentiation is evident in central Labrador such that all watersheds have their own unique cluster assignments. Interestingly, No Boat Pond in Traverspine River watershed and Nippard's Lake in Eagle River have very similar cluster assignments, each completely dominated by the same cluster. A similar pattern of watershed differentiation is persistent throughout southeastern Labrador; however here the Gilbert River watershed is distinct from the rest of the region in that it is dominated by a single genetic cluster not represented in any other region. The pattern of differentiation across Nova Scotia, New Brunswick and Québec is interesting. Alder Brook in Nova Scotia is more similar to Saint John River in New Brunswick and to Québec than it is to Farnham Brook in Nova Scotia. Farnham Brook is more similar to Kennebecasis River in New Brunswick. These patterns are interpreted as being influenced mainly by relatively contemporary processes, in particular limited ongoing gene flow among watersheds and possible exchanges among watersheds due to stocking.

## DISCUSSION

Population genetic structure of brook trout in Labrador and the broader region of northeastern North America, including insular Newfoundland, Nova Scotia, New Brunswick and Québec, was investigated by measuring variation at six microsatellite loci. I have collected the first genetic data from the nuclear genome on brook trout in Labrador, providing us with baseline data about current population structure needed to monitor the future impacts of the recent opening of the Trans-Labrador Highway, and enabling us to make inferences about patterns of post-Wisconsinan colonization of Labrador by brook trout. Levels and patterns of genetic variability were investigated within populations. Contemporary patterns of limited ongoing gene flow were revealed by pairwise $F_{ST}$ values, neighbour-joining analysis of Nei's genetic distance, AMOVA and STRUCTURE analyses. Finally, evidence for dual routes of post-Wisconsinan recolonization was indicated by the multi-dimensional scaling plot (MDSP) of pairwise $F_{ST}$ measures, and STRUCTURE results.

### *Patterns of intra-population genetic variability*

Genetic diversity found in brook trout populations in Labrador, Canada, (weighted average $H_E$ among regions = 0.622) was observed to be within the spectrum of variability found in other brook trout populations across their northeastern range (weighted average $H_E$ = 0.690 in the Indian Bay watershed, Newfoundland [Adams *et al.* 2003]; weighted

average $H_E$ = 0.380 in Gros Morne, Newfoundland [Poissant *et al.* 2005]; weighted average $H_E$ = 0.629 in Newfoundland in our study; weighted average $H_E$ = 0.662 in Nova Scotia in our study; weighted average $H_E$ = 0.778 in New Brunswick [Rogers *et al.* 2004]; weighted average $H_E$ = 0.643 in New Brunswick in our study; average $H_E$ = 0.49 in Québec [Angers & Bernatchez 1998]; $H_E$ = 0.659 in Québec in our study). This was somewhat unexpected as Labrador is located at the northernmost edge of the brook trout's native range, where levels of diversity would be expected to be reduced due to relatively recent bottlenecks and founder events. One possibility is that this discrepancy may be due to the use of different loci throughout the mentioned studies. Regardless, these moderate levels of genetic diversity are in contrast to the only other study to our knowledge that investigated genetic variation in brook trout populations in Labrador. In an assessment of the intraspecific population genetic structure of Atikonak Lake in western Labrador (S. Carr, unpublished report) only one mtDNA haplotype (401 bp of the cytochrome b gene) was found among 23 individuals, which was in agreement with a wide-scale survey of brook trout RFLP variation that found 98% of fish that colonized northeastern North America was of a single haplotype (Danzmann *et al.* 1998). Clearly, due to the relatively rapid evolution of microsatellites, these nuclear markers provide more variability for investigating population structure across this recently colonized area than mtDNA.

One-third of the lakes (11 of 33) have a significant heterozygote excess ($P <$ 0.01), of which the majority (6 of 11) were located in southeastern Labrador. Negative $F_{IS}$ is not observed as frequently as positive $F_{IS}$ and has therefore not been as fully

explored in the literature. One process that can result in a heterozygote excess is biased sampling, such that by chance the sample is not representative of the entire mixed population (Rasmussen 1979). Alternatively allelic frequencies in the population may differ between male and female parents (Pudovkin *et al.* 1996). Another possible reason for a significant heterozygote excess in southeastern Labrador is negative assortative mating (outbreeding). There is no particular reason to expect either of these phenomena to be more likely in southeastern Labrador than in any of our other sampling localities. Overdominance, such that there is selection either for heterozygotes or against deleterious homozygous individuals, can also drive excess heterozygosity in a population (Mitton 1989). To address this possibility $F_{IS}$ values were examined for each locus in each lake which revealed that deviations from Hardy Weinberg equilibrium in southeastern Labrador are mainly attributable to one locus (*Sfo*C86). Although this locus does not display a significant heterozygote excess in any of the individual lakes, $F_{IS}$ values for this locus are negative within individual lakes, and overall this locus does display a significantly negative $F_{IS}$ ($P < 0.05$). Strobeck (1979) reported that excess heterozygosity at specific loci can be caused by linkage to other loci influenced by heterozygote advantage. Alternatively, the possibility that scoring errors explain some of the observed excess of heterozygotes cannot be ignored. As a precaution, analyses were performed without this locus, but no difference in the interpretation of the data was found (results not shown).

When considering allelic richness, brook trout in eastern Canada have a characteristic 'southern richness, northern purity' pattern of variability wherein

populations in Nova Scotia have the highest levels of allelic richness ($A$ = 5.2) and those in northern Labrador have the lowest ($A$ = 3.8) and there is a general northern gradient throughout Labrador itself. When considering expected heterozygosity a similar pattern emerges with one notable exception. Unexpectedly the lowest level of heterozygosity was found in central ($H_E$ = 0.613) rather than northern Labrador ($H_E$ = 0.620). The reduced level of expected heterozygosity across central Labrador could be attributed to the increased elevation of the region as it is on a plateau which could have a number of effects. Higher elevations will be covered by glaciers for a longer period of time after the last glaciation, therefore populations are expected to settle there later, leaving less time to accumulate variation. A higher elevation may also have been more challenging to colonize, with a consequent lower number of founders and smaller effective population size. In addition, populations at higher altitudes are more likely to be isolated by waterfalls, with emigration downstream outweighing immigration upstream, thus reducing variation over time by genetic drift due to reduced population size and isolation. Reduced heterozygosity of populations at higher altitudes has been reported in a number of fish species (Trinidadian guppies, *Poecilia reticulate* (Shaw *et al.* 1991); eastern mosquitofish, *Gambusia holbrooki* (Hernandez-Martich & Smith 1990); brown trout, *Salmo trutta* (Hamilton *et al.* 1989)) including in brook trout in Québec (Angers *et al.* 1999).

*Contemporary patterns reveal limited ongoing gene flow*

Pairwise $F_{ST}$ measures in Labrador revealed levels of differentiation (mean $F_{ST}$ = 0.131; data not shown) that are similar to those of other brook trout populations, including those in Newfoundland (mean $F_{ST}$ = 0.148), Nova Scotia (mean $F_{ST}$ =0.099), New Brunswick (mean $F_{ST}$ = 0.101), Maine ($F_{ST}$ = 0.216; Castric *et al.* 2001) and Québec ($F_{ST}$ = 0.373; Angers & Bernatchez 1998). On a broader geographical level, pairwise $F_{ST}$ values through eastern Canada (mean $F_{ST}$ = 0.149; data not shown) are slightly larger than levels of differentiation along the coast of Maine, New Brunswick, Nova Scotia and Québec ($F_{ST}$ = 0.107; Castric *et al.* 2003). This could be a result of the broader sampling range considered in our study in addition to the Castric *et al.* (2003) study sampling primarily coastal populations, which may exchange migrants as part of an anadromous life history, thus resulting in less genetic differentiation between populations.

The level of divergence of brook trout in Labrador suggests there are limited levels of ongoing gene flow among populations of fish. The hierarchical AMOVA identifies watershed as the limiting feature, explaining the highest amount of variation and differentiation among all comparisons performed (8.38%, $F_{CT}$ = 0.084). This pattern is also evident in the STRUCTURE analysis. Assignment of individuals into $K$ = 3 clusters separates northern Labrador from central Labrador, and further differentiates watersheds within central Labrador, exhibiting limited gene flow between northern and central Labrador, as well as limited gene flow between watersheds within central Labrador. The greater the numbers of genetic groups that individuals are forced into, the greater the differentiation between watersheds. Generally cluster assignments are

homogeneous throughout each watershed (e.g. St. Augustine River in central Labrador), which indicates some ongoing gene flow within watersheds. An exception to this pattern is No Boat Pond (Traverspine River); brook trout in this lake are more genetically similar to Nippard's Lake (Eagle River) than to other lakes in its own watershed. This could be a result of having a shared founder population either via transfer by anglers, or due to differences in watershed structure at the time of colonization. Although bootstrap support is low, the pattern of ongoing gene flow within watersheds is also supported by the neighbour-joining dendrogram which clusters lakes in the same watershed together.

One of the goals of this study was to collect baseline information about genetic variation of brook trout populations in Labrador needed to monitor the possible future impacts of the Trans-Labrador Highway. Our analyses found enough variation to detect structure; if required in the future, it is important to make conservation efforts to maintain this diversity so brook trout populations preserve their ability to persevere through environmental changes such as the recent opening of the Trans-Labrador highway. Ongoing gene flow was found within but not among watersheds based on the AMOVA and STRUCTURE, so it may be important to base any future management strategies at the scale of the watershed.

*Evidence for dual routes of post-Wisconsinan colonization*

Microsatellites have been successfully utilized to infer colonization patterns (Angers & Bernatchez 1998; Taylor & McPhail 2000; Lu *et al.* 2001; Koskinen *et al.*

2002), and can reveal patterns that may be undetectable using mtDNA or other phylogeographic markers (Flanders *et al.* 2009). Only one mtDNA haplotype was found in brook trout during a previous survey of genetic variation of freshwater fish in Labrador (S. Carr, unpublished report). Furthermore, Danzmann *et al.* (1998) found the entire northern range to be colonized by mainly a single haplotype. Here, inferences are made about possible colonization routes based on patterns of microsatellite variation.

Evidence for dual colonization routes is present in by both the multi-dimensional scaling plot of pairwise $F_{ST}$ values and the geographical pattern revealed by STRUCTURE at $K = 2$. The MDSP identifies two groups; one that includes lakes in central and northern Labrador, and another that includes lakes in all other regions. STRUCTURE analysis at $K = 2$ identifies the same two genetic groups; northern and central Labrador are grouped separately from all other regions. This deep level of genetic differentiation suggests that central and northern Labrador were colonized from a different refugial source than the rest of the northeastern range. One colonization route that has been suggested is that brook trout from the Atlantic and Mississippian refugia dispersed through Québec then into Labrador, while other anadromous fish from the Atlantic refugium invaded coastal regions (Black *et al.* 1986). Alternatively, Danzmann *et al.* (1998) suggested that brook trout in the northeastern parts of their native range may have origins from a more north-easterly Acadian refugium, in combination with contribution from the Atlantic and/or Mississippian refugia.

The results provide more support for the latter hypothesis; northern and central Labrador may have been founded by some combination of fish from the Atlantic and/or

Mississippian refugia, while more southern populations were founded by those from an Acadian refugium. The populations from Québec and western Labrador are an exception to this pattern, as they would be expected to cluster with northern and central Labrador populations. However these samples were collected separately by different personnel from the remainder of the lakes in this study. In general, some caution should be taken based on the sampling limitations as west-Labrador and Québec are represented by only one population (in addition, the Atlanic provinces, insular Newfoundland, and northern Labrador are represented by only two to three populations). The Québec sample in particular is unusual with a much higher heterozygosity than reported by Angers and Bernatchez (1998) for the same lake. These samples were obtained from the Bernatchez lab with the intention of standardizing allele sizes; therefore it is possible that the Québec brook trout used in this study did not represent a random sample of the population. Samples may have been chosen with the intention of representing a wide range of alleles, thus artificially increasing heterozygosity. The pattern of heterozygosity observed is not consistent with concurrent colonization of central Labrador and northern Labrador from the Atlantic and/or Mississippian refugia, or with colonization of central prior to northern Labrador. However, reduced heterozygosity in central Labrador may simply be due to genetic drift subsequent to founding due to its location on a plateau, as discussed previously. This pattern is also evident from the STRUCTURE analysis at $K = 3$, where central Labrador has a different cluster assignment than northern Labrador. A third hypothesis is that northern and central Labrador lakes were colonized from a northern coastal refugium/nunatak located in the Torngat Mountains in northern Labrador (Pielou 1991). Under this scenario central Labrador lakes would have been colonized after

northern ones, accounting for their reduced diversity. The likelihood of this northern refugium sustaining a brook trout population is unknown.

A major finding here is that microsatellite variation is sufficient to detect the possibility of two glacial refugia, whereas the only previous study on genetic variation in brook trout revealed no variation at all in the mtDNA cytochrome *b* gene. To verify that the patterns of microsatellite variation may be more influenced by contemporary processes, my second chapter focuses on documenting the genetic variation of a larger portion of the mitochondrial genome.

**Unusual pattern of nucleotide variation in the mitochondrial genome of brook trout in eastern Canada: evidence for genetic recombination and dual routes of post-glacial recolonization**

## ABSTRACT

Levels and patterns of mitochondrial DNA (mtDNA) variation were characterized in order to investigate the population structure and possible routes of post-glacial recolonization of brook trout (*Salvelinus fontinalis*) through eastern Canada, specifically in Labrador. Sequence diversity of a 1960-bp portion of the mitochondrial genome (the entire NADH dehydrogenase 1 gene and a portion of the cytochrome oxidase 1 gene) of 126 brook trout from 32 lakes through eight regions in eastern Canada was analysed. Overall patterns of intraspecific haplotype and nucleotide diversity revealed low levels of mtDNA diversity of brook trout in eastern Canada, a characteristic trait of populations at the northern end of their range. Upon comparison of eastern brook trout populations to the only available brook trout whole mitochondrial genome in GenBank (AF154850.1), it was revealed that the reference brook trout represents a fish with mtDNA that has recombined with Arctic charr mtDNA. A single haplotype dominated the eastern Canadian region (composing 61.1% of populations), and the timing of evolution of the majority of clades coincide with the retreat of glacier ice during the Wisconsinan glaciation. Pairwise $F_{ST}$ values and the AMOVA results revealed very little to no significant differentiation among populations, however, the haplotype distribution was

able to reveal evidence of two different post-glacial founding groups contributing to brook trout populations. Overall, our results exemplify the strong influence that historical processes, such as glaciations, have on shaping modern population genetic structure of brook trout.

## INTRODUCTION

Brook trout (*Salvelinus fontinalis* Mitchill 1814) are ubiquitous across their native range of eastern North America (Power 1980), existing as far north as Labrador, the mainland component of the province of Newfoundland and Labrador, Canada, and as far south as Georgia, USA. Populations have successfully adapted to a broad range of habitats, from small creeks to large rivers and lakes of various depths (Power 1980), where fish may exhibit anadromous or resident freshwater life history forms. These salmonids are intriguing not only in terms of their ecology and life history, but as a northern species whose population structure and distribution have been significantly affected by historical events, such as the glaciations, are also interesting from an evolutionary perspective.

The last glacial maximum, the Wisconsinan, which occurred ~ 18 000 years ago (Dyke & Prest 1987), had a profound effect on the present-day distribution of genetic variation across species ranges, especially those populations at the species' range extremes. As glacial ice advanced, northern species were restricted to southern unglaciated refugia. Relative to present-day northern populations, populations that

persisted in isolation in southern refugia have had more time to accumulate variation (Hewitt 1996, 1999; Bennett 1997) while simultaneously diverging from populations in other refugia (Hewitt 1996; Hewitt 2004). The signature of this is evident in the distribution and genetic structure of populations today. Northern regions typically have fewer species (Pielou 1991), with lower genetic diversity, a pattern that is referred to by Hewitt (1996, 1999) as 'southern richness, northern purity'. This is exemplified by freshwater fish; in northern, formerly glaciated regions, the number of species is relatively low and populations within species, including brook trout, possess less genetic variation than in southern locations (Danzmann *et al.* 1998).

Patterns of mitochondrial DNA (mtDNA) variation in some brook trout populations are remarkable in that fish with completely introgressed Arctic charr (*Salvelinus alpinus*) mtDNA have been identified at the northern end of their range. Originally this was documented by Bernatchez *et al.* (1995) in a brook trout population in Lake Alain, eastern Québec. Further instances of introgression of Arctic charr mtDNA were later identified in populations of brook trout throughout the Rocheuse River subdrainage of the Portneuf basin (Glémet *et al.* 1998). Since introgression in brook trout has been restricted to drainages where Arctic charr are absent, it was deduced that the timing of introgression was historical, and possibly due to differences in founding population sizes of brook trout and Arctic char. Across the region of introgression, brook trout have a higher level of genetic diversity than Arctic charr (Anderson *et al.* 1983; Kornfield *et al.* 1981; McGlade 1981), indicative of a larger long-term effective population size. The likelihood of hybridization is higher when one parental population is

substantially smaller (Hubbs 1955; Avise *et al.* 1988), in which case the smaller population (in this case Arctic charr) may attempt to out-cross as the adequate number of mates is not available. Additionally, freshwater fish have an increased chance of introgression compared to other vertebrates attributable to competition for suitable habitat for spawning, external fertilization, and the tendency for species to have overlapping ranges (Hubbs 1995). Since the introgression event was likely historical, yet current brook trout populations still possess Arctic charr mtDNA, it is thought that the arctic char mtDNA genome may confer a selective advantage to these brook trout. The respiratory enzyme complexes of these fish are encoded by brook trout nuclear DNA and Arctic charr mtDNA – mtDNA which has evolved in a cold environment hence possibly providing an advantage for brook trout found at the northern end of their range where conditions are much cooler (Glémet *et al.* 1998). This event is a powerful illustration of how historical processes exert a significant influence on shaping modern genetic structure.

Mitochondrial DNA restriction fragment length polymorphism (RFLP) variation has been used to investigate widespread patterns of re-colonization of brook trout throughout their native range (Danzmann *et al.* 1998). Following retreat of glacial ice from eastern North America after the Wisconsinan glacial maximum, watersheds are hypothesized to have been re-colonized by fish from three refugia: the Mississippian, Atlantic and Acadian. Fish with Mississippian origins dominate the southern Great Lakes region, while fish from the Atlantic refugium appear to have re-colonized the Lake Ontario watershed, Atlantic coastal region, and also had some contribution to Ontario,

Québec and eastern Canada. Throughout eastern Canada a single haplotype, haplotype '1', is predominant, and is postulated to have been present in both the Mississippian and Atlantic refugia. However, there is also evidence in this region for additional contribution from an Acadian refugium based on the presence of 10 private haplotypes combined with the lack of haplotype '2' fish. Despite the comprehensive examination of brook trout mtDNA variation across much of the native range, the phylogeography of brook trout in Labrador, Canada, the most northern part of the range, remains unexplored.

One possible scenario for the colonization of Labrador by brook trout is that fish from the Atlantic and Mississippian refugia dispersed through Québec then into Labrador, while other fish from the Atlantic refugium invaded coastal regions (Black *et al.* 1986). Support for dual routes of colonization of Labrador by brook trout was found when microsatellite population genetic structure of this species was examined across eastern Canada (Pilgrim *et al.* 2010, in review). This was evidenced by two distinct genetic clusters represented disproportionately in north-central Labrador versus west and southeastern Labrador, the island of Newfoundland, and the remainder of the Atlantic provinces. These two clusters could correspond to the two routes of colonization hypothesized by Black *et al.* (1986). However, the possibility of a northern coastal refugium/nunatak raised by Pielou (1991) is an alternate explanation for the north-central cluster.

Microsatellites have been successfully employed to elucidate historical factors, such as colonization patterns (Angers & Bernatchez 1998; Koskinen *et al.* 2002; Lu *et al.* 2001), but microsatellite population structure is also strongly influenced by contemporary

processes, such as ongoing gene flow. In order to discriminate between historical and contemporary features of brook trout population structure in Labrador and neighbouring eastern Canada, 1960 bp sequences from two mtDNA genes among 126 fish from eight regions in eastern Canada were examined in the present study. In addition to phylogeographic analyses aimed at elucidation of the pattern of post-glacial recolonization, patterns of intraspecific haplotype and nucleotide diversity within and among these regions are also described.

## METHODS AND MATERIALS

### *Sample collection*

Brook trout samples were collected by employees from the Wildlife Division of the Department of Environment and Conservation via gill netting. Sampling took place in the summer months during the years 2003 – 2008. Morphometric data, maturity, sex and stomach contents for each sample were recorded. Fin clippings were also collected and stored in envelopes at -20°C. Six regions across eastern Canada were sampled: northern Labrador, central Labrador, southeastern Labrador, insular Newfoundland, Nova Scotia and New Brunswick, for a total of 35 lakes sampled within 22 watersheds (Table 2.1; Figure 2.1). Representative Québec and western Labrador samples were provided by Dr. Louis Bernatchez (Université Laval) and Dr. Steven M. Carr (Memorial University of Newfoundland and Labrador), respectively.

Table 2.1. Locations sampled for brook trout populations across eastern Canada with sample size from each lake (*N*), as well as latitude and longitude of each lake.

| Region | Watershed | Lake | Population Code | N | Latitude | Longitude |
|---|---|---|---|---|---|---|
| Northern Labrador | | | | | | |
| | Saputit Brook | Saputit Lake | SBS | 2 | 57° 27' 39 | 62° 36' 26 |
| | Kogaluk River | Cabot Lake | KOC | 2 | 56° 25' 4 | 63° 38' 21 |
| | Iladlivik Brook | Walkabout Lake | IBW | 2 | 56° 19' 39 | 63° 9' 34 |
| | Anaktalik Lake | Anaktalik Brook | ABA | 1 | 56° 29' 50 | 62° 51' 38 |
| | Konrad Brook | Konrad Lake | KLK | 2 | 56° 13' 22 | 62° 43' 24 |
| | No Name | Shallow Lake | NNS | 2 | 57° 41' 16 | 63° 21' 56 |
| | No Name | Mistake Lake | NNM | 1 | 56 25' 1 | 63° 38' 21 |
| Western Labrador | | | | | | |
| | Atikonak River | Atikonak Lake | AAL | 17 | 52° 40' 0 | 64° 34' 60 |
| Central Labrador | | | | | | |
| | Traverspine River | No Boat Pond | TRN | 3 | 53° 8' 5 | 60° 38' 5 |
| | Traverspine River | The Right Lake | TRR | 2 | 53° 0' 8 | 60° 45' 54 |
| | Kenamu River | Mercier Lake | KRM | 1 | 52° 55' 5 | 60° 43' 25 |
| | Kenamu River | Brennan Lake | KRB | 2 | 52° 57' 22 | 60° 15' 25 |
| | Kenamu River | Nikki's Pond | KRN | 3 | 52° 36' 18 | 60° 25' 55 |
| | Eagle River | Fred's Lake | ERF | 2 | 52° 52' 1 | 59° 43' 39 |
| | Eagle River | No Name Lake | ERNN | 3 | 52° 40' 58 | 65° 26' 35 |
| | Eagle River | NAP Pond | ERNA | 3 | 52° 35' 6 | 59° 4' 44 |
| | Eagle River | Osprey Lake | ERO | 3 | 52° 44' 39 | 58° 35' 1 |
| | Eagle River | Dead Dog Pond | ERD | 2 | | |
| | Eagle River | Nippard's Lake | ERN | 3 | 53° 5' 10 | 58° 50' 29 |
| | St. Augustine River | St. Augustine | SASA | 4 | 52° 35' 15 | 59° 18' 21 |
| | St. Augustine River | Bog Lake | SAB | 3 | 52° 31' 45 | 59° 4' 3 |
| Southeast Labrador | | | | | | |
| | Paradise River | Keith's Lake | PRK | 3 | 52° 59' 23 | 57° 49' 15 |
| | Paradise River | Crooked Lake | PRC | 3 | 53° 20' 28 | 57° 34' 44 |
| | Alexis River | Alexis Pond | ARA | 2 | 52° 31' 59 | 57° 3' 16 |
| | Alexis River | Handkerchief Pond | ARH | 2 | 52° 31' 4 | 57° 1' 34 |
| | Alexis River | Feeder Pond | ARF | 3 | 52° 32' 10 | 56° 29' 36 |

Table 2.1 continued

| | | | | | | |
|---|---|---|---|---|---|---|
| Southeast Labrador | | | | | | |
| | Gilbert River | Tilt Pond | GRT | 1 | 52° 42' 17 | 56° 18' 39 |
| | Gilbert River | Gilbert Lake | GRG | 2 | 52° 41' 14 | 56° 11' 57 |
| | St. Lewis River | Curl's Pond | SLC | 2 | 52° 24' 50 | 56° 0' 52 |
| | St. Mary's River | Mary's Harbour Big Pond | SMM | 3 | 52° 18' 53 | 56° 1' 24 |
| Newfoundland | | | | | | |
| | Salmonier River | Little Gull Pond | SRL | 18 | 47° 15' 34 | 53° 19' 33 |
| | Middle Brook | Butt's Pond | MBB | 15 | 48° 49' 22 | 54° 16' 10 |
| Nova Scotia | | | | | | |
| | Alder Brook | River Denys | ALD | 1 | 45° 58' 12 | 64° 35' 10 |
| | Farnham Brook | Salmon River | FBS | 1 | 45° 22' 55 | 63° 16' 14 |
| New Brunswick | | | | | | |
| | Saint John River | Moose Lake | SJM | 1 | 46° 51' 54 | 66° 47' 16 |
| | Kennebecasis River | Walton Lake | KEW | 2 | 45° 36' 39 | 65° 19' 16 |
| Québec | | | | | | |
| | Rupert River | Mistassini Lake | RRM | 4 | 50° 59' 49 | 73° 38' 21 |
| Overall | | | | 126 | | |

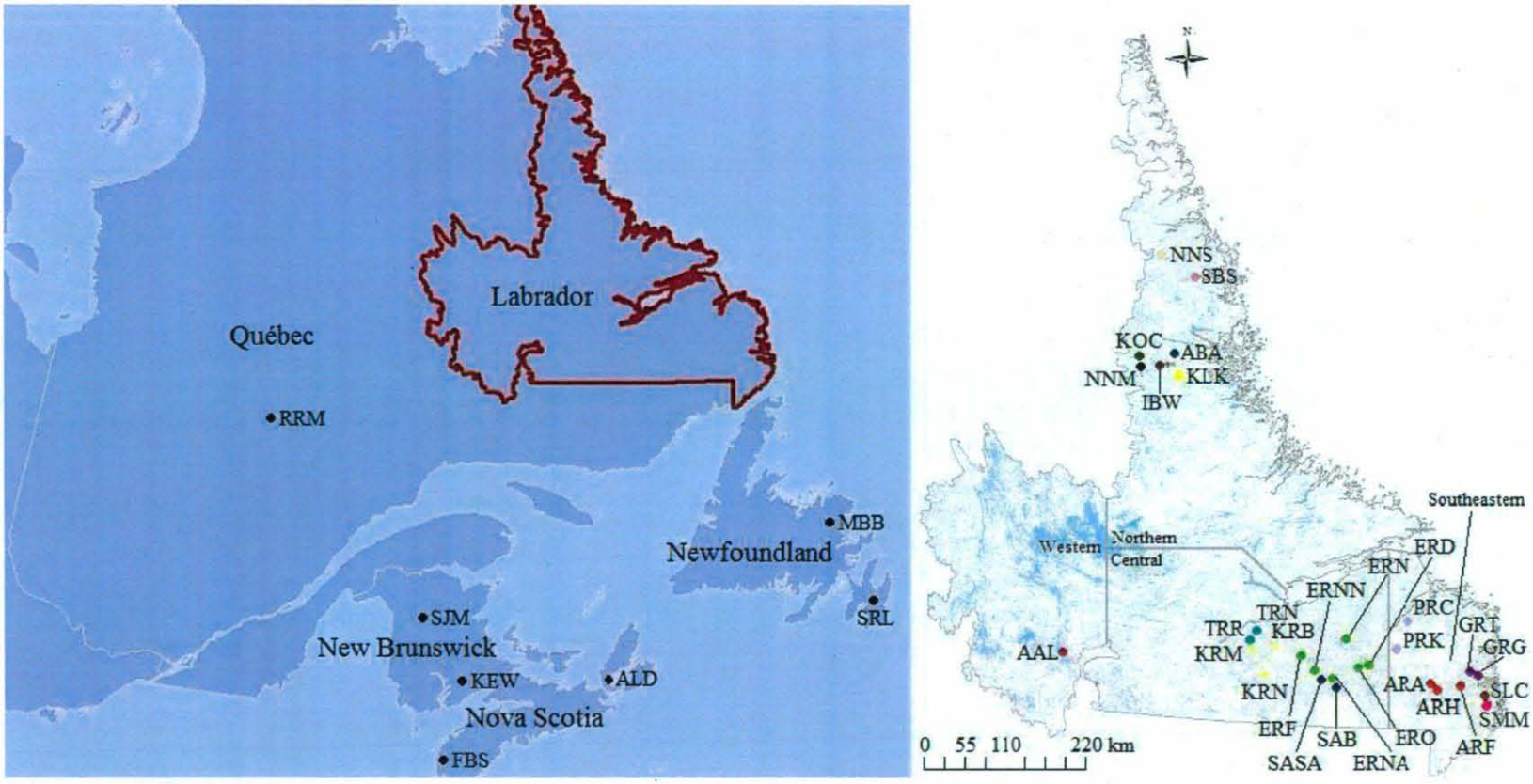


Figure 2.1. Lakes sampled for brook trout across eastern Canada. Population codes are given in Table 2.1.

*DNA extraction, PCR amplification and sequencing*

Genomic DNA was extracted from (~4 $mm^2$ of caudal fin tissue) using the Qiagen QIAamp DNA Mini Kit (Qiagen Inc., Missisauga, Canada) following the tissue protocol. Two gene regions (NADH dehydrogenase 1 [ND1] and cytochrome oxidase 1 [CO1]) from the mitochondrial genome were amplified using the three sets of primers described in Table 2.2 (all primers were tailed by attaching one of the following M13 sequences to the 5' end: forward primers – 5'-CAGGAAACAGCTATGAC-3'; reverse primers – 5'-GTAAAACGACGGCCAGT-3'). 25 µL PCR reactions containing 2.5 µL 10X buffer (Qiagen Inc.), 0.5 µL 10 mM dNTPs (New England Biolabs, Ontario, Canada), 1µL of each of the 10 µM forward and 10 µM reverse primer (Operon, Alabama, USA), 0.2 µL of 5U/µL HotStar *Taq* Plus DNA Polymerase (Qiagen Inc., Valencia, USA) and 2 µL of DNA template (2 – 570 ng) were subjected to the following thermal profile on an Applied Biosystems GeneAmp PCR System 9700 Thermal Cycler (Foster City, CA): 95℃ for 5 minutes followed by 40 cycles of 93℃ for 30 seconds, 55℃ for 30 seconds and 72℃ for 2 minutes, ending with a final extension of 72℃ for 5 minutes. Successfully amplified PCR products were purified using a multi-well filter plate (Pall Life Sciences, New York, USA) according to manufactuer's instructions. Sequencing reactions were carried out in 20 µL reactions composed of 0.5 µL BigDye® Terminator v3.1 Cycle Sequencing Ready Reaction – 100 mix (Applied Biosystems Inc., CA, USA), 2 µL of 1.6 pmol/µL forward and reverse M13 primer, 5 µL 5X sequencing buffer and 2 µL of DNA template. The following PCR profile was utilized: 96°C for 6 minutes, 25 cycles of 96°C for 10 seconds followed by 50°C for 5 seconds, and 60°C for 4 minutes. Sequencing reaction products

Table 2.2. Sequence of each forward and reverse primer utilized. Note that the M13 sequences 5'-CAGGAAACAGCTATGAC-3' and 5'-GTAAAACGACGGCCAGT-3' were added to the 5' end of the forward and reverse primers, respectively.

| Gene | Forward Primer | Reverse Primer | # of bp |
|---|---|---|---|
| ND1 | 5'-AAAGTGGCAGAGCCCGGTAATTG-3' | 5'-AATGGTCCTCCAGCGTATTCTAC-3' | 1305 |
| | 5'-CGTAGCTCAAGAAAGCATCTGAC-3' | 5'-TATTTGGGGTATGGGCCCGAAAG-3' | |
| CO1 | 5'-TCAACCAACCACAAAGACATTGGCAC-3' | 5'-TAGACCTCTGGGTGGCCAAAGAATCA-3' | 655 |

were purified by ethanol precipitation such that products were washed with a series of decreasing concentrations of ethanol (95% followed by 70%), centrifuged between each wash to precipitate DNA, then finally resuspended in 10 uL of Hi-Di Formamide (Applied Biosystems Inc., CA, USA). Samples were electrophoresed in an Applied Biosystems 3730 DNA Analyzer using Sequencing Analysis v. 5.2 Software (Applied Biosystems Inc., CA, USA) at the Genomics and Proteomics Facility, CREAIT Network, Memorial University of Newfoundland.

*Data analysis*

Sequences were edited and aligned by comparison to the homologous ND1 and CO1 genes from the *Salvelinus fontinalis* complete mitochondrial genome available from GenBank (AF154850) using Sequencher v4.8 (Gene Codes, Ann Arbour, MI), and contiguous sequences for each individual were assembled. Numbers of variable sites, transitions, transversions, synonymous and nonsynonymous substitutions, and average pairwise sequence diversity were determined for each gene and the total sequence using MEGA 4.0 (Tamura *et al.*, 2007).

Individual haplotypes were identified using DnaSP v.5 (Librado and Rozas, 2009) in combination with visual inspection to account for gap polymorphisms. The program Sneato v.2.0 (http://www.xmission.com/ ~wooding/Sneato/index.html) was used to construct a haplotype network, and the minimum spanning network option under AMOVA in Arlequin 3.11 (Excoffier *et al.*, 2005) was used to find alternate connections.

A neighbour-joining phylogeny of haplotypes was constructed from a maximum composite likelihood distance matrix using pairwise deletion, and tested by 1000 bootstraps, in MEGA 4.0. Haplotype 13 was excluded from the phylogeny due to its high level of missing data.

Dates were inferred at the nodes of major clusters using the divergence time and linearized tree options in MEGA 4.0. We employed an evolutionary rate of 3.018% per million years calculated from the sequence divergence between brook trout from this study and Arctic charr reference sequence AF154851 (overall mean distance = 0.06036) and the time of divergence (~ 1 million years) between brook trout and Arctic charr reported by Brunner *et al.* (2001).

A hierarchical Analysis of Molecular Variance (AMOVA) was carried out in ARLEQUIN v.3.11 to investigate the partitioning of genetic variation in Newfoundland and Labrador. Brook trout were grouped in three ways: i) according to watershed, then into regions; ii) according to lake, then into regions; iii) according to lake, then into watershed (for watersheds that included more than one sampled lake). Estimates of the variance components were quantified using conventional *F*-statistics over 1000 bootstrap replicates. ARLEQUIN was also utilized to calculate pairwise $F_{ST}$ values among watersheds in Newfoundland and Labrador based on conventional *F*-statistics ($F_{ST}$). The pairwise $F_{ST}$ matrix was utilized to construct a Neighbour-Joining dendrogram in MEGA 4.0. Differentiation between groups revealed by the dendrogram was investigated in ARLEQUIN by measuring the pairwise $F_{ST}$ between distinct lineages.

In order to investigate the possibility of demographic expansion of brook trout populations into eastern Canada, DnaSP v.5.0 was utilized to calculate Tajima's (1989) *D*, a measure indicative of directional selection, and the pairwise mismatch distribution of all haplotypes, as well as the most predominant haplotype and its derivatives.

## RESULTS

### *Low mtDNA diversity of brook trout in eastern Canada*

Contiguous 1960 bp sequences containing a 49 bp segment of the tRNA-Leu 1, the complete NADH dehydrogenase 1 gene, the complete tRNA-Ile and tRNA-Gln, an 18 bp segment of the tRNA-Met, and a 655 bp segment of the cytochrome oxidase 1 gene were assembled for 126 brook trout. A total of 22 variable sites defining 13 haplotypes were identified (Figure 2.2); one is located in the tRNA-Leu 1, 10 are localised to the NADH dehydrogenase 1 gene, one is located in the tRNA-Ile, and the remaining 10 were found in the cytochrome oxidase 1 gene (Table 2.3). Of the 22 substitutions, 18 are transitions while four are transversions. All protein-coding gene substitutions are synonymous with the exception of two, both found in the ND1 gene (Val ↔ Ile at position 350 and Val ↔ Met at position 988) (Table 2.3). Average pairwise sequence diversity among all individuals was 0.00087 in the ND1 gene and 0.00080 in the CO1 gene (Table 2.3). Overall pairwise sequence diversity measured 0.00094.

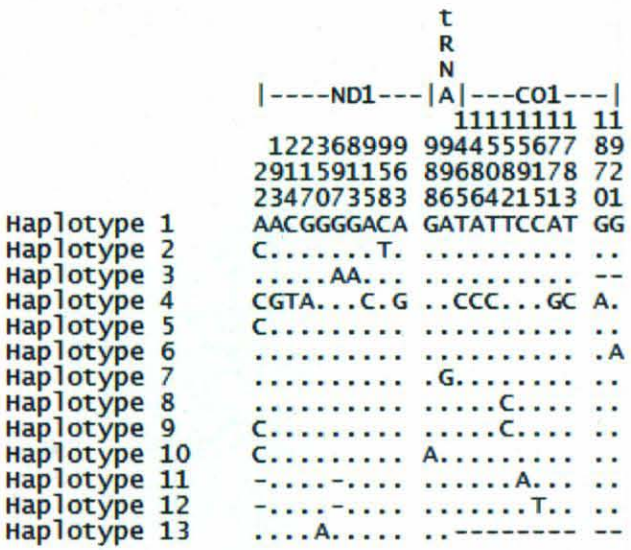


Figure 2.2. Variable nucleotide positions with their corresponding haplotypes among 126 brook trout sequenced. For haplotype 1, each position is given. As compared to haplotype 1, periods indicate the same nucleotide at that position, and dashes indicate missing data (due to unsuccessful amplification of the CO1 gene).

Table 2.3. Number of variables sites, transitions, transversions, synonymous and nonsynonymous substitutions, and average pairwise diversity obtained within each protein-coding region characterized, and over all brook trout sampled in eastern Canada.

| Gene | No. of variable sites | Transitions | Transversions | Synonymous Substitutions | Nonsynonymous substitutions | Average pairwise diversity |
|---|---|---|---|---|---|---|
| ND1 | 12 | 10 | 2 | 9 | 2 | 0.00087 |
| CO1 | 10 | 8 | 2 | 10 | 0 | 0.0008 |
| Overall | 22 | 18 | 4 | 19 | 2 | 0.00094 |

*Unusual pattern of variability in the brook trout mtDNA genome*

While comparing our sequences to Genbank sequence AF154850, the whole mtDNA genome of a brook trout from a population in Québec (Doiron *et al.* 2002), we discovered an unusual region of high diversity. There were 11 third position base changes within a 220-bp region of the ND1 gene where all the fish we sequenced differed from the reference sequence (Figure 2.3). The majority (nine) of these positions were transitions, while two were transversions. Only a single transition was nonsynonymous at position 1084, resulting in the conversion of Val ↔ Ile (a conservative amino acid change). Interestingly, the 11 variable sites in the 220-bp region in the reference sequence perfectly match Arctic charr (GenBank AF154851), while the nucleotides at those positions in the fish in this study match the sequence of a brook trout (GenBank AF126000) from a naturalised British Columbian population (Taylor *et al.* 1999) (Figure 2.3).

*Relationships among haplotypes*

The haplotype network depicted in Figure 2.4 demonstrates that the majority of haplotypes (3, 5, 6, 7, 8, 11, 12, 13) are connected to haplotype 1, the most frequent haplotype, through only one or two bases changes. Haplotype 9 is connected to haplotype 1 by two changes through haplotype 8 while haplotype 10 is connected by two changes through haplotype 5. Haplotype 2, the second most frequent haplotype, is also connected to haplotype 1 by two base changes through haplotype 5. A notable exception to this

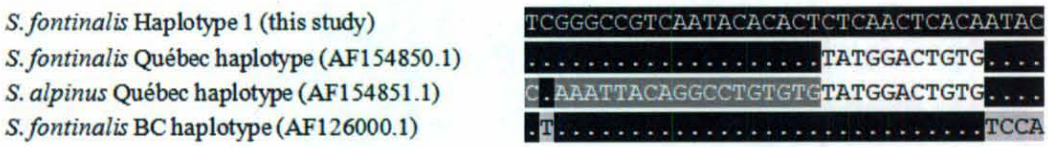


Figure 2.3. Alignment of a portion of the ND1 gene in two *Salvelinus* species, with conserved portions of sequence highlighted in the same shade. When considering the 220 bp region of high diversity between our haplotype 1 brook trout and the reference brook trout sequence from GenBank (AF154850), the sequence is conserved between the reference brook trout and Arctic charr.

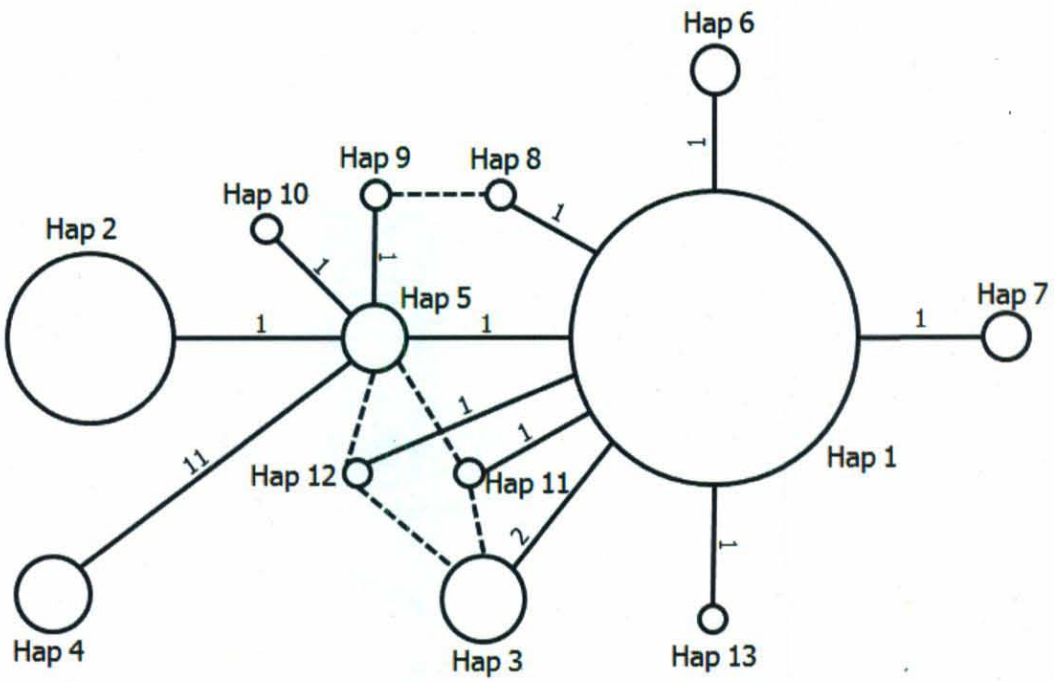


Figure 2.4. Haplotype network of the ND1 gene and partial CO1 gene of brook trout in eastern Canada. Haplotypes are listed in Figure 2.2. Nodes are sized according to relative frequency of the haplotype as listed in Table 2.4, and numbers on the connections indicate the number of mutational steps between haplotypes. Dashed lines indicate alternate connections.

pattern is haplotype 4; this haplotype differs from haplotype 1 by 12 bases; it is again connected through haplotype 5. Also notable is that only haplotype 1 of the frequent haplotypes has a number of derivative haplotypes; haplotype 2, the second most frequent haplotype, does not. Assuming that midpoint rooting identifies the most ancestral haplotype, the neighbour-joining dendrogram of haplotypes (Figure 2.5), indicates that haplotype 4 is ancestral, and that haplotypes 2, 5, and 10 are basal to a clade including haplotypes 1, 3, 6, 7, 8, 9, 11 and 12 which is hence the most recently-evolved clade. Based on the evolutionary rate of 3.018% per million years, the most ancestral haplotype, haplotype 4, evolved 92,977 years ago. The most basal clade (haplotypes 2, 5 and 10) arose 12,282 years before present, while the clade including haplotypes 1, 3, 6, 7, 11 and 12 dates back to 10,262 years ago. The most recently-evolved clade including haplotypes 8 and 9 dates to 8,501 years ago.

*Haplotype distribution in eastern Canadian brook trout regions*

The frequency of each of the 13 haplotypes found in this study in each of the eight regions investigated is given in Table 2.4; haplotypes are numbered according to frequency. There are seven shared haplotypes; the remaining six are singleton haplotypes found once each. The most common haplotype, haplotype 1, was found in 61.1% of brook trout and was distributed among all Newfoundland and Labrador regions, as well as Québec. Haplotype 2 was the next most common haplotype, representing 18.2% of fish, and was present in all Labrador regions except southeastern, and also in New Brunswick.

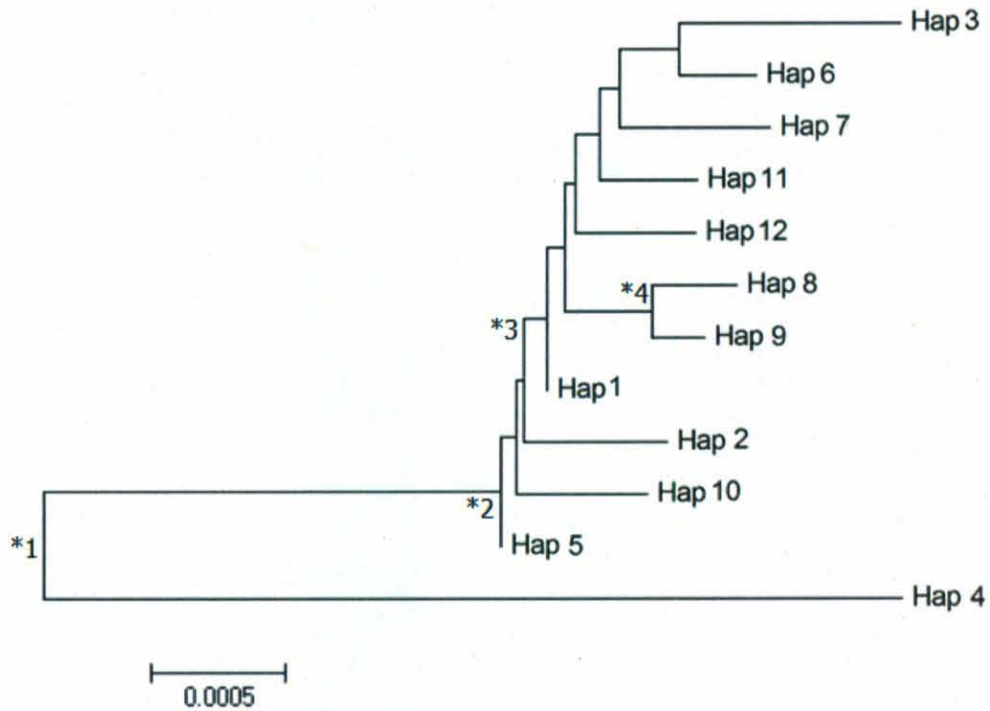


Figure 2.5. Neighbour-joining dendrogram relating twelve of the mtDNA haplotypes of brook trout in eastern Canada, based on sequence analysis of the ND1 gene and a portion of the CO1 gene. The dendrogram was constructed from a Maximum Composite Likelihood distance matrix using pairwise deletion, and bootstrap values are based on 1000 replicates. Numbered asterisks refer to the following period of time since divergence: *1 = 92 977 years ago, *2 = 12 282 years ago, *3 = 10 262 years ago, *4 = 8 501 years ago.

Table 2.4. Haplotype frequency (*N*) and distribution among *Salvelinus fontinalis* populations in eastern Canada.

| Haplotype | *N* | Regions |
|---|---|---|
| 1 | 77 | Northern Labrador, West Labrador, Central Labrador, Southeast Labrador, Newfoundland, Québec |
| 2 | 23 | Northern Labrador, West Labrador, Central Labrador, New Brunswick |
| 3 | 7 | Newfoundland |
| 4 | 5 | West Labrador, Newfoundland |
| 5 | 4 | New Brunswick, Québec |
| 6 | 2 | Central Labrador |
| 7 | 2 | West Labrador |
| 8 | 1 | Newfoundland |
| 9 | 1 | New Brunswick |
| 10 | 1 | Nova Scotia |
| 11 | 1 | Central Labrador |
| 12 | 1 | Southeast Labrador |
| 13 | 1 | Nova Scotia |

Haplotype 3 fish (5.6%) were found only in Newfoundland, while haplotype 4 fish (4.0%) were found in west Labrador and Newfoundland. Haplotype 5 was found in New Brunswick and Québec, and represented 3.2% of fish. Haplotype 6 was found twice in central Labrador, accounting for 1.6% of fish. West Labrador and Newfoundland each had one haplotype 7 fish, again representing 1.6% of fish.

To further investigate the distribution of genetic diversity in Newfoundland and Labrador brook trout populations, three separate hierarchical AMOVAs were carried out (Table 2.5). First brook trout were grouped by watershed, then into regions. Most of the variation was detected within watersheds (72.59%; $F_{ST} = 0.2741$; $P < 0.00001$), followed by among watersheds within regions (27.84%; $F_{SC} = 0.2772$; $P < 0.00001$). The amount of variation partitioned among regions (-0.43%) was not significantly different from zero ($F_{CT} = -0.0043$; $P > 0.1$). Brook trout were then grouped according to lake, then into regions. Most of the variation was partitioned within lakes (78.27%; $F_{ST} = 0.2173$; $P < 0.00001$), followed by among lakes within regions (15.91%; $F_{SC} = 0.1690$; $P < 0.00001$). Again, there was no significant differentiation among regions (5.81%; $F_{CT} = 0.0581$; $P > 0.1$). Finally, grouping brook trout according to lake then subsequently into corresponding watershed revealed most of the variation partitioned among watersheds (62.89%; $F_{CT} = 0.6289$; $P < 0.00001$) followed by 28.49% of variation partitioned among watersheds ($F_{SC} = 0.7683$; $P < 0.00001$). The least amount of variation was detected within lakes (8.6152%; $F_{ST} = 0.9139$; $P < 0.00001$).

Table 2.5. Hierarchical AMOVA of brook trout populations in eastern Canada, with populations grouped according to watershed then into corresponding regions, and grouped according to lake, then into corresponding regions and watersheds. Variance was quantified by percentage of total variance (%) and conventional *F*-statistics ($F_{CT}$, $F_{SC}$, $F_{ST}$) in for each hierarchical level.

| | Among regions/ watersheds | | Among populations within watersheds/regions | | Within populations | |
|---|---|---|---|---|---|---|
| Comparison | % | $F_{CT}$ | % | $F_{SC}$ | % | $F_{ST}$ |
| i. Watersheds grouped regionally | -0.43 | -0.0043 | 27.84 | 0.2777 | 72.59 | 0.2741 |
| ii. Lakes grouped regionally | 5.18 | 0.0581 | 15.91 | 0.1690 | 78.27 | 0.2173 |
| iii. Lakes grouped into watersheds | 62.89 | 0.6289 | 28.49 | 0.7678 | 8.62 | 0.9139 |

*Regional distribution of haplotypes and population differentiation*

The geographical distribution of haplotypes was then investigated to identify any obvious patterns of heterogeneity that may indicate distinctive post-glacial founding lineages (Figure 2.6). Haplotypes 1 (represented in blue) and 2 (represented in red) are clearly heterogeneously distributed; haplotype 1 is the most common haplotype in Newfoundland and everywhere in Labrador except western where haplotype 2 dominates. Haplotype 2 is also common in northern and central Labrador but absent from southeastern Labrador and Newfoundland. Albeit with small sample sizes haplotype 1 is present in Québec, haplotype 2 in New Brunswick and neither in Nova Scotia. Newfoundland is characterized by a high frequency of haplotype 3 (represented in green), not seen anywhere else, and the presence of the highly divergent haplotype 4 (purple), also seen in west Labrador, but nowhere else. Lakes sampled in Québec and New Brunswick share the basal common haplotype 5 (turquoise). Newfoundland and all southern regions of Labrador contain unique or at least region-specific haplotypes, as do the lakes sampled New Brunswick. Nova Scotia contains only two unique haplotypes.

Regional heterogeneity in haplotype distribution was supported by the neighbour-joining dendrogram of pairwise $F_{ST}$ values among watersheds in Newfoundland and Labrador (Figure 2.7). Some north and central Labrador populations appear to be differentiated from the rest of Newfoundland and Labrador. A second cluster was comprised of watersheds from all other regions of Newfoundland and Labrador. Pairwise $F_{ST}$ between the two clusters measured 0.4464 and was significant ($P < 0.00001$).

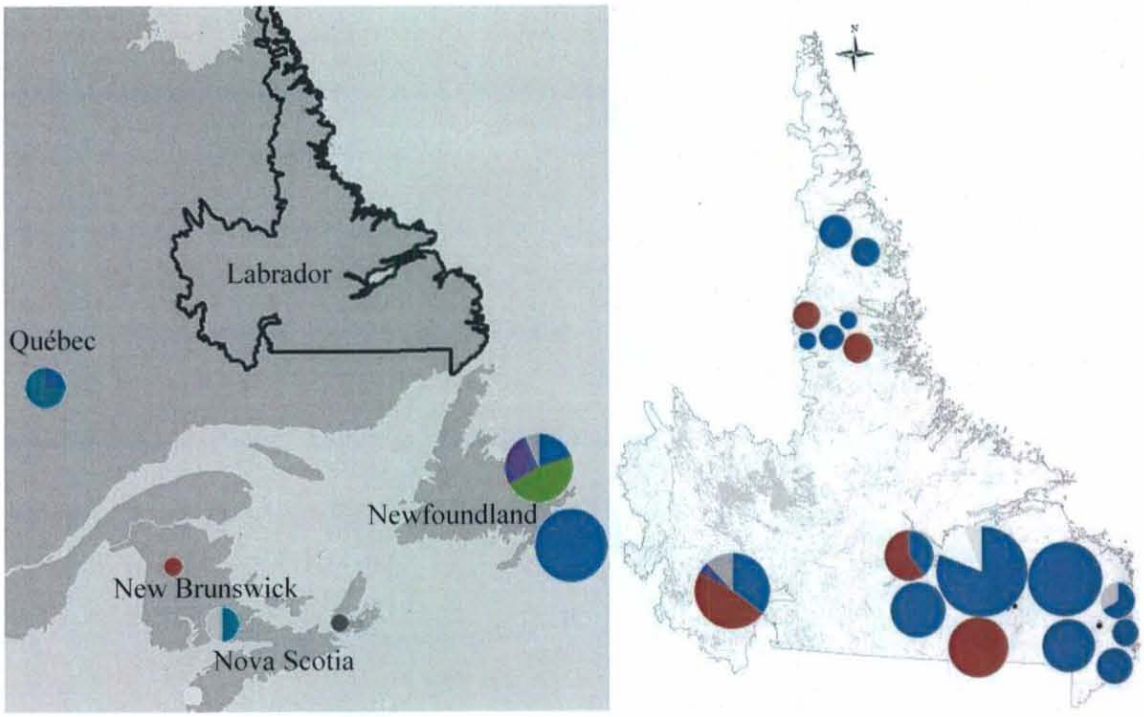


Figure 2.6. Geographical distribution according to watershed of the thirteen haplotypes among 126 brook trout in eastern Canada. Singly occurring haplotypes are shaded whereas shared haplotypes appear in color as follows: 1 = blue, 2 = red, 3 = green, 4 = purple, 5 = turquoise. Pie charts are sized according to relative sample size.

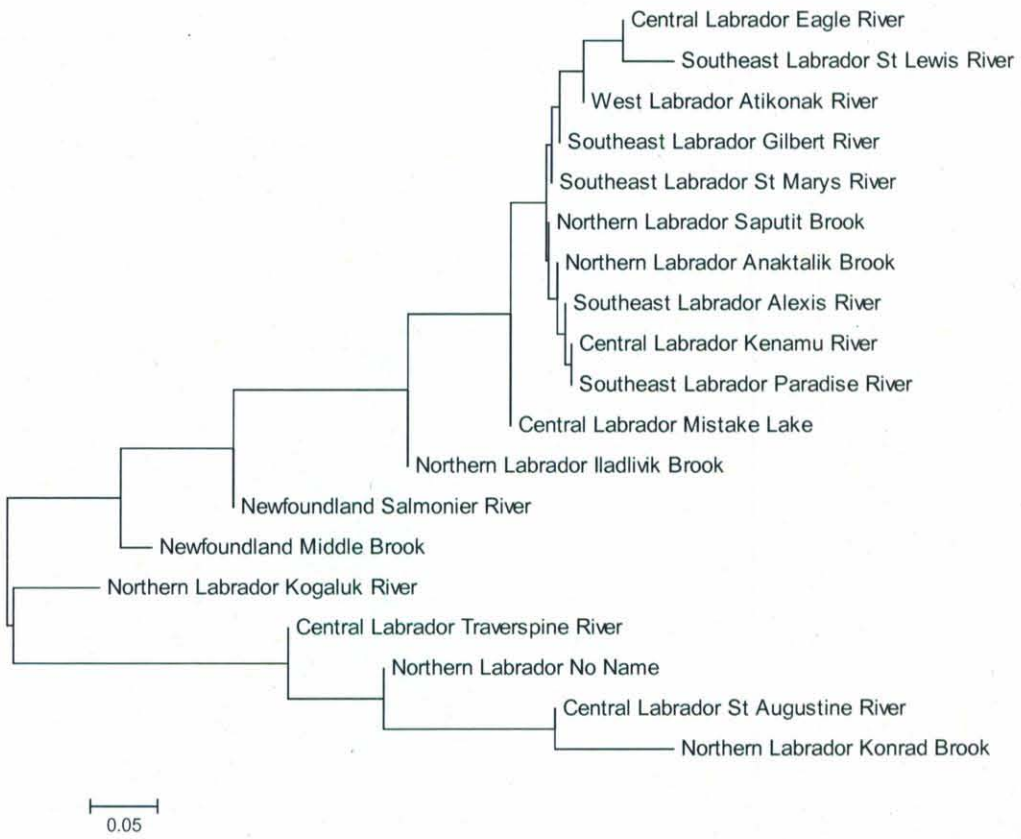


Figure 2.7. Neighbour-joining dendrogram of watersheds throughout Newfoundland and Labrador inferred from pairwise $F_{ST}$ values among watersheds, based on conventional $F$-statistics.

*Demographic history*

Tajima's (1989) *D* (a measure indicative of directional selection and demographic expansion) was measured in order to investigate the possibility of demographic expansion of brook trout populations in Labrador. Tajima's *D* for all brook trout across eastern Canada measured -1.33, however was not significantly negative ($P > 0.10$). To investigate the possibility of demographic expansion associated with haplotype 1 and its derivatives, Tajima's *D* was re-calculated for these haplotypes alone (1 plus its derivatives) and measured -1.30, remaining non-significantly negative ($P > 0.10$). In addition, the pairwise mismatch distribution or all haplotypes, as well as just haplotype 1 and its derivatives, followed that of a model of constant population size (Figure 2.8).

## DISCUSSION

Population genetic structure of brook trout in the northern part of their range was investigated by sequencing two mtDNA genes in 126 fish from eight regions in northeastern Canada. Levels of genetic diversity are described and patterns of variation within and among populations are used to make inferences about post-glacial colonization and demographic history of brook trout in this region. To complement a previous wider-scale study of brook trout (Danzmann *et al.* 1998), this study focuses particularly on Labrador, a large but un-investigated portion of the range.

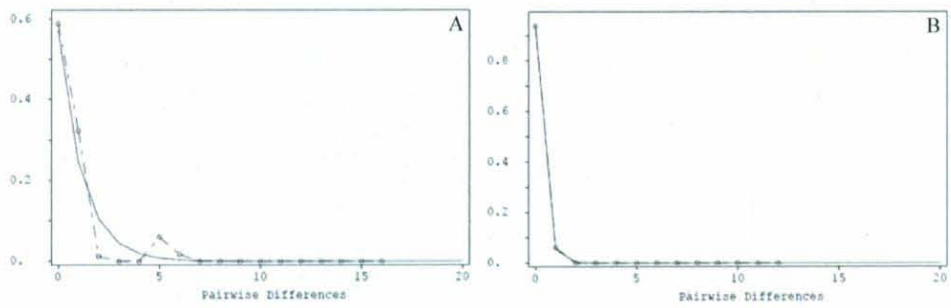


Figure 2.8. The observed (solid line) and expected (dashed line) pairwise mismatch distribution under constant population size model of brook trout in eastern Canada of all haplotypes (A), and of haplotype 1 and its derivatives (B).

*Low genetic variability characterizes eastern Canadian brook trout populations*

Pairwise sequence divergences among brook trout populations through eastern Canada are higher in the ND1 gene (0.00087) than the CO1 gene (0.00080), but overall are very low (0.00094). Despite this, within the coding regions, there are two nonsynonymous changes (Table 2.3) both in the ND1 gene, one of which (Val <-> Met) is a non-conservative change. Low genetic diversity in eastern Canadian brook trout populations was also illustrated by the haplotype distribution. 78.6% of brook trout are accounted for by only two haplotypes (haplotype 1 and 2), while 76.2% are represented by haplotype 1 and its derivatives.

The level of diversity observed here is low relative to both brook trout in general, and to other North American freshwater fishes. For example, in a brook trout population located in Ontario, nine RFLP haplotypes were identified among 33 individuals (Bernatchez & Danzmann 1993), compared to 14 haplotypes among 126 individuals in our study. In the Ontario population, average pairwise sequence diversity measured 0.010, an order of magnitude higher than what is observed in brook trout throughout eastern Canada. Taylor *et al.* (1999) characterized mtDNA sequence variation of a bull trout (*Salvelinus confluentus*) population in northwest North America, and reported a pairwise sequence diversity of 0.079, almost two orders of magnitude higher than what is observed in eastern brook trout. Arctic charr populations across their entire range (including northern regions such as Alpine Europe, Norway and Alaska) were also much more diverse (average pairwise sequence diversity = 0.0203; Brunner *et al.* 2001) than those found in our study of brook trout. The pairwise sequence diversity of lake whitefish

(*Coregonus clupeaformis*), throughout their northern distribution from Alaska to Labrador, was 0.007 (Bernatchez & Dodson 1991), again much higher than northeastern brook trout.

Low mtDNA diversity is characteristic of northern populations so this was not an unexpected finding here, but it is notable that levels are much lower than those observed in other northern-distributed freshwater fish, such as Arctic charr and whitefish. However, both of these species likely colonized their current northern distribution from a northern refugium (Bernatchez & Dodson 1991; Brunner *et al.* 2001), whereas brook trout populations have been hypothesized to have originated from more southern refugia (Danzmann *et al.* 1998). As a result, northern brook trout may have experienced a larger loss of alleles due to bottlenecking during expansion from more distant source populations than northern Arctic charr and whitefish populations which are in closer geographic proximity to their founding population.

*Unusual region of high diversity in the ND1 gene of brook trout*

Upon comparison of the brook trout sequences in this study with the mitochondrial genome of brook trout in GenBank (AF15480), a 220 bp region was found in the ND1 gene in which there were 11 differences between the brook trout in this study and the reference sequence. This corresponds to 5% sequence diversity, a very high level compared with the average pairwise sequence diversity reported intra-specifically in brook trout (this study and Bernatchez & Danzmann 1993). The nucleotide variants in our

fish match a GenBank entry of brook trout from a naturalised population in British Columbia, Canada (GenBank AF126000; the only brook trout mtDNA sequence available in GenBank other than the whole mtDNA genome sequence), while the nucleotide variants in the whole mtDNA genome sequence (GenBank AF15480) exactly match Arctic charr (GenBank AF154851) (Figure 2.2). Except for this 220 bp region, the consensus brook trout mtDNA sequence obtained in this study is very similar (~1.6%) to the whole mtDNA genome over the 1960 bp sequenced.

Complete replacement of brook trout mtDNA with Arctic charr mtDNA has been reported in brook trout populations in eastern Québec (Bernatchez *et al.*, 1995; Doiron *et al.*, 2002). Therefore, a reasonable explanation for this phenomenon is that the whole mtDNA genome sequence represents a brook trout whose mtDNA has recombined with Arctic charr mtDNA. Evidence of mtDNA recombination has been documented in a number of other species including vertebrates (Buroker *et al.* 1990; Awadalla *et al.* 1999; Hoarau *et al.* 2002; Slate & Phua 2003; Kraytsberg *et al.* 2004; Guo *et al.* 2006; Ujvari *et al.* 2007), invertebrates (Lunt & Hyman 1997; Ladoukakis & Zouros 2001; Burzynski *et al.* 2003; Van Wormhoudt *et al.* 2010), plants (Städler & Delph 2002; Jaramillo-Correa & Bousquet 2005; Manchekar *et al.* 2009) and fungi (Polashock *et al.* 1997; Saville *et al.* 1998). The most common mechanism for mtDNA recombination suggested is by means of paternal leakage of mitochondria during fertilization (Lunt & Hyman 1997; Ladoukakis & Zouros 2001; Städler & Delph 2002; Jaramillo-Correa & Bousquet 2005), which could be likely between brook trout and Arctic char as we know the two species interbreed. Paternal leakage is more likely to occur in hybridization zones because within

these regions reproductive barriers are not always entirely developed, and often the mechanism that eliminates male mitochondria may not be fully functional (Wagner *et al.* 1991; Jaramillo-Correa & Bousquet 2005). This provides further support of recombination between brook trout and Arctic charr mtDNA because, as previously mentioned, the instance documented in this study is in a region (Québec) where hybridization of the two species has been found (Bernatchez *et al.*, 1995; Doiron *et al.*, 2002).

Brook trout and Arctic charr hybridization has also been reported in northern Labrador (Hammar *et al.*, 1991), where the hybrids are physically most similar to brook trout, but could be distinguished by protein electrophoresis. It is interesting that none of the brook trout in our study appear to contain any Arctic charr mtDNA, as the habitats of these two species do overlap in our study area, and are located in a northern environment.

Doiron *et al.* (2002) suggested that the introgression event could be advantageous for brook trout in northern regions as these fish would have mitochondrial respiratory enzymes encoded by Arctic charr mtDNA that has evolved in a cold environment. In support of this, they found evidence that some regions of the mtDNA genome are under selection, especially in the ND2 and ND5 genes, but not ND1. We found only one conservative amino acid substitution (Val ↔ Ile) in the 220-bp region of high diversity, also suggesting no influence of selection in this gene.

The possibility that the 220 bp region of high diversity represents a pseudogene is unlikely as the nucleotide variants show the typical pattern of mtDNA variation; that is

they do not introduce any frameshifts or stop codons and, with the exception of two transversions, they are third position transitions. Finally, the possibility of laboratory error resulting in this phenomenon cannot be discounted as the entire brook trout mitochondrial genome submitted to GenBank was from a study in which the entire mitochondrial genome of Arctic charr was also sequenced with the same primers (Doiron *et al.* 2002). However, there is no strong evidence for this occurrence as none of the combinations of primers used by Doiron *et al.* (2002) to amplify these mtDNA sequences would produce PCR amplicons specifically of the size of the region of unusual diversity (220 bp).

*Differentiation among Newfoundland and Labrador brook trout populations and inferences about colonization history*

Although there are relatively low levels of variation within populations of brook trout in eastern Canada, significant structure was revealed among watersheds by the hierarchical AMOVA. This is a similar pattern to that reported by Danzmann *et al.* (1998). Significant variation primarily among watersheds may indicate that founding populations of brook trout were similar within but not among watersheds.

Patterns of mtDNA variation were also used in order to investigate the re-colonization of brook trout in eastern Canada, specifically through Labrador. The contribution of two founding sources was revealed by the distribution of haplotypes. Watersheds in north, west and central Labrador appear similar as they are dominated by

the presence of haplotype 1 and 2 fish, implying that these populations may have been founded by the same refugial source. The similarity of northern and central Labrador was supported by the neighbour-joining dendrogram in which two significantly differentiated groups were revealed ($F_{ST} = 0.4464$; $P < 0.00001$) - one that included some northern and central Labrador watersheds and another that included all regions. The absence of haplotype 2 fish in southeastern Labrador and insular Newfoundland and the high prevalence of restricted haplotypes suggest that these regions may have an alternate founding source than that which re-colonized the rest of Labrador; the similarity between southeastern Labrador and Newfoundland populations is also illustrated by the neighbour-joining dendrogram as watersheds in these regions are in the same cluster. However, this alternate source must have also had some contribution to west Labrador as both Newfoundland and west Labrador have haplotype 4 fish, the most ancestral haplotype as revealed by the neighbour-joining dendrogram of haplotypes. Post-glacial founding brook trout populations migrating from east to west (Newfoundland to west Labrador) would also likely colonize central Labrador, perhaps contributing the distribution-restricted haplotypes 6 and 12 fish, and some haplotype 2 fish. This is supported by the population dendrogram in that some central Labrador watersheds group with southeastern Labrador and insular Newfoundland.

Analysis of mtDNA reveals that the evolution of the majority of the haplotypes coincides with the timing of re-colonization of brook trout into Labrador, reported to have occurred anywhere from 13 000 to 5 000 years ago (Dyke & Prest 1987; Lamb 1980; Shaw 2006). The most frequent haplotype, haplotype 1 (along with haplotypes 3, 6, 7, 8,

9, 11 and 12 in this clade) originated approximately 10 262 years ago, while the clade composed of haplotypes 2, 5 and 10 evolved 12 282 years ago. The only exception is haplotype 4 which is ancestral, having evolved approximately 92 977 years ago. Although the evolution of these haplotypes coincides with glacier retreat from Labrador, there is little evidence that recolonization was coupled with population expansion. The star-like network of haplotype 1 and its derivatives, as well as the magnitude of Tajima's *D* (for all haplotypes = -1.33, for haplotype 1 and its derivatives = -1.30) are indicative of population expansion, however Tajima's D was not significant, and the pairwise mismatch distribution follows that of constant population size. It could be possible that the levels of variation in this portion of the mtDNA genome are simply too low to allow significance.

The previous chapter utilized microsatellite markers to make inferences on post-glacial colonization and also identified two founding populations, where northern and central Labrador were colonized by a different refugial source than the rest of Newfoundland and Labrador, possibly from some combination of Mississippian and/or Atlantic refugia. This pattern is supported based on mtDNA variation, with the exception that west Labrador appears to have the same founding source as northern and central Labrador. Based on microsatellite variation, west Labrador did not cluster with northern and central Labrador; this could be attributed to contemporary processes while observed patterns of mtDNA variation could reflect historical relationships. MtDNA variation also reveals genetic homogeneity among southeastern Labrador and Newfoundland, suggesting that these regions had the same founding populations, possibly from a north-

easterly Acadian source. Watersheds in these regions also group together on the neighbour-joining dendrogram. There is also evidence that the Acadian refuge may have somewhat contributed to west and central Labrador based on the presence of shared haploytpes in west Labrador and Newfoundland, as well as the restricted distribution of independent haplotypes in central and western Labrador. Similar colonization routes have been suggested by Black *et al.* (1986) and Danzmann *et al.* (1998). It has been postulated that brook trout from the Atlantic and Mississippian refugium migrated through Québec to recolonize Labrador, while subsequently the Atlantic refugia was a source for fish invading coastal regions (Black *et al.* 1986). It has also been suggested that northeastern brook trout populations were founded by fish from an Acadian refugia (Danzmann *et al.* 1998).

Evidence from this study supports an intermediate hypothesis: brook trout from some combination of an Atlantic and/or Mississippian refugia colonized north, west and central Labrador, while fish from an Acadian refugium invaded Newfoundland and southeastern Labrador, with some contribution to central and west Labrador populations. This hypothesis coincides well with evidence in the study carried out by Danzmann *et al.* (1998). They found that much of the brook trout's northern range was colonized by a single haplotype, haplotype '1', from an Atlantic and/or Mississippian refugia, which would correspond to the widespread haplotype 1 in our study. They attributed the presence of a number of restricted haplotypes, plus the absence of their haplotype 2 fish in eastern populations, to contribution from an alternate refugium, the Acadian. Based on the small sample sizes in this study, it is difficult to infer patterns of colonization through

Québec, New Brunswick and Nova Scotia, but the presence of haplotypes 1 and 2 throughout these regions support the contribution of all three refugia.

*Conclusions*

This study revealed that very low mtDNA diversity and structure at the level of watershed characterize populations of brook trout in eastern Canada, specifically in Labrador. The evolution of the majority of the haplotypes coincide with the timing of glacier retreat from Labrador, and my study reveals evidence for two source populations: one source may have recolonized north, west and central Labrador populations from some combination of Atlantic and/or Mississippian refugium, while another Acadian source dominates Newfoundland and southeastern Labrador, still with some contribution to west and central Labrador. My study also revealed that the only whole mtDNA genome sequence available on GenBank at the time (AF15480) may represent a brook trout mtDNA recombinant with that of Arctic charr mtDNA. Overall, the levels and patterns of mtDNA variation, as well as the timing of the evolution of the majority of the haplotypes characterized in this study demonstrate the strong influence that historical processes have on shaping modern population genetic structure of brook trout.

## SUMMARY

As a part of a ten year monitoring program by the Newfoundland and Labrador Department of Environment and Conservation, this study aimed to investigate the population genetic structure of brook trout (*Salvelinus fontinalis*) in Labrador, as well as neighbouring regions of eastern Canada. In order to determine the influence that both contemporary and historical factors have on genetic structure, two molecular markers, microsatellites and mitochondrial DNA (mtDNA), were utilized.

Overall genetic variation was found to be moderate to low based on microsatellites and mtDNA variation, respectively. Overall, heterozygosity ($H_E$) based on microsatellite diversity measured 0.622 which was within the spectrum of variability of other brook trout studies in their native range (Adams *et al.* 2003; Angers & Bernatchez 1998; Poissant *et al.* 2005; Rogers *et al.* 2004). The pairwise sequence divergence based on mtDNA variability was low, measuring 0.00094. Additionally, only two haplotypes dominated eastern Canadian brook trout populations (accounting for 78.6% of all fish). MtDNA variability was lower than that documented in a study of brook trout in Ontario (Bernatchez & Danzmann 1993), as well as lower than that of other freshwater fish with a similar range, such as Arctic charr (*S. alpinus*) (Brunner *et al.* 2001) and lake whitefish (*Coregonus clupeaformis*) (Bernatchez & Dodson 1991). The low diversity revealed by mtDNA is characteristic of populations at their range extremes, especially northern

populations in previously glaciated regions (Hewitt 1996, 1999), such as brook trout in Labrador.

While investigating intraspecific mtDNA nucleotide diversity, a high level of sequence diversity in the ND1 gene was found (5% as compared to 1.6% over the entire 1960 bp sequenced) when comparing our brook trout sequences to the only available brook trout whole mitochondrial genome in GenBank (AF15480). It has been suggested that the ND2, ND3 and ND5 genes of brook trout are under selection (Doiron *et al.* 2002), however, selection does not appear to play a role in the divergence of the ND1 gene in our fish as there was only a single conservative amino acid substitution in this region of high diversity. A possible explanation for this region of high diversity is that the whole mitochondrial genome of brook trout available on GenBank (AF15480) is from a fish whose mtDNA has recombined with the mtDNA of Arctic charr. In support of this, the related phenomenon of complete mtDNA introgression of Arctic char mtDNA has been reported (Bernatchez *et al.* 1995; Doiron *et al.* 2002).

Population genetic structure of brook trout populations based on both nuclear and mtDNA markers was also investigated. Overall, there was sufficient variation to detect population structure with microsatellites. Pairwise $F_{ST}$ values and STRUCTURE results revealed limited levels of ongoing gene flow. Additionally, the hierarchical AMOVA and STRUCTURE analysis showed that watershed was the limiting factor, a detail that has important conservation implications and suggests that future management decisions should based at this level. Levels of mtDNA variation also detected significant

structuring at the level of watershed, which suggests that founding populations of brook trout were similar within but not among watersheds.

Finally, patterns of microsatellite and mtDNA variation were utilized to make inferences about post-Wisconsinan recolonization of brook trout through eastern Canada, specifically in Labrador, and to investigate the demographic history of this area. Although microsatellite variation is influenced more by contemporary factors, reflecting fine-scale structure, microsatellites have also been successfully employed to reveal patterns of post-glacial recolonization (Angers & Bernatchez 1998; Taylor & McPhail 2000; Lu *et al.* 2001; Koskinen *et al.* 2002). The multidimensional scaling plot of $F_{ST}$ values and STRUCTURE results revealed that central and northern Labrador constitute a separate genetic cluster from the rest of the sampled regions (west Labrador, southeastern Labrador, insular Newfoundland, Nova Scotia, New Brunswick and Québec), providing evidence for dual routes of recolonization. This could be attributed to two different post-glacial founding groups, one with some combination of fish from the Atlantic and/or Mississippian refugia colonizing northern and central Labrador, and another with fish from the Acadian refugium colonizing the remainder of the regions. A similar pattern of two contributing sources was revealed by the mtDNA haplotype distribution. This analysis revealed one group composed of north, west and central Labrador populations, which may been founded by brook trout populations from some combination of Atlantic and/or Mississippian refugia and suggested that the Acadian made a major contribution to brook trout populations in insular Newfoundland and southeastern Labrador, and a minor contribution to central and west Labrador. Analysis of mtDNA sequence variation also

revealed that the timing of the evolution of the majority of clades timed with the retreat of ice from Labrador, anywhere from 13 000 to 5 000 before present (Dyke & Prest 1987; Lamb 1980; Shaw 2006), following the Wisconsinan glaciation.

In general, this study exemplifies how different factors, both contemporary and historical, have had a significant influence on shaping population genetic structure of brook trout populations in Labrador, and in eastern Canada. Microsatellites were able to detect significant fine-scale structure, revealing limited on-going gene flow throughout Labrador, while mtDNA variation was able to detect deeper structuring influenced by the Wisconsinan glaciation, and revealed evidence that the northernmost portion of the brook trout's native range may have been re-colonized by three different refugia. Overall, this study illuminates how contemporary and historical events have shaped the genetic structure of brook trout populations – essential information with important implications, such as for defining conservation and evolutionary significant units, in order to make responsible management decisions.

## REFERENCES


Adams BK, Hutchings JA (2003) Microgeographic population structure of brook charr: a comparison of microsatellite and mark-recapture data. *Journal of Fish* Biology, 62, 517-533.

Anderson L, Ryman N, Ståhl G (1983) Protein loci in the arctic charr *Salvelinus alpinus* L.: electrophoretic expression and genetic variability patterns. *Journal of Fish Biology*, 23, 75-94.

Angers B, Bernatchez L (1998) Combined use of SMM and non-SMM methods to infer fine structure and evolutionary history of closely related brook charr (*Salvelinus fontinalis* Salmonidae) populations from microsatellites. *Molecular Biology and Evolution*, 15, 143-159.

Angers B, Bernatchez L, Angers A, Desgroseillers L (1995) Specific microsatellite loci for brook charr reveal strong population subdivision on a microgeographic scale. *Journal of Fish Biology*, 47, 177-185.

Angers B, Magnan P, Plantes M, Bernatchez L (1999) Canonical correspondence analysis for estimating spatial and environmental effects on microsatellite gene diversity in brook charr (*Salvelinus fontinalis*). *Molecular Ecology*, 8, 1043-1053.

Avise JC, Arnold J, Ball RM (1987) Intraspecific phylogeography: the mitochondrial DNA bridge between population genetics and systematic. *Annual Review of Ecology and Systematics*, 18, 489-522.

Avise JC, Ball RM, Arnold J (1988) Current versus historical population sizes in vertebrate species with high gene flow: a comparison based on mitochondrial DNA lineages and inbreeding theory for neutral mutations. *Molecular Biology and Evolution,* 5, 331-344.

Avise JC, Neigel JE, Arnold J (1984) Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *Journal of Molecular Evolution*, 20, 99-105.

Avise JC (1989) Gene trees and organismal histories: a phylogenetic approach to population biology. *Evolution*, 43, 1192-1208.

Avise JC (2000) *Phylogeography: the history and formation of species*. Harvard University Press, Cambridge.

Avise JC 2004 *Molecular Markers Natural History and Evolution, 2nd ed.* Sinauer Associates, Sunderland.

Awadalla PA, Eyre-Walker A, Maynard Smith J (1999) Linkage disequilibrium and recombination in hominid mitochondrial DNA. *Science*, 284, 2090-2091.

Behnke RJ (1972) The systematics of salmonid fishes of recently glaciated lakes. *Journal of the Fisheries Research Board of Canada*, 29, 639-671.

Bennett KD (1997) *Evolution and Ecology: The Pace of Life*. Cambridge University Press, Cambridge.

Bermingham E, Forbes SH, Friedland K *et al.* (1991) Discrimination between Atlantic salmon (*Salmo salar*) of North American and European origin using restriction analyses of mitochondrial DNA. *Canadian Journal of Fisheries and Aquatic Sciences,* 48, 884-893.

Bernatchez L, Danzmann RG (1993) Congruence in Control-Region Sequence and Restriction Site Variation in Mitochondrial DNA of Brook Charr (*Salvelinus fontinalis* Mitchill). *Mololecular Biology and Evolution*, 10, 1002-1014.

Bernatchez L, Dodson JJ (1991) Phylogeographic structure in mitochondrial DNA of the lake whitefish (*Coregonus clupeaformis*) and its relation to Pleistocene glaciations. *Evolution*, 45, 1016-1035.

Bernatchez L, Dodson JJ, Boivin S (1989) Population bottlenecks: influence on mitochondrial DNA diversity and its effect in coregonine stock discrimination. *Journal of Fish Biology*, 35, 233-244.

Bernatchez L, Glemet H, Wilson CC, Danzmann R (1995) Introgression and fixation of arctic char (*Salvelinus alpinus*) mitochondrial genome in an allopatric population

of brook trout (*Salvelinus fontinalis*). *Canadian Journal of Fisheries and Aquatic Sciences*, 52, 179-185.

Birky CW, Fuerst P, Maruyama T (1989) Organelle gene diversity under migration mutation and drift: equilibrium expectations approach to equilibrium effects of heteroplasmic cells and comparison to nuclear genes. *Genetics*, 121, 613-627.

Black GA, Dempson JB, Bruce WJ (1986) Distribution and postglacial dispersal of freshwater fishes of Labrador. *Canadian Journal of Zoology*, 64, 21-31.

Boula D, Castric V, Bernatchez L, Audet C (2002) Physiological endocrine and genetic bases of anadromy in brook charr (*Salvelinus fontinalis*) of the Laval River (Québec Canada). *Environmental Biology of Fishes*, 64, 229-242.

Bradshaw AD (1965) Evolutionary significance of phenotypic plasticity in plants. *Advances in Genetics*, 13, 115-155.

Brunner PC, Douglas MR, Osinov A, Wilson CC, Bernatchez L (2001) Holarctic phylogeography of Acrtic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. *Evolution*, 55, 573-586.

Buroker NE, Brown JR, Gilbert TA, O'Hara TJ, Beckenbach AT, Thomas WK, Smith MJ (1990) Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. *Genetics*, 124, 157-163.

Burzynski A, Zbawicka M, Skibinski DO, Wenne R (2003) Evidence for recombination of mtDNA in the marine mussel *Mytilus trossulus* from the Baltic. *Molecular Biology and Evolution*, 20, 388-392.

Castric V, Bernatchez L (2003) The rise and fall of isolation by distance in the anadromous brook charr (*Salvelinus fontinalis* Mitchill). *Genetics*, 163, 983-996.

Castric V, Bonney F, Bernatchez L (2001) Landscape structure and hierarchical genetic diversity in the brook charr, *Salvelinus fontinalis*. *Evolution*, 55, 1016-1028.

Danzmann RG, Morgan R, Jones M, Bernatchez L (1998) A major sextet of mitochondrial DNA phylogenetic assemblages extant in eastern North American brook charr (*Salvelinus fontinalis*): distribution and postglacial dispersal patterns. *Canadian Journal of Zoology*, 76, 1300-1318.

DFO (Economic Analysis and Statistics Policy Sector Canadian Department of Fisheries and Oceans) (2007) *2005 Survey of Recreational Fishing in Canada*, Ottawa, Ontario.

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

Dizon AE, Lockyer C, Perrin WF, Demaster DP, Sisson J (1992) Rethinking the stock concept: A phylogeographic approach. *Conservation Biology,* 6, 26-36.

Doiron S, Bernatchez L, Blier PU (2002) Comparative mitogenomic analysis of the potential adaptive value of Arctic charr mtDNA introgression in brook charr populations. *Molecular Biology and Evolution*, 19, 1902-1909.

Dyke AS, Prest VK (1987) Late Wisconsinan and Holocene history of the Laurentian ice sheet. *Géographie physique et Quaternaire*, 41, 237-263.

Dynes J, Magnan P, Bernatchez L, Rodriguez MA (1999) Genetic and morphological variations between two forms of lacustrine brook charr. *Journal of Fish Biology*, 54, 955-972.

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

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.

Felsenstein J (1989) PHYLIP-phylogeny inference package (Version 3.2). *Cladistics,* 5, 164-166.

Field-Dodgson MJ (1988) Size characteristics and diet of emergent Chinook salmon in a small, stable, New Zealand stream. *Journal of Fish Biology*, 32, 27-40.

Flanders J, Jones G, Benda P, Dietz C, Zhang S (2009) Phylogeography of the greater horseshoe bat *Rhinolophus ferrumequinum*: contrasting results from mitochondrial and microsatellite data. *Molecular Ecology*, 18, 306-318.

Frankham R, Ballou JD, Briscoe DA (2002) *Introduction to Conservation Genetics, 1st ed.* Cambridge University Press, New York.

Fraser DJ, Lippé C, Bernatchez L (2004) Consequences of unequal population size asymmetric gene flow and sex-biased dispersal on population structure in brook charr (*Salvelinus fontinalis*). *Molecular Ecology*, 13, 67-80.

Glémet H, Blier P, Bernatchez L (1998) Geographical extent of Arctic charr (*Salvelinus alpinus*) mtDNA introgression in brook charr populations (*S. fontinalis*) from eastern Québec, Canada. *Molecular Ecology,* 7, 1655-1662.

Goudet J (1995) FSTAT (Version 1.2): A Computer Program to Calculate F-Statistics. *Journal of Heredity*, 86, 485-486.

Guo X, Liu S, Liu Y (2006) Evidence for Recombination of Mitochondrial DNA in Triploid Crucian Carp. *Genetics*, 172, 1745-1749.

Hamilton KE, Ferguson A, Taggart JB, Tomasson T (1989) Postglacial colonization of brown trout *Salmo trutta* L.: *Ldh-5* as a phylogeographic marker locus. *Journal of Fish Biology*, 35, 651-664.

Hammar J, Dempson JB, Verspoor E (1991) Natural hybridization between Arctic char (*Salvelinus alpinus*) and brook trout (*S. fontinalis*): evidence from northern Labrador. *Canadian Journal of Fisheries and Aquatic Sciences*, 48, 19-32.

Hardy OJ, Vekemans X (2002) SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, 2, 618-620.

Hardy OJ, Charbonnel N, Freville H, Heuertz M (2003) Microsatellite allele sizes: a simple test to assess their significance on genetic differentiation. *Genetics*, 163, 1467-1482.

Hernandez-Martich JD, Smith MH (1990) Patterns of genetic variation in eastern mosquitofish (*Gambusia holbrooki* Girard) from the piedmont and coastal plain of three drainages. *Copeia*, 3, 619-630.

Hewitt GM (1996) Some genetic consequences of ice ages and their role in divergence and speciation. *Biological Journal of the Linnaean Society*, 58, 247-276.

Hewitt GM (1999) Post-glacial recolonization of European biota. *Biological Journal of the Linnaean Society*, 68, 87-112.

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

Hocutt CH, Wiley EO (1986) *The Zoogeography of North American Freshwater Fishes*. John Wiley and Sons, New York.

Hoarau G, Holla S, Lescasse R, Stam WT, Olsen JL (2002) Heteroplasmy and evidence for recombination in the mitochondrial control region of the flatfish *Platichthys flesus*. *Molecular Biology and Evolution*, 19, 2261-2264.

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

Hutchings JA (1996) Adaptive phenotypic plasticity in brook trout, *Salvelinus fontinalis*, life histories. *Ecoscience*, 3, 25-32.

Jaramillo-Correa JP, Bousquet J (2005) Mitochondrial genome recombination in the zone of contact between two hybridizing conifers. *Genetics*, 171, 1951-1962.

Jones MW, Danzmann RG, Clay D (1997) Genetic relationships among populations of wild resident and wild hatchery anadromous brook charr. *Journal of Fish* Biology, 51, 29-40.

Knopp T, Merila J (2009) The postglacial recolonization of Northern Europe by *Rana arvalis* as revealed by microsatellite and mitochondrial DNA analyses. *Heredity*, 102, 174-181.

Kornfield I, Beland KF, Moring JR, Kircheis FW (1981) Genetic similarity among endemic arctic char (*Salvelinus alpinus*) and implications for their management. *Canadian Journal of Fisheries and Aquatic Sciences*, 38, 32-39.

Koskinen MT, Nilsson J, Veselov A, Potutkin AG, Ranta E, Primmer CR (2002) Microsatellite data resolve phylogeographic patterns in European grayling, *Thymallus thymallus*, Salmonidae. *Heredity*, 88, 391-401.

Kraytsberg Y, Schwartz M, Brown TA, Ebralidse K, Kunz WS, Clayton DA, Vissing J, Khrapko K (2004) Recombination of human mitochondrial DNA. *Science*, 304, 981.

Ladoukakis ED, Zouros E (2001) Direct evidence for homologous recombination in mussel (*Mytilus galloprovincialis*) mitochondrial DNA. *Molecular Biology and Evolution*, 18, 1168-1175.

Lamb HF (1980) Late Quaternary Vegetational history of southeastern Labrador. *Arctic, Antarctic and Alpine Research*, 12, 117-135.

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

Lu G, Basley DJ, Bernatchez L (2001) Contrasting patterns of mitochondrial DNA and microsatellite introgressive hybridization between lineages of lake whitefish (*Coregonus clupeaformis*): relevance for speciation. *Molecular Ecology*, 10, 965-985.

Lunt DH, Hyman BD (1997) Animal mitochondrial DNA recombination. *Nature*, 387, 247.

Manchekar M, Scissum-Gunn K, Song D, Khazi F, McLean SL, Nielsen BL (2006) DNA recombination activity in soybean mitochondria. *Journal of Molecular Biology*, 356, 288-299.

McAllister DE, Platania SP, Schueler FW *et al.* (1986) Ichthyofaunal patterns on a geographic grid. In: *The Zoogeography of North American Freshwater Fishes* (eds Hocutt CH, Wiley EO), pp. 17-51. John Wiley and Sons, New York.

McCrimmon HR, Campbell JS (1969) World distribution of brook trout *Salvelinus fontinalis*. *Journal of the Fisheries Research Board of Canada*, 26, 1699-1725.

McGlade JM (1981) Genotypic and phenotypic variation in brook trout *Salvelinus fontinalis* (Mitchill). PhD Thesis. University of Guelph, Guelph, Ontario.

Mitton JB (1989) Physiological and demographic variation associated with allozyme variation in *Isozymes in Plant Biology*. Dioscorides Press, Portland, OR, pp. 127-145.

Moritz C, Dowling TE, Brown WM (1987) Evolution of animal mitochondrial DNA: relevance for population biology and systematic. *Annual Review of Ecology and Systematics*, 18, 269-292.

Moritz C, Faith DP (1998) Comparative phylogeography and the identification of genetically divergent areas for conservation. *Molecular Ecology*, 7, 419-429.

Nei M, Tajima F, Tateno Y (1983) Accuracy of estimated phylogenetic trees from molecular data. *Journal of Molecular Evolution*, 19, 153-170.

Paetkau D, Waits LP, Clarkson PL, Craighead L, Strobeck C (1997) An empirical evaluation of genetic distance statistics using microsatellite data from bear (Ursidae) populations. *Genetics*, 147, 1943-1957.

Page RDM (1996) Treeview: an application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*, 12, 357-358.

Pielou EC (1991) *After the Ice Age: the Return of Life to Glaciated North America*. University of Chicago Press, Chicago.

Poissant J, Knight TW, Ferguson MM (2005) Nonequilibrium conditions following landscape rearrangement: the relative contribution of past and current hydrological landscapes on the genetic structure of a stream-dwelling fish. *Molecular Ecology*, 14, 1321-1331.

Polashock JJ, Bedker PJ, Hillman BI (1997) Movement of a small mitochondrial double-stranded RNA element of *Cryphonectria parasitica*: ascospore inheritance and implications for mitochondrial recombination. *Molecular Genetics and Genomics*, 256, 566-571.

Power G (1980) The brook charr, *Salvelinus fontinalis* in *Charrs: salmonid fishes of the genus Salvelinus*. Dr. W. Junk bv Publishers, The Hague, The Netherlands, pp. 141-203.

Presa P, Guyomard R (1996) Conservation of microsatellites in three species of salmonids. *Journal of Fish Biology*, 49, 1326-1329.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics,* 155, 945-959.

Pudovkin AI, Zaykin DV, Hedgecock D (1996) On the potential for estimating the effective number of breeders from heterozygote-excess in progeny. *Genetics*, 144, 383-387.

Rasmussen DI (1979) Sibling clusters and genotypic frequencies. *American Naturalist*, 113, 948-951.

Raymond M, Rousset F (1995) GENEPOP (version 1.2): population genetics software for exact test and ecumenicism. *Journal of Heredity*, 86, 248-249.

Rogers SM, Curry RA (2004) Genetic population structure of brook trout inhabiting a large river watershed. *Transactions of the American Fisheries Society*, 133, 1138-1149.

Ryder OA (1986) Species conservation and systematic: the dilemma of subspecies. *Trends in Ecology and Evolution*, 1, 9-10.

Saville BJ, Kohli Y, Anderson JB, (1998) MtDNA recombination in a natural population. *Proceedings of the National Academy of Sciences of the United States of America*, 95, 1331-1335.

Schmidt RE (1986) Zoogeography of the northern Appalachians. In: *The zoogeography of North American freshwater* fishes (eds Hocutt CH, Wiley EO), pp. 137-159. John Wiley Sons, Toronto, Ontario.

Scott W, Crossman E (1973) Freshwater fishes of Canada. *Bulletin of the Fisheries Research Board of Canada,* 184.

Sefc KM, Payne RB, Sorenson MD (2007) Genetic differentiation after founder events: an evaluation of $F_{ST}$ estimators with empirical and simulated data. *Evolutionary Ecology Research*, 9, 21-39.

Shaw J (2006) Palaeography of Atlantic Canadian Continental Shelves from the last glacial maximum to the present, with an emphasis on Flemish Cap. *Journal of Northwest Atlantic Fishery Science*, 37, 119-126.

Shaw PW, Carvalho GR, Magurran AE, Sheghers BE (1991) Population differentiation in Trinidadian guppies (*Poecilia reticulata*): patterns and problems. *Journal of Fish Biology,* 39, 203-209.

Slate J, Phua H (2003) Patterns of linkage disequilibrium in mitochondrial DNA of 16 ruminant populations. *Molecular Ecology*, 12, 597-608.

Snucins, EJ, Curry RA, Gunn JM (1992) Brook trout (*Salvelinus fontinalis*) embryo habitat and timing of alevin emergence in a lake and a stream. *Canadian Journal of Zoology*, 70, 423-427.

Städler T, Delph LF (2002) Ancient mitochondrial haplotypes and evidence for intragenic recombination in a gynodioecious plant. *Proceedings of the National Academy of Sciences of the United States of America*, 99, 11730-11735.

Strobeck C (1979) Partial selfing and linkage: the effect of a heterotic locus on a neutral locus. *Genetics*, 92, 305-315.

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 EB, Pollard S, Louie D (1999) Mitochondrial DNA variation in bull trout (*Salvelinus confluentus*) from northwestern North America: implications for zoogeography and conservation. *Molecular Ecology,* 8, 1155-1170.

Taylor EB, Kuiper A, Troffe PM, Hoysak DJ, Pollard S (2000) Variation in developmental biology and microsatellite DNA in reproductive ecotypes of kokanee, *Oncorhynchus nerka*: implications for declining populations in a large British Columbia lake. *Conservation Genetics*, 1, 231-249.

Tessier N, Bernatchez L, Presa P, Angers B (1995) Gene diversity analysis of mitochondrial DNA microsatellites and allozymes in landlocked Atlantic salmon (*Salmo salar* L.).*Journal of Fish Biology,* 47, 156-163.

Thomas WK, Withler RE, Beckenbach AT (1986) Mitochondrial DNA analysis of Pacific salmonid evolution. *Canadian Journal of Zoology,* 64, 1058-1064.

Thomas WK, Beckenbach AT (1989) Variation in Salmonid Mitochondrial DNA: Evolutionary Constraints and Mechanisms of Substitution. *Journal of Molecular Evolution*, 29, 233-242.

Ujvari B, Dowton M, Madsen T (2007) Mitochondrial DNA recombination in a free-ranging Australian lizard. *Biology Letters*, 3, 189-192.

Van Wormhoudt A, Le Bras Y, Huchette S (2009) *Haliotis marmorata* from Senegal; a sister species of *Haliotis tuberculata*: morphological and molecular evidence. *Biochemical Systematics and Ecology*, 37, 747-757.

Wagner DB, Dong J, Carlson MR, Yanchuk AD (1991) Paternal leakage of mitochondrial DNA in *Pinus*. *Theoretical and Applied Genetics*, 82: 510-514.

Ward RD, Woodwark M, Skibinski DOF (1994) A comparison of genetic diversity levels in marine freshwater and anadromous fishes. *Journal of Fish Biology*, 44, 213-232.

Weir BS, Cockerham CC (1984) Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 1358-1370.

Wilson AC, Cann RL, Carr SM, *et al.* (1985) Mitochondrial DNA and two perspectives on evolutionary genetics. *Biological Journal of the Linnaen Society*, 26, 375-400.

Wilson GM, Thomas WK, Beckenbach AT (1985) Intra- and inter- specific mitochondrial DNA sequence divergence in *Salmo*: rainbow steelhead and cutthroat trouts. *Canadian Journal of Zoology*, 63,

Wilson GM, Thomas WK, Beckenbach AT (1987) Mitochondrial DNA analysis of Pacific Northwest populations of *Oncorhynchus tshawytscha*. *Canadian Journal of Fisheries and Aquatic Sciences*, 44, 1301-1305.

Wilson CC, Hebert PDN, Reist JD, Dempson JB (1996) Phylogeography and postglacial dispersal of arctic charr, *Salvelinus alpinus* in North America. *Molcular Ecology*, 5, 187- 197.

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.

Woodruff DS (1989) The problems of conserving genes and species. In: *Conservation for the Twenty First Century* (eds Western D, Pearl M), pp. 2088-2094. Oxford University Press, New York.