Using Major Histocompatibility Genes Polymorphism to Identify Arctic Charr (*Salvelinus alpinus*) Populations

by

Pablo Andrés Conejeros Abraham

A thesis

presented to the University of Waterloo

in fulfillment of the

thesis requirement for the degree of

Doctor of Philosophy

in

Biology

Waterloo, Ontario, Canada, 2007

I hereby declare that I am sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be electronically available to the public.

Abstract

Arctic charr is the most northerly distributed salmonid and the most abundant fish in high latitude postglacial lakes. Arctic charr lives in oligotrophic water bodies where it has been able to adapt and thrive due, in part, its noted outstanding phenotypic plasticity. Throughout its geographic range, the Arctic charr has had to specialize to get the most of each ecosystem, to the point that Arctic charr were originally described as 56 different species and only later considered as many phenotypic variations of the same group, called the Arctic charr complex. With the aim of using the resources available in areas with very low primary production, Arctic charr often specialize to become different morphotypes within the same water body. Each morphotype can follow different life histories that can be anadromous or non-migratory. In several lakes, the non-migratory stocks may also differentiate further, each form with its own trophic and/or reproductive behavior. Adult sympatric forms may differ in depth distribution, size-at-maturity, time and place of spawning, color and/or other meristic characters that include differential gill raker and vertebrae numbers. The two typical forms that are found in sympatry are a small, profoundal form often termed “dwarf” Arctic charr and a large, littoral or pelagic zone resident often termed “Normal” Arctic charr.

The Arctic charr colonized most of its current habitat very recently, after the ice retreat in the late Pleistocene, 10000-15000 years ago. The reproductive isolation of stocks, if it has occurred at all, occurred so recently that the accumulated genetic drift often does not yield enough data to support the genetic separation of the stocks. Since the geographic borders of the stocks tend to be unclear and because the Arctic charr is a migratory species, the management of fisheries can be difficult in light of these issues, this thesis examines the potential for identifying Arctic charr populations using Major Histocompatibility (MH) genes as molecular markers.

MH genes are useful because they are not neutral markers, but are subject to selection. MH receptors present peptides to T-lymphocytes and from that interaction the immune system defines what is self or non-self and thus whether or not immune reactions should be initiated. Due to the large variety of potential pathogenic peptides to be presented, the domain of the MH receptor that binds the peptide, the peptide binding region, is the most polymorphic coding region known. Each individual has a limited number of MH alleles. Given the high degree of polymorphism in populations it is virtually impossible that two individuals will share the same set, of MHC alleles with the exception of monozygotic twins. Since MH receptors present peptides derived from pathogens, they are related to disease resistance, and some MH alleles are more effective at presenting certain peptides than others. Therefore, populations settled in a specific niche will interact with a defined variety of pathogens that will select for certain patterns in the MH alleles of the population. The selection of these MH allelic patterns occurs rapidly, since they determine the survival of the individuals during disease outbreaks. Rapid selection means that MH allelic patterns they can be used to differentiate populations that have been separated for relatively short periods of time.

The MH genes of Arctic charr had not been characterized before the publication of this thesis, so the first step was their isolation and characterization. We found the MH sequences obtained to have typical characteristics of classical MH receptors, sharing similarities with other salmonids and having most of their variation in the peptide binding region. We next characterized populations of Arctic charr selected from the global distribution using the three polymorphic MH receptors. For all of the receptor we found most of the polymorphisms distributed equally amongst the populations, but the interpopulation diversity was generally enough to differentiate at least some of the studied populations.

For the MH Class I we studied three non classical (UCA, UGA, UEA) and one classical (UBA) gene. For UBA and UCA we found a large degree of polymorphism while UGA and UEA were not very polymorphic. Despite the fact that the UGA gene was also not polymorphic in studies of rainbow trout, we found the gene to be the best Class I population marker for Arctic charr because it had the highest relative rates of interpopulation diversity. Thus, UGA may be exhibiting some antigen presentation functions in Arctic charr. The population analysis using MH Class II α and Class II β genes were the most successful. Particularly in the case of Class II β, the analyses arose capable of differentiating all the populations chosen for this study. Both genes showed high levels of polymorphism and high rates of non-synonymous/synonymous substitution in the exon that encodes the peptide binding region. Lastly, we used MHC Class II α and Class II β to differentiate two separate sets of morphotypes living in sympatry in Lake Kiryalta in Russia and Gander Lake in Canada. The morphotypes in Gander Lake were successfully differentiated using both MH Class II α and β allele data, while the morphotypes in Lake Kiryalta were separated only with the MH Class II β allele data.

Given that the use of one or more MH genes used allowed us to differentiate the populations studied, MH genes seem to be extremely useful as population markers for Arctic charr. Since MH genes not only characterize populations according to their phylogenetic relationships, but also according to their specific adaptation to inhabited niches, we concluded that all the Arctic charr populations studied are independent evolutionary significant units of the Arctic charr species. The conclusion implies that although different stocks might be living in sympatry, they should be considered as separate species for fishery and other management purposes, because their specific adaptations to the pathogens in their ecological niche might not allow them to

cross-repopulate the other stock if it were removed by over-fishing or other anthropogenic stresses.

Acknowledgements

Like a “Minga Chilota”, this work is product of so many people and moments. Aware that I’m missing so many, I’d like to thank,

My supervisor Dr. Brian Dixon and the members of my committee, Dr. Michael Power and Dr. Jonathan Witt. My Lab Partners, Marsela, Suchita, Anna, Anathea, Steve, Kazu, Shathi, Stacey, Darah, Leandro.

My friends and brothers of snow and forest, Sairah, Greg, Margaret, Megan, Vicky, Jamie, Tara, Reagan, Jen, Ted, Paul, Sarah, Scott, Mohamed, Jim, Richard.

My old and new Lab Partners, Dr. Sergio Marshall, Tali, Vero, Marcela, Anita, Pato, Paulina, Sandra. Amigos, Ivan, Eulogio, Gloria, Alfonso, Paola.

A mi amiga y compañera, Orietta. A mi Padre, Patricio y a mi Madre, Julia. A mis hermanos Macarena y Esteban.

A mi Tata Rene y a mi Yeya Lucia. Al Tata Pepe, y Miguel, la abuelita Tina, y Julia.

A mis primos hermanos Alejandro, Rene, Daniel, Fernando, Sole. A mi tio Rene.

A mi tia Vivi, Mirtha, Olga, Yola, tio Fico. A Matilda, Amaru, Noa, Mia, Camila.

A mis amigos hermanos Cristian, Palma, Lopez, Saabedra, Pacheco, Ortega, Roberto, Quilo, Zuñiga. A la Gala, Toteles, Lobo, Mitsco, Capitan, la Flor.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AA	Amino acid
AIDS	Acquired immune deficiency syndrome
APC	Antigen presenting cell
CD4	Cluster of differentiation 4
CD8	Cluster of differentiation 8
cDNA	Complementary deoxyribonucleic acid
DGGE	Denaturant gradient gel electrophoresis
DIG	Dioxigenin
DNA	Deoxyribonucleic acid
EDTA	ethylenediaminetetraacetic acid
ESU	Evolutionary significant unit
HERV	Human endogenous retrovirus
HLA	Human Leukocyte Antigen
IgA	Immunoglobulin A
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ISAV	Infectious salmon anemia virus
LB	Luria-Bertani media
mDNA	Mitochondrial DNA
MH	Major histocompatibility

MHC	Major histocompatibility complex
mRNA	Messenger RNA
NOAA	National Oceanic and Atmospheric Administration
PBR	Peptide binding region
PCR	Polymerase chain reaction
PIPES	Piperazine-N,N′-bis(2-ethanesulfonic acid)
RACE	Rapid amplification of cDNA ends
RER	Rough endoplasmic reticule
RFLP	Restriction fragment length polymorphism
RNA	Ribonucleic acid
RT-PCR	Reverse transcription polymerase chain reaction
SDS	Sodium dodecyl sulfate
SINE	Short interspersed nuclear element

Chapter 1. General Introduction

General Introduction

Distribution of Arctic charr

Arctic charr (*Salvelinus alpinus* L.) is the most northerly distributed salmonid, and the most abundant fish in Arctic postglacial lakes. It has an holarctic distribution, inhabiting all the landmasses that surround the Arctic Ocean (Johnson 1980). Most of the area inhabited by Arctic charr is of post glacial origin and has only emerged recently from under the Pleistocene glaciers that started to retreat 18000 years ago (Svendsen and Mangerud 1992). Several lakes where Arctic charr are currently found were still covered by ice as recently as 1100 years ago (Johnson 1980). The southern parts of the continental landmasses, that might have been refugia for Arctic charr during the Pleistocene, are now dominated by species better adapted to warmer conditions, so the southernmost Arctic charr population now occurs in Maine at approximately 45 º latitude. On the other hand Arctic charr live as far north as 82 º N in Ellesmere Island (Parker and Johnson 1991). Throughout their range, Arctic charr usually inhabit oligotrophic and ultraoligotrophic lakes (Klemetsen *et al*. 2003) where it is often the only fish species.

Arctic charr are extremely polymorphic, presenting several different forms and behaviors throughout the geographic distribution. The morphological differences are apparent in color and size, but also in traditional meristic characters such as the number of vertebrae and gill rakers. The different morphotypes caused the original taxonomic classification of the Arctic charr to be divided into 56 different species (Savvaitova 1993), that publication after publication have slowly redefined as only one group named the Arctic charr complex.

Morphotypes and stocks of Arctic charr

There are around 50,000 populations of Arctic charr around the world, 1300 of them anadromous (Klemetsen *et al*. 2003). Anadromous populations of Arctic charr tend to be more constant in shape and behavior, although there are cases where two or more morphotypes of anadromous charr are found in sympatry, with well differentiated adult sizes (Johnson 1980). Anadromous Arctic charr tend to inhabit the middle portion of the distribution, with non-migratory stock dominating at the northern and southern populations of the range. Migration of Arctic charr has taken the species across long ocean passages, for example the colonization of Jan Mayen Island, 450 km from the nearest land (Johnson 1980). Nevertheless, it is generally considered that Arctic charr have a remarkable homing capacity, and tagged fish moved to different locations tend to return quickly to the capture point even when they have been mixed with other charr populations (Svenning and Grotnes 1991). Non-anadromous Arctic charr populations tend to inhabit southern waters, and they usually spawn for the first time at younger ages than anadromous charr. They spawn in lakes or rivers during the fall and exclusively in lakes during spring. Non-anadromous charr often develop different forms within the same water body, even in some lakes where the ice retreated very recently (Skreslet 1973). The specialization of Arctic charr in different sympatric forms occurs through the geographic range and to an extent that has not been witnessed in any other species (Klemetsen *et al*. 2003).

Two or more forms of Arctic charr are usually distinguishable from each other because they present differential depth distributions, sizes at maturity, times and/or places of spawning, color and/or other meristic characters. The most common sympatric Arctic charr morphotypes are the Normal and dwarf forms. Normal Arctic charr are dark in color and bright red along the sides and belly at maturation, similar to other adult salmonids. Normal Arctic charr usually live at

depths of less than 60 m (Klemetsen *et al*. 2003) and they feed pelagically or littorally, consuming fish (often smaller Arctic charr), as well as hatching and flying insects. Dwarf Arctic charr tend to mature at an earlier ontogenic state than normal Arctic charr, thus growth stops prematurely as energy is channelled into reproductive development. Dwarf charr are uniformly pale white, have larger upper than the lower jaws, have eyes that are larger with respect to the head and longer pectoral fins than Normal charr (Jonsson and Jonsson 2001). The dichotomy in mouth position and pectoral fin size usually corresponds with trophic adaptation (Lavin and McPhail 1986). Dwarf Arctic charr usually live at depths below 90 m and are bottom feeders, where much of the primary production in Arctic lakes occurs. They consume mainly zoobenthos, but they can also be planktivores in some cases (Knudsen *et al*. 2006).

The growth rate of the normal morphotype exceeds that of the dwarf morphotype, and normal Arctic charr mature at around 300-700 mm, while the dwarfs mature at around 200-400 mm. Nevertheless, mature dwarf females of 70 mm have been captured (Klemetsen *et al*. 2003). Normal Arctic charr live longer than dwarf Arctic charr, and the age at which the normal morphotype is usually captured is around 6 years, while most dwarf Arctic charr captured are around 3 years old (Johnson 1980).

Despite these general considerations, in several lakes Arctic charr are found with various morphological or behavioral differences, and authors have described as many as five sympatric stocks living in the same water body (Johnson 1980). Spawning times have been observed to be asynchronous, sometimes with one morphotype spawning in the fall while the other spawns in spring (Frost 1965), and the spawning locations have been found to vary from shallow water, 1-3 m, to water as deep as 20-30 m (Johnson 1980), with sites separated by more than 10 km (Hartley *et al*. 1992).

The origins of bimodality

The differentiation of stocks into bimodal populations is thought to be initiated in part by density dependent factors, where the density of older charr might allow or inhibit the maturation of juveniles (Johnson 1980). In some lakes a decrease in adult numbers does not affect the recruitment rate, so environmental factors must play an important role as well. Bimodality increases with latitude and increase in food supply (Griffiths, 1994 *fide in* Klemetsen *et al*. 2003), and niche segregation between stocks is often broken when there is high abundance of food (Jonsson and Jonsson 2001), so the most important environmental factors must be related to trophic competition. Morphologically specialized morphotypes feed more effectively than intermediate forms (Jonsson and Jonsson 2001), which by natural selection creates a best fitting population that is bimodal. A habitat shift from the profundal zone by dwarf Arctic charr as they grow is probably a tradeoff between feeding and predation risk (Klemetsen *et al*. 2003). Developing morphological differentiation is probably maintained over time by the fidelity of homing to the same spawning site, which it is known to be remarkable in Arctic charr (Svenning and Grotnes 1991). Differentiation is probably also maintained by behavioural patterns, since morphologically similar individuals attract each other (Klemetsen *et al*. 2003). Although spawning usually occurs within morphotypes, alternative male mating behavior can occur (Jonsson and Jonsson 2001), which might provoke a situation where an incomplete assortive mating could maintain a level of semi-speciation (Klemetsen *et al*. 2003), hence avoiding the complete separation of the morphotypes.

Genetic differentiation of Arctic charr populations

Several studies using genetic markers have defined different populations of Arctic charr. One of these used mitochondrial DNA markers to divide the whole Arctic charr complex into five homogeneous phylogeographic lineages: an Atlantic lineage including southeastern Canadian and European Arctic charr; an Arctic lineage for northern North America including western Greenland; a Beringian lineage for north Pacific Dolly Varden (*Salvelinus malma*); a Siberian lineage for northern Russia, and an Acadian group for landlocked populations found largely in southern Québec and Maine (Brunner *et al*. 2001). Subsequent analysis using microsatellites, more sensitive molecular markers, have further defined differentiation among populations within those lineages (Bernatchez *et al*. 1998; Bernatchez *et al*. 2002; Wilson *et al*. 2004).

The discussion of whether or not known sympatric Arctic charr morphotypes are genetically distinct is still open to debate. Occasionally it has been observed that hatchery raised charr from dwarf stocks grow well and become large (Eliassen *et al*. 1998). In general, however, descendants of dwarf and normal Arctic charr raised under the same conditions have different growth performance (Svedang 1990) and noticeable differences are found at the age of sexual maturity. The offspring of different morphotypes retain morphological and ecological characters such as mouth position and fin length (Nordeng 1993; Skualson *et al*. 1996; Svedang 1990). It has also been observed that higher proportions of the offspring of anadromous parents smoltified when released, in comparison to descendants of resident parents (Hindar 1986), so a genetic component is expected in the behavioral tendencies of individuals. On the other hand growth rate and age of sexual maturity also depend on environmental variables such as rate of food consumption and water temperature (Nordeng 1993; Svedang 1990), so the maximum size of any morphotype might be plastic depending on the environmental conditions.

Arctic charr are remarkably faithful in their time and place of spawning and it has been observed that spring spawners give raise to spring spawners and fall spawners to fall spawners (Johnson 1980), so initial phenotypic plasticity might have caused genetic differentiation. Perhaps one of the most extreme cases of stock separation is given in Lake Nashikinskoe, Russia, where the two morphotypes that are present have completely different niches and differ in morphology, ecology and even karyotype (Savvaitova 1995).

Genetic differentiation among sympatric morphotypes has been defined using microsatellites in a few cases (Westgaard *et al*. 2004), but in most cases it has been impossible to differentiate among morphotypes using traditional genetic markers (Danzmann *et al*. 1991; Hindar 1986; Magnusson and Ferguson 1987; Volpe and Ferguson 1996), likely due to the short isolation time (if any) of the morphotypes.

Since in general, the heritability of the physiological characters that differentiate the morphotypes is unclear, it is not possible to estimate the effective isolation of the morphotypes in terms that indicate the potential of one to replace the other in cases of selective fishing. Cross breeding of dwarf and normal morphotypes do generate fertile offspring, and given the interactive relationship of morphotypes in some locations, the exploitation of one of them could promote the expansion of the other. It is unknown whether the genetic differentiation among the morphotypes small enough to allow the transformation of one into another given favorable conditions. Questions also remain as to whether there are other genetic characters that make the differentiation strong enough to prevent cross repopulation. In systems where there has been a reduction of the normal Arctic charr population by selective fishing, the numbers of the dwarf morphotype have increased, but they have not repopulated the normal Arctic charr population

(Johnson & Campbell 1976 *fide in* Johnson 1980), so the recovery of the large form is likely dependent solely on its numbers.

Arctic charr's anadromous behavior makes it vulnerable to modern fishing techniques, but both lacustrine and anadromous stocks are affected by human activity and many populations are threatened (Maitland *et al*. 2006). An extreme case is Zugersee, Switzerland, where the annual catch dropped from 100,000 charr at the beginning of the 20th century to 5000 individuals in 1950. This was caused by heavy fishing but also by eutrophication, a condition that is not conducive to Arctic charr spawning (Jonsson and Jonsson 2001). Arctic charr is not only sensitive to eutrophication, but it is also to increased temperature (Johnson 1980), two factors that have increased dramatically in the last few decades due to human impact.

Arctic charr populations are endangered in several of the lakes that have been studied by researchers, particularly the ones that hold commercial fisheries, which has often decreased the stocks dramatically (Department of Fisheries and Oceans 1999; Department of Fisheries and Oceans 2001; Department of Fisheries and Oceans 2004; Gudbergsson 2004; Igoe *et al*. 2003). However, management strategies that might counteract these declines are hard to develop, since it is difficult to estimate what are the effective limits of Arctic charr populations. Although in some cases genetic markers have been able to differentiate populations demonstrating reproductive isolation (Westgaard *et al*. 2004), since the assays are based on DNA markers that do not respond to natural selection, it is not clear if the level of adaptation is enough for the different morphotypes to be treated as separate species. For these reasons, this thesis examines a molecular marker that represents the differentiation of groups according with the natural selection exerted by the environment.

The Major Histocompatibility Complex

The Major Histocompatibility complex (MHC), originally described as a collection of several genes in a relatively small area of the chromosome 6 in humans (HLA) and 17 in mice (H-2) (Dausset and Contu 1980), encodes 224 functional genes of which 63 are related with the immune function. MHC genes are present in all jawed vertebrates studied so far, and arose with the evolution of the adaptive immune system, 500 million years ago (Danchin *et al*. 2004). Contrary to what it has been observed in other vertebrates, teleosts have the different MHC receptors encoded by genes located in different chromosomes, such that they do not form a complex and are hence called simply "MH genes" (Stet *et al*. 2003).

MHC regions

In mammals, the MHC is divided into three well characterized regions. The Class I area is, the region that contains the three Class I α chain genes, called A, B and C. Class I α is the main subunit of the MHC Class I dimer and it is present on most nucleated cells. Its expression requires dimerizing with the second subunit, β-2- microglobulin, in the Rough Endoplasmic Reticulum (RER), followed by the binding of a peptide derived from the degradation of endogenous proteins by the proteasome (Steinmetz and Hood 1983). The Class I protein migrates to the cell surface and presents those peptides to CD8+ T-lymphocytes. During embryonic development, T-lymphocytes that bind to MHCs carrying self peptides are negatively selected, so the remaining T-cells only react against MHC containing non-self particles (Kasahara *et al*. 1995). Thus, upon binding of the MHC and the peptide, the T-lymphocyte will activate and induce the apoptosis of the cell containing the "unknown" particles, either through receptor

signals or by secreting chemicals that destroy the infected cell's membrane (Dausset and Contu 1980). The MHC Class I α protein binds 8-9 amino acid antigenic peptides in a groove that is formed between the first and the second domain of the molecule (Hemmer *et al*. 2000) (Figure 1-1). The Class I alpha molecule also possesses a third structural domain, which holds the binding domains above the membrane, as well as transmembrane and a cytoplasmic domain. Each domain is encoded by a separated exon in the MHC Class I α gene, and the second and third exon encode the area that binds the antigen, called the peptide binding region (PBR).

The second area of the MHC contains, among other genes, the three encoding Class II proteins. MHC class II receptors are also dimers, comprising two structurally very similar molecules, Class II α and Class II β. Both have a transmembrane, a cytoplasmic domain and a structural extracellular domain called α2 and β2 respectively (Steinmetz and Hood 1983). The first domain of both molecules (α1 and β1) are the ones that bind and present peptides to T-cells, but this time the groove is formed by one domain from both polypeptides and is open at each end, so the PBR can bind peptides 12-15 amino acids long (Hemmer *et al*. 2000).

Class II α and Class II β molecules are encoded by separated genes that are only expressed in antigen presenting cells (APC), such as macrophages, B-lymphocytes and dendritic cells. Each domain of the molecule is also encoded by a different exon and the second exon is the one that generates the first domain, which contains the PBR. In the class II region of the human MHC, there are three expressed Class II α genes paired with three expressed Class II β (Steinmetz and Hood 1983).

When the Class II dimer is initially produced in the RER, a conserved molecule, the MHC Class II associated invariant chain, blocks the groove, so the MHC Class II will not acquire RER

antigens, but will rather obtain them later in the endocytic pathway. The peptides that the MHC Class II binds are exogenous, usually obtained by phagocytosis, and they are digested in a lysosome that fuses with the MHC compartment, replacing the invariant chain with the exogenous peptide in the PBR. The MHC Class II presents these peptides to CD4+ T-lymphocytes. As in the case of the Class I molecule, CD4+ T-cells will only react with non-self peptides presented in the MHC groove. When they do, they start secreting cytokines that promote a variety of immune processes such as the activation of B-lymphocytes to produce specific antibodies (Kasahara *et al*. 1995).

A third area of the mammalian MHC, also called Class III, contains a variety of genes that encode several secreted proteins with immune function, including cytokines and components of the complement system (Steinmetz and Hood 1983).

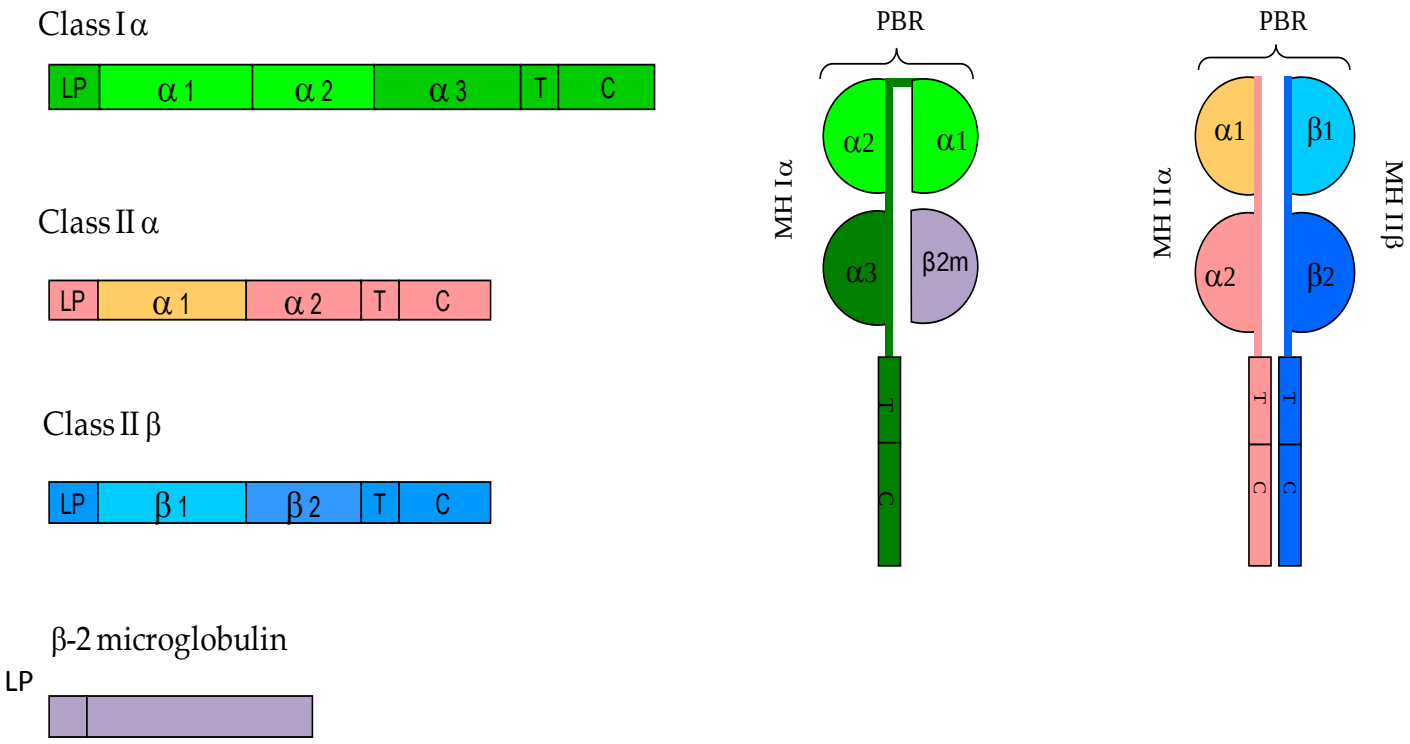


Figure 1-1. Major Histocompatibility receptors.

The general genomic structure of the four genes that encode MH receptors is shown in the left: Class I, Class II α, Class II β and β 2 m. Each block represents an exon, named according to the domain that it encodes in the corresponding MH receptor shown in the right: Leader peptide (LP); α/β 1, 2, 3; transmembrane (T) and cytoplasmic (C) domain.

MHC polymorphism

Although all MHC Class I and Class II molecules share the same basic structure, there is a lot of variation in the specific amino acid sequences, particularly in the domains encoding the PBRs. With samples from only a few thousands human beings analysed, there are 904 different known alleles of Class I MHC and 620 known alleles of MHC Class II. Therefore, although each individual human carries a maximum of six alleles of Class I and six of Class II, the possible combinations are so large that is virtually impossible that two individuals will share the same set of MHC alleles.

The divergence of the human Class I α genes, HLA-A, HLA-B and HLA-C, reaches 5.42%, 5%, and 3.64% respectively when considering the whole gene, and 9.05%, 8.88% and 5.99% if only the α1 and α2 domains are considered. Although most amino acid positions in the PBR are conserved, in fact, the most polymorphic amino acids are always part of the PBR (Reche and Reinherz 2003).

Human Class II α on the other hand, only has one polymorphic gene, HLA-DQA, with a 6.56% total divergence and 16.9% variation in the α1 domain. The other two genes, HLA-DPA and HLA-DRA have very low polymorphism, with divergences of 1.1% and 0.4% respectively. Also, HLA-DPA and HLA-DRA do not have any highly variable residues, while HLA-DQA has six in the α1 domain, although they do not contact the peptide (Reche and Reinherz 2003).

The three human Class II β genes, called HLA-DPB, HLA-DQB and HLA-DRB are highly polymorphic, which overall divergences are 3.4%, 5.27% and 4.15% respectively, and 8.7%, 14.4% and 11.96% in the β1 domain, which contain 8, 9 and 14 highly variable residues (Reche and Reinherz 2003). In contrast with what occurs in Class II α, most of the highly variable

residues correspond to amino acids that contact the peptide in Class II β, so it is inferred that in humans, most of the variability in the MHC Class II PBR is given by the divergence of the β subunit.

Consequently, most studies on MHC polymorphism have utilized Class II β. In the year 2000 there were 85 different allele sequences reported for chimpanzees, 21 for goats, 61 for domestic cats, among others (O'Brien and Yuhki 1999). Interestingly, of 251 human individuals examined in one study, 125 Class II β alleles were found, but the polymorphism of Class II β genes can be much higher in other species such as ocelots, where in only 35 individuals, 111 alleles were described (Bodmer *et al*. 1994). At the other extreme is the cheetah, where only 4 alleles were found in 6 individuals (O'Brien and Yuhki 1999).

Trans-species polymorphism

A very interesting characteristic of MHC polymorphism is that it usually predates the species that contains it (Klein 1987). This phenomenon, called trans-species polymorphism, implies that some of the alleles will be closer in sequence to an orthologous allele from another species, than to other alleles within the same species. The divergence of the alleles therefore passes though speciation events, and daughter species inherit per each MH loci, a set of alleles from their common ancestor (Klein 1987).

The trans-species polymorphism of MHC has been documented in most vertebrate species studied, so for example in domestic cat out of a divergence of 8% in Class I, 1% is species-specific and the rest of the divergence belongs to the whole Felidae family (O'Brien and Yuhki 1999). In Chinook salmon, all Class I alleles found belonged to 2 lineages that are common for

all Pacific salmon, with a common ancestor approximately15 million years ago (Garrigan and Hedrick 2001).

MHC genes and disease

Since MHC genes are directly involved in recognizing pathogenic peptides, its high variability is meant to increase the possibilities for a whole species to defend itself from a wider variety of pathogens, which are usually organisms undergoing constant change due to short generation times. Therefore, some specific MHC alleles are better than others at binding peptides from certain pathogens, and thus are related with the relative degree of disease resistance. In fact, the fitness difference given by one or another MHC allele has been observed to be as high as +0.5 (Lohm *et al*. 2002) while other alleles have been found related to disease susceptibility (Grimholt *et al*. 2003). Some relationships seen in humans are remarkable, like the presence of the haplotype HLA-DR15 or HLA-DQ1, which is correlated with autoimmune diseases and also with a faster progression of AIDS, a worse response to Hepatitis B vaccination and IgA deficiency (Dawkins *et al*. 1999; Vejbaesya *et al*. 1998). Naturally, the most remarkable evidence of specific MHC alleles related to disease resistance or susceptibility is found in non-human animals, since they can be challenged with specific pathogens and under controlled conditions. In homozygous mice, it was observed that those carrying H2K-bm8 as a unique allele would survive much higher doses of an Herpesvirus hominis type 1 infection, than homozygous mice carrying a H2K-B6 allele (Messaoudi *et al*. 2002). In wild passerine birds two particular alleles were significantly more common in the population, one of them related with resistance to malaria and the other related with resistance to *Plasmodium* (Bonneaud *et al*. 2006).

One of the most significant findings at this respect was seen in chicken, where the allele B21 conferred 7 times more resistance to Marek's disease. This trait is currently used widely in commercial chicken which are bred to carry the B21 allele (Kaufman *et al*. 1995).

Sources of MHC polymorphism

The high levels of MHC gene polymorphism could not be generated by accumulation of single point mutations, it is thought that the process must also involve recombination and gene conversion (Parham and Ohta 1996). MHC genes contain kilobase stretches (blocks) that have been shuffled recently to yield specific haplotypes (Dawkins *et al*. 1999). However, the blocks themselves have not changed much. Recombination occurs between the blocks but rarely within the blocks. Therefore, MHC alleles and even related loci, share stretches of DNA sequences and the resulting exons are a patchwork of motifs. Comparisons between closely related species have shown differentiation among MHC gene regions is mostly provided by deletion and insertions (Tsukamoto *et al*. 2005), demonstrating the high level of instability in the MHC region. Products of several duplication and deletion events, the MHC genes are usually present in multiple copies in the genome (Sato *et al*. 2001). One of the possible reasons for the high instability of the MHC region is that it contains a high density of retroelements, such as L1, or HERV and Alu. These retroelements could mediate recombinations taking the MHC fragments with them (Dawkins *et al*. 1999). Also, short repeated sequences might be playing a role as recombination hotspots, increasing the chance of generating polymorphism in the MHC genes. Short repeated sequences have been found in every second intron of MHC Class II β investigated so far (Schwaiger *et al*. 1994), and their basic structure, (GT)n or/and (GA)n, has remained in the same position for 100

million years in mammals. Interestingly, pseudogenes tend to degenerate these repeats over time (Schwaiger and Epplen 1995). More complex potential hotspots for recombination have been found in Chamois where the recombination rates of Class II β second exon exceeds the rates of point mutations by 10 times (Schaschl *et al*. 2005). In Chamois Class II β genes there are two segments of 98 and 87 bp respectively that are kept intact during recombination (Schaschl *et al*. 2005). Also, conserved repeated sequences have been found in closely related cichlid species (Klein *et al*. 1993a; Ono *et al*. 1993): twelve bp stretches that are differentially repeated, in different individuals are fully conserved, indicating a potential function as hotspots for recombination of Class II β exons.

Non-PBR polymorphism

Another source of MHC polymorphism aside from the PBR patterns is the fact that some MHC gene promoters also have high levels of divergence. In MHC Class II genes, this divergence is extensive enough to suggest selective pressure (Mitchison 1997). Expression of MHC Class II β is genetically correlated with parasite load in some species (Wegner *et al*. 2006), and in these cases a lower allelic diversity is compensated with a higher expression of the alleles that are present.

As previously mentioned, one of the most important sources of sequence polymorphism in MHC genes is the fact that these genes have been duplicated several times during evolution, which has caused the presence of multiple MHC gene copies in the genome of most vertebrates. The high number of gene copies permits higher loads of alleles per individual and also increases the total variability in the population. The multiple gene copies improve the capacity for adaptation,

which allows rapid colonization of new niches (Klein 1991). However, after the periods of expansion and settling in new environments, the pathogen-host relationships tends to become more stable, and many of the copies of MHC genes lose relevance for the presentation of antigens (Miller *et al*. 2002). Some of these extra MHC genes decay and finally convert into pseudogenes that do not express a functional molecule (Shiina *et al*. 2005). Other MHC genes modify their original function and become part of different biochemical pathways that might or might not be related to peptide presentation. These are called non-classical MHC genes, and they are generally non-polymorphic and have different patterns of expression than those of classical MHC genes (Shum *et al*. 1999). There are several pseudo-MHC genes and non-classical MHC genes in most vertebrate genomes. For example in humans DP, DQ and DR are used to present pathogenic peptides while mouse uses only the DQ gene primarily for this purpose, while the DP and DR loci contain mostly pseudogenes (O'Brien and Yuhki 1999). In rainbow trout (*Oncorhynchus mykiss*) there is only one Class I classical gene, however more than ten different non-classical genes have been described. Non-classical genes in the case of rainbow trout seem to play a major role in the polymorphism of the classical MHC, donating fragments that recombine inter-locus with the sole classical gene (Miller *et al*. 2006).

In teleosts, which have undergone several genome expansion and contraction events, the classical MH genes for Class I α and Class II α and Class II β are located on different chromosomes. This atypical arrangement probably occurred via chromosomal duplications that increased the number of MH genes which were initially located on the same chromosome, but differential gene usage led to the random evolution of some copies into non-classical genes or in pseudogenes (Stet *et al*. 2003), leaving the classical genes on separate chromosomes. The unique MH distribution on teleosts in which the different MH gene classes are located in different

linkage groups allows them to evolve and recombine more independently than they would if they were in a single complex, as in the case of mammals. This independence probably increases their potential for a wider range of their polymorphism (Stet *et al*. 2003).

Natural selection

According to Neutral Theory (Kimura 1995), most changes in the individual's DNA are selectively neutral, this is, they do not affect the fitness of the organism. However, some changes will increase or decrease the fitness of individuals, and some schools of thought consider that many mutations will be under selective pressure (Kreitman and Akashi 1995).

In classical population genetics, two allele models (usually denoted A and a) can produce three genotypes: AA, Aa and aa. Selection on one of the alleles occurs if the different genotypes provide differential fitness to the individuals that carry them. Differential fitness given by directional selection will show the pattern AA > Aa > aa, when the "A" allele is related with increased fitness, and the opposite tendency if the "a" allele is the one that increases fitness (Nielsen 2005). Directional selection can act in two fashions. On one hand if a new mutation increases fitness it will be under "positive selection", that is its frequency will tend to increase in the population, while a new mutation that decreases fitness will be under "negative" or "purifying selection" (Kreitman and Akashi 1995), that is its frequency will tend to decrease in the population, . In absence of selection, the substitution of the allele "A" by allele "a" is a neutral change and then in terms of fitness, each genotype will be equal, or in the example given above, AA=Aa=aa.

At the amino acid level, neutral selection happens in coding regions, when mutations that cause amino acid substitutions ("non-synonymous") occur at the same rate than those that do not change amino acids ("synonymous"). On the other hand when the mutation reduces fitness, i. e. under purifying selection, mutations at non-synonymous sites will be negatively selected, and therefore they will occur at a lower rate than synonymous mutations (Nielsen 2005). When the mutation increases fitness, non-synonymous substitutions will be positively selected and thus will occur at a higher frequency than synonymous ones.

When multiple alleles are maintained through positive selection, the rate of non-synonymous substitutions will surpass the synonymous substitution rate (Meyer and Thomson 2001). If the fitness of the multiple alleles are additive, then heterozygous individuals will be more fit (a condition called heterozygote advantage), and the pattern of the classical example will be AA<Aa>aa (Nielsen 2005). There are cases where several alleles that provide similar fitness are maintained in the population, kept through evolutionary time by heterozygote advantage. This evolutionary process called "balancing selection", keeps the multiplicity of alleles for much longer time than predicted under neutrality and therefore the coalescence time of the alleles often pre-dates speciation (Klein et al 1993b). Alleles under balancing selection tend to be present in similar proportions within the species, but the ratio can sometimes be skewed by the differential fitness of some alleles under specific circumstances or in specific populations (Hedrick 1999; Meyer and Thomson 2001).

MHC gene evolution has been largely described as a classical example of balancing selection where the existence multiple MHC alleles is maintained as a selected character (Potts and Slev 1995; Meyer and Thomson 2001). Balancing selection pressure on the variable MHC gene PBRs can be found in most vertebrates, sometimes reaching very high values as in the case of

Humboldt penguins (Kikkawa *et al*. 2005). Higher non-synonymous substitutions indicate selection for the polymorphism itself within the population rather than for a particular amino acid, presumably as a character that has historically facilitated the survival of the species.

MHC genes as population markers

Genomic areas with high variability such as the MHC can be used as population markers, since they will be more divergent in populations that are separated by longer time. Studies have shown a deeper separation in the differentiation of populations using MH genes than is detectable with microsatellites in some cases (Bernatchez and Landry 2003). One particular feature of using the MHC as a population marker is that its polymorphism occurs in the coding region and thus the particular alleles of each individual within the population are selected by nature. This allows the use of MHC polymorphism to define conservation units, populations with unique characteristics that make them an integral and indispensable part of a certain ecosystem. Named by the National Oceanographic and Atmospheric Administration (NOAA) as evolutionary significant units (ESU), they are supposed to receive independent management strategies (Miller *et al*. 2001). MHC is a marker that gives information not only on the relatedness of individuals but also on their interaction with the environment, because many of the interactions will force individuals to deal with different, and possibly new, pathogens. Since pathogen behavior and distribution are relatively specific for particular ecosystems, a settled population will tend to present a consistent MHC allelic pattern. Consequently, MHC gene differentiation has been used in cases where the isolation of populations has occurred very recently, as in the case of two populations of sheep introduced to St. Kilda Island in 1932, that presented distinguishable patterns in their MHC due

differences in the parasite fauna parasitic present on the different sides of the island (Charbonnel and Pemberton 2005).

MH gene polymorphism on fish populations

MH polymorphism is high among all studied fish species. In Atlantic salmon (*Salmo salar*) the level of divergence found in the first domain of MH Class II β reached 6%, similar to the variation found in microsatellites (Landry and Bernatchez 2001). In sockeye salmon (*Oncorhynchus nerka*) Class II β genes, the β1 domain had an 8.5% average variation reported (Miller *et al*. 2001). In rainbow trout, the β 1 domain had a maximum variation of 22% (Dorschner *et al*. 2000) and the divergence in striped bass (*Morone saxatilis*) was 27% for Class II α and 32% for Class II β in the first domains (Hardee *et al*. 1995). Similar degrees of divergence can be seen in the MH class I gene; in Atlantic salmon for example the maximum divergence was 18% for Class I α gene (Grimholt *et al*. 2002).

Moreover in salmonids, specific MH alleles have been found associated with resistance or susceptibility to certain diseases. In Atlantic salmon, fish carrying specific MH alleles were significantly more resistant to ISAV, a relationship found for 11 alleles for Class I α, and 7 for Class II α, in some cases decreasing the mortality of families carrying these alleles more than 20% (Grimholt *et al*. 2003). In similar challenges of Atlantic salmon against ISAV, 2 alleles of Class I α were found that increased survival more than 10% (Kjoglum *et al*. 2006). Also, synergic effects on disease resistance were found in certain combinations of two Class I α alleles as well as two Class II α alleles. One allele of Class I α and one of Class II α were directly related to susceptibility to the virus, while synergistic combinations of alleles were also reported to cause susceptibility (Kjoglum *et al*. 2006). Also in Atlantic salmon, resistance to *Aeromonas salmonicida* was found to be associated with specific MH Class II β alleles (Langefors *et al*. 2001), and in Chinook salmon, it was found that specific alleles as well as the degree of MH

Class II β heterozygosity were related to the resistance to Infectious Hematopoietic Necrosis Virus (Arkush *et al*. 2002).

The relationship of high MH polymorphism to disease resistance has allowed successful differentiation of fish populations even when they inhabit similar geographical areas. For example, a comparison of anadromous and land-locked Atlantic salmon from adjacent areas using MH Class II β polymorphism yielded significant divergence, with higher heterozygosity and sequence divergence levels than was detected using microsatellites (Landry and Bernatchez 2001). Four sympatric barbel morphotypes of Lake Tana in Ethiopia were successfully differentiated using MH Class II β (Dixon *et al*. 1996), and two striped bass populations from the Roanoke River, North Carolina, were isolated according with MH Class II α polymorphism (Hardee *et al*. 1995). Chinook salmon populations from the upper and interior Fraser River in British Columbia were differentiated using MH Class II β and microsatellites. The levels of differentiation were four times higher in the MH genes, and the intrapopulation diversity was much lower than for the microsatellites, showing strong selection for MH adaptation far beyond basal genetic drift (Miller *et al*. 2001). In steelhead populations along the California coast the MH Class II β differentiation was also higher than in microsatellites, and when the populations were separated in three geographic regions the MH loci was the best marker of population differentiation (Aguilar and Garza 2006).

Using MH gene polymorphism to distinguish Arctic charr stocks

Since diet and ecological specializations are often long lasting in Arctic charr, fish from different stocks are infected with defined parasite species. For example, 73% of normal Arctic charr in

Lake Fjellfrosvatn, Norway, contained 5-8 parasite species, while 72% of dwarf Arctic charr contained only one species (Knudsen *et al*. 1997). Trophic segregation between pelagic and benthic feeders caused the Normal to be infected by copepod-transmitted parasites and the dwarf to be infected by parasites transmitted by amphipods (Knudsen *et al*. 1997). Moreover it is expected that fish feeding in different geographic locations confront different pathogen species (Knudsen 1995) and infection with different species of pathogens should select a different specific set of MH alleles for each stock, which in turn would allow the detection of isolation of the populations (if it exists). A differentiation based on MH alleles would also indicate that the different stocks are adapted to their specific environment, in a character that is vital for their survival, as it is key to the defense against pathogens.

This thesis will examine the potential use of MH gene polymorphisms to distinguish, on the one hand, populations separated geographically from the global range of Arctic charr, and on the other, to distinguish among sympatric morphotypes inhabiting the same lakes, the separation of which may have occurred only after the retreat of the last glaciation.

References

Aguilar, A. and Garza, J. C.: A comparison of variability and population structure for major histocompatibility complex and microsatellite loci in California coastal steelhead (*Oncorhynchus mykiss* Walbaum). Molecular Ecology 15: 923-37, 2006

Arkush, K., Giese, A., Mendonca, H., McBride, A., Marty, G., and Hedrick, P.: Resistance to three pathogens in the endangered winter-run chinook salmon (*Oncorhynchus tshawytscha*): effects of inbreeding and major histocompatibility complex genotypes. Canadian Journal of Fisheries and Aquatic Sciences 59: 966-975, 2002

Bernatchez, L., Dempson, B., and Martin, S.: Microsatellite gene diversity analysis in anadromous arctic char, *Salvelinus alpinus*, from Labrador, Canada. Canadian Journal of Fisheries and Aquatic Sciences 55: 1264-1272, 1998

Bernatchez, L. and Landry, C.: MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? Journal of Evolutionary Biology 16: 363-77, 2003

Bernatchez, L., Rhydderch J, and F, Kircheis.: Microsatellite gene diversity analysis in landlocked Arctic Charr from Maine. Transactions of the American Fisheries Society 131: 1106-1118, 2002

Bodmer, J. G., Marsh, S. G., Albert, E. D., Bodmer, W. F., Dupont, B., Erlich, H. A., Mach, B., Mayr, W. R., Parham, P., Sasazuki, T., and *et al*.: Nomenclature for factors of the HLA system, 1994. Human Immunology 41: 1-20, 1994

Bonneaud, C., Chastel, O., Federici, P., Westerdahl, H., and Sorci, G.: Complex Mhc-based mate choice in a wild passerine. Proceedings of the Royal Society of London. Series B. 273: 1111-6, 2006

Brunner, P. C., Douglas, M. R., Osinov, A., Wilson, C. C., and Bernatchez, L.: Holarctic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. Evolution 55: 573-86, 2001

Charbonnel, N. and Pemberton, J.: A long-term genetic survey of an ungulate population reveals balancing selection acting on MHC through spatial and temporal fluctuations in selection. Heredity 95: 377-88, 2005

Danchin, E., Vitiello, V., Vienne, A., Richard, O., Gouret, P., McDermott, M. F., and Pontarotti, P.: The major histocompatibility complex origin. Immunological Reviews 198: 216-32, 2004

Danzmann, R., Ferguson, M., Skualson, S., Snorranson, S., and Noakes, D.: Mitochondrial DNA diversity among four sympatric morphotypes of Arctic charr, *Salvelinus alpinus* L, from Thingvallavatn, Iceland. Journal of Fish Biology 39: 649-659, 1991

Dausset, J. and Contu, L.: Is the MHC a general self-recognition system playing a major unifying role in an organism? Human Immunology 1: 5-17, 1980

Dawkins, R., Leelayuwat, C., Gaudieri, S., Tay, G., Hui, J., Cattley, S., Martinez, P., and Kulski, J.: Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. Immunological Reviews 167: 275-304, 1999

Department of Fisheries and Oceans, C.: Central and Arctic regions - Status report - Hornaday River Arctic charr. 1999

Department of Fisheries and Oceans, C.: Newfounland region - Stock status report - North Labrador Arctic Charr. 2001

Department of Fisheries and Oceans, C.: Central and Arctic - Stock status report - Cambridge bay Arctic charr. 2004

Dixon, B., Nagelkerke, L. A., Sibbing, F. A., Egberts, E., and Stet, R. J.: Evolution of MHC class II beta chain-encoding genes in the Lake Tana barbel species flock (*Barbus intermedius* complex). Immunogenetics 44: 419-31, 1996

Dorschner, M. O. D., T. Bronte, C. R. Burnham Curtis, M. K. Phillips, R. B.: High levels of MHC class II allelic diversity in lake trout from Lake Superior. Journal of Heredity 91: 359-63, 2000

Eliassen, R., Johnsen, H., Mayer, I., and Jobling, M.: Contrasts in osmoregulatory capacity of two Arctic charr, *Salvelinus alpinus* (L.), strains from northern Norway. Aquaculture 168: 255-269, 1998

Frost, W.: Breeding Habits of Windermere Charr, *Salvelinus Willughbii* (Gunther), and their Bearing on Speciation of these Fish. Proceedings of the Royal Society of London. Series B, Biological Sciences 163: 232-284, 1965

Garrigan, D. and Hedrick, P. W.: Class I MHC polymorphism and evolution in endangered California Chinook and other Pacific salmon. Immunogenetics 53: 483-9, 2001

Grimholt, U., Drablos, F., Jorgensen, S. M., Hoyheim, B., and Stet, R. J.: The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. Immunogenetics 54: 570-81, 2002

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebo, S., and Stet, R. J.: MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. Immunogenetics 55: 210-9, 2003

Gudbergsson, G.: Arctic charr in Lake Myvatn: The centennial catch record in the light of recent stock estimates. Aquatic Ecology 38: 271-284, 2004

Hardee, J. J., Godwin, U., Benedetto, R., and McConnell, T. J.: Major histocompatibility complex class II A gene polymorphism in the striped bass. Immunogenetics 41: 229-38, 1995

Hartley, S., Mcgowan, C, Greer R, Walker, A.: The genetics of sympatric Arctic charr (Salvelinus alpinus (L.)) populations from Loch Rannoch, Scotland. Journal of Fish Biology 41: 1021-1031, 1992

Hedrick, P.: Balancing selection an MHC. Genetica 104: 207-214, 1999

Hemmer, B., Kondo, T., Gran, B., Pinilla, C., Cortese, I., Pascal, J., Tzou, A., McFarland, H. F., Houghten, R., and Martin, R.: Minimal peptide length requirements for CD4(+) T cell clones--implications for molecular mimicry and T cell survival. International Immunology 12: 375-83, 2000

Hindar, K.: Genetic differenciation among local populations and morphotypes of Arctic charr, *Salvelinus alpinus*. Biologicl journal of the Linnean society 27: 269-285, 1986

Igoe, F., O'Grady, D., Tierney, D., and Fitzmaurice, P.: Arctic char *Salvelinus alpinus* (L.) in Ireland - A millennium review of its distribution and status with conservation recommendations. Biology and Environment: Proceedings of the Royal Irish Academy 103 B: 9-22, 2003

Johnson, L.: The Arctic charr, *Salvelinus alpinus*. In E. K. Balon (ed.): Charrs: salmonid fishes of the genus Salvelinus, pp. 19-98, Dr. W. Junk, The Hague, 1980

Jonsson, B. and Jonsson, N.: Polymorphism and speciation in Arctic charr. Journal of Fish Biology 58: 605-638, 2001

Maitland, P., Winfield, I., McCarthy, I., and Igoe, F.: The status of Arctic charr Salvelinus alpinus in Britain and Ireland. Ecology of Freshwater Fish 16: 6-19, 2006

Kasahara, M., Flajnik, M. F., Ishibashi, T., and Natori, T.: Evolution of the major histocompatibility complex: a current overview. Transplant Immunology 3: 1-20, 1995

Kaufman, J., Volk, H., and Wallny, H. J.: A "minimal essential Mhc" and an "unrecognized Mhc": two extremes in selection for polymorphism. Immunological Reviews 143: 63-88, 1995

Kikkawa, E. F., Tsuda, T. T., Naruse, T. K., Sumiyama, D., Fukuda, M., Kurita, M., Murata, K., Wilson, R. P., LeMaho, Y., Tsuda, M., Kulski, J. K., and Inoko, H.: Analysis of the sequence variations in the Mhc DRB1-like gene of the endangered Humboldt penguin (Spheniscus humboldti). Immunogenetics 57: 99-107, 2005

Kimura, M.: Limitations of Darwinian selection in a finite population. Proceedings of the National Academy of Science 92: 2343-2344, 1995

Kjoglum, S., Larsen, S., Bakke, H. G., and Grimholt, U.: How specific MHC class I and class II combinations affect disease resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*). Fish and Shellfish Immunology 21: 431-41, 2006

Klein, D., Ono, H., O'HUigin, C., Vincek, V., Goldschmidt, T., and Klein, J.: Extensive MHC variability in cichlid fishes of Lake Malawi. Nature 364: 330-4, 1993

Klein, J.: Of HLA, tryps, and selection: an essay on coevolution of MHC and parasites. Human Immunology 30: 247-58, 1991

Klein, J., Satta, Y., and O'hUigin, C.: The molecular descent of the Major Histocompatibility Complex. Annual reviews in Immunology 11: 269-295, 1993b

Klemetsen, A., Amundsen, P., Dempson, B., Jonsson, B., Jonsson, N., O'Connell, F., and Mortensen, E.: Atlantic salmon Salmo salar L., brown trout *Salmo trutta* L. and Arctic charr *Salvelinus alpinus* (L.): a review of aspects of their life histories. Ecology of Freshwater Fish 12: 1-59, 2003

Knudsen, R.: Relationships between Habitat, Prey Selection and Parasite Infection in Arctic Charr (*Salvelinus alpinus*). Nordic Journal of Freshwater Research 71: 333-344, 1995

Knudsen, R., Klemetsen, A., Amundsen, P. A., and Hermansen, B.: Incipient speciation through niche expansion: an example from the Arctic charr in a subarctic lake. Proceedings of the Royal Society of London. Series B. 273: 2291-8, 2006

Knudsen, R., Kristoffersen, R., and Amundsen, P. A.: Parasite communities in two sympatric morphotypes of Arctic charr, Salvelinus alpinus (L.), in northern Norway. Canadian Journal of Zoology 75: 2003-2009, 1997

Kreitman, M. and Akashi, H.: Molecular evidence for natural selection. Annual Reviews in Ecology and Systematic 26: 403-422, 1995

Landry, C. and Bernatchez, L.: Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*). Molecular Ecology 10: 2525-39, 2001

Langefors, A., Lohm, J., Grahn, M., Andersen, O., and von Schantz, T.: Association between major histocompatibility complex class IIB alleles and resistance to *Aeromonas salmonicida* in Atlantic salmon. Proceedings of the Royal Society of London. Series B. 268: 479-85, 2001

Lavin, P. and McPhail, J.: Adaptive divergence of trophic phenotype among freshwater populations of the threespine stickleback (*Gasterosteus aculeatus*). Canadian Journal of Fisheries and Aquatic Sciences 43: 2455-2463, 1986

Lohm, J., Grahn, M., Langefors, A., Andersen, O., Storset, A., and von Schantz, T.: Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. Proceedings of the Royal Society of London. Series B. 269: 2029-33, 2002

Magnusson, K. and Ferguson, M.: Genetic analysis of four syrnpatric morphotypes of Arctic charr, *Salvelinus alp*inus, from Thingvallavatn, Iceland. Environmental Biology of Fishes 20: 67-73, 1987

Messaoudi, I., Guevara Patino, J. A., Dyall, R., LeMaoult, J., and Nikolich-Zugich, J.: Direct link between mhc polymorphism, T cell avidity, and diversity in immune defense. Science 298: 1797-800, 2002

Meyer, D. and Thomson, G.: How selection shapes variation of the human major histocompatibility complex: a review. Annals of Human Genetics 65: 1-26, 2001

Miller, K. M., Kaukinen, K. H., Beacham, T. D., and Withler, R. E.: Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. Genetica 111: 237-57, 2001

Miller, K. M., Kaukinen, K. H., and Schulze, A. D.: Expansion and contraction of major histocompatibility complex genes: a teleostean example. Immunogenetics 53: 941-63, 2002

Miller, K. M., Li, S., Ming, T. J., Kaukinen, K. H., and Schulze, A. D.: The salmonid MHC class I: more ancient loci uncovered. Immunogenetics 58: 571-89, 2006

Mitchison, A.: Partitioning of genetic variation between regulatory and coding gene segments: the predominance of software variation in genes encoding introvert proteins. Immunogenetics 46: 46-52, 1997

Nordeng, H.: Solution to the "charr problem" based on Arctic charr (Salvelinus alpinus) in Norway. Canadian Journal of Fishery and Aquatic Science 40: 1372-1387, 1993

Nielsen, R.: Molecular signatures of natural selection. Annual reviews in Genetics 39: 197-218, 2005

O'Brien, S. J. and Yuhki, N.: Comparative genome organization of the major histocompatibility complex: lessons from the Felidae. Immunological Reviews 167: 133-44, 1999

Ono, H., O'HUigin, C., Tichy, H., and Klein, J.: Major-histocompatibility-complex variation in two species of cichlid fishes from Lake Malawi. Molecular Biology and Evolution 10: 1060-72, 1993

Parham, P. and Ohta, T.: Population biology of antigen presentation by MHC class I molecules. Science 272: 67-74, 1996

Parker, H. and Johnson, L.: Population structure, ecological segregation and reproduction in non-anadromous Arctic charr, Salvelinus alpinus (L), in four unexploited lakes in the Canadian high Arctic. Journal of Fish Biology 38: 123-147, 1991

Potts, W. K. and Slev, P. R.: Pathogen-based models favoring MHC genetic diversity. Immunological Reviews 143: 181-97, 1995

Reche, P. A. and Reinherz, E. L.: Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. Journal of Molecular Biology 331: 623-41, 2003

Sato, A., Mayer, W. E., Tichy, H., Grant, P. R., Grant, B. R., and Klein, J.: Evolution of Mhc class II B genes in Darwin's finches and their closest relatives: birth of a new gene. Immunogenetics 53: 792-801, 2001

Savvaitova, K.: Arctic Charrs. The structure of population systems and prospect for economic use. In N. R. Council (ed.), Canada Institute for Scientific and Technical Information, 1993

Savvaitova, K.: Patterns of diversity and processes of speciation in Arctic Charr. Nordic Journal of Freshwater Research 71: 81-91, 1995

Schaschl, H., Suchentrunk, F., Hammer, S., and Goodman, S. J.: Recombination and the origin of sequence diversity in the DRB MHC class II locus in chamois (*Rupicapra spp.*). Immunogenetics 57: 108-15, 2005

Schwaiger, F. W. and Epplen, J. T.: Exonic MHC-DRB polymorphisms and intronic simple repeat sequences: Janus' faces of DNA sequence evolution. Immunological Reviews 143: 199-224, 1995

Schwaiger, F. W., Weyers, E., Buitkamp, J., Ede, A. J., Crawford, A., and Epplen, J. T.: Interdependent MHC-DRB exon-plus-intron evolution in artiodactyls. Molecular Biology and Evolution 11: 239-49, 1994

Shiina, T., Dijkstra, J. M., Shimizu, S., Watanabe, A., Yanagiya, K., Kiryu, I., Fujiwara, A., Nishida-Umehara, C., Kaba, Y., Hirono, I., Yoshiura, Y., Aoki, T., Inoko, H., Kulski, J. K., and Ototake, M.: Interchromosomal duplication of major histocompatibility complex class I regions in rainbow trout (*Oncorhynchus mykiss*), a species with a presumably recent tetraploid ancestry. Immunogenetics 56: 878-93, 2005

Shum, B. P., Rajalingam, R., Magor, K. E., Azumi, K., Carr, W. H., Dixon, B., Stet, R. J., Adkison, M. A., Hedrick, R. P., and Parham, P.: A divergent non-classical class I gene conserved in salmonids. Immunogenetics 49: 479-90, 1999

Skreslet, S.: The ecosystem of the arctic lake nordlaguna, jan mayen island: III. Ecology of arctic char, *Salvelinus alpinus* (L.) ASTARTE 6: 43-54, 1973

Skulason, S., Snorrason, S., Noakes, D., and Ferguson, M.: Genetic basis of life history variations among sympatric morphotypes of Arctic char, *Salvelinus alpinus*. Canadian Journal of Fisheries and Aquatic sciences 53: 1807-1813, 1996

Steinmetz, M. and Hood, L.: Genes of the major histocompatibility complex in mouse and man. Science 222: 727-33, 1983

Stet, R. J., Kruiswijk, C. P., and Dixon, B.: Major histocompatibility lineages and immune gene function in teleost fishes: the road not taken. Critical Reviews in Immunology 23: 441-71, 2003

Svedang, H.: Genetic basis of life-history variation of dwarf and normal Arctic charr, *Salvelinus alpinus* (L.), in Stora Rosjon, central Sweden. Journal of Fish Biology 36: 917-932, 1990

Svendsen, J. and Mangerud, J.: Paleoclimatic inferences from glacial fluctuations on Svalbard during the last 20000 years. Climate Dynamics 6: 213-220, 1992

Svenning, M. and Grotnes, P.: Stationarity and homing ability of landlocked Arctic charr. Nordic Journal of Freshwater Research 66: 36-43, 1991

Tsukamoto, K., Hayashi, S., Matsuo, M. Y., Nonaka, M. I., Kondo, M., Shima, A., Asakawa, S., Shimizu, N., and Nonaka, M.: Unprecedented intraspecific diversity of the MHC class I region of a teleost medaka, Oryzias latipes. Immunogenetics 57: 420-31, 2005

Vejbaesya, S., Eiermann, T. H., Suthipinititharm, P., Bancha, C., Stephens, H. A., Luangtrakool, K., and Chandanayingyong, D.: Serological and molecular analysis of HLA class I and II alleles in Thai patients with psoriasis vulgaris. Tissue Antigens 52: 389-92, 1998

Volpe, J. P. and Ferguson, M. M.: Molecular genetic examination of the polymorphic Arctic charr *Salvelinus alpinus* of Thingvallavatn, Iceland. Molecular Ecology 5: 763-72, 1996

Wegner, K. M., Kalbe, M., Rauch, G., Kurtz, J., Schaschl, H., and Reusch, T. B.: Genetic variation in MHC class II expression and interactions with MHC sequence polymorphism in three-spined sticklebacks. Molecular Ecology 15: 1153-64, 2006

Westgaard, J., Klemetsen, A., and Knudsen, R.: Genetic differences between two sympatric morphotypes of Arctic charr confirmed by microsatellite DNA. Journal of Fish Biology 65: 1185-1191, 2004

Wilson, A. J., Gislason, D., Skulason, S., Snorrason, S. S., Adams, C. E., Alexander, G., Danzmann, R. G., and Ferguson, M. M.: Population genetic structure of Arctic charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales. Molecular Ecology 13: 1129-42, 2004

Chapter 2. Cloning and characterization of Arctic charr (*Salvelinus alpinus*) Major Histocompatibility receptor genes

Brief communication

Arctic charr (*Salvelinus alpinus*) is the most northerly distributed salmonid and its habitat encircles the north polar circle (Johnson 1980). Arctic charr is the main fish species in Arctic post glacial lakes and it inhabits areas with low nutrient load (Klemetsen *et al*. 2003) that were covered by glacial ice until only 10000-15000 years ago (Bernatchez *et al*. 2002). These fish have successfully colonized these marginal environments due to their extraordinary phenotypic plasticity. Different populations of Arctic charr show distinct morphology, to the point that the first taxonomic studies defined 56 different species around the world (Savvaitova 1995). This was corrected later by genetic studies that showed that the different morphologies corresponded to just one species now called the Arctic charr complex (Brunner *et al*. 2001). The phenotypic plasticity of the Arctic charr, which allows it to exploit different ecological niches within the same water body, has generated in several postglacial lakes at least two distinct morphotypes with markedly different adult sizes and shapes, asynchronous reproductive seasons, and different feeding behaviors (Jonsson and Jonsson 2001). Given the short time in evolutionary terms that has elapsed since the Arctic charr started postglacial colonization, the separation among morphotypes or populations cannot always be detected by genetic markers (Volpe and Ferguson 1996).

Human impacts are progressively damaging the Arctic charr ecosystem, and fisheries are often overexploited. However, given Arctic charr's high levels of phenotypic polymorphism, the distinction between morphotypes or populations are obscure, and current management strategies lack a reliable method to evaluate the effective impact of fishing (Jonsson and Jonsson 2001).

With the aim of developing a technique to differentiate the Arctic charr morphotypes and populations, we are studying its Major Histocompatibility genes (MH). The MH receptors present peptides produced by protein degradation to T-lymphocytes, which detect the peptide. If the peptide is not derived from self proteins they will start immune reactions that will eliminate the pathogen that produced the peptide (Steinmetz and Hood 1983). The MH genes are thus the molecules responsible for the distinction between self and non-self, and are the first step in pathogen recognition and the immune reaction. In order to bind and present the enormous variety of peptides produced by a very large number of constantly evolving pathogens, the MH receptors' peptide binding regions (PBR) are encoded by the most polymorphic nuclear genes known. MH gene polymorphism is maintained at the population level, so individuals that carry alleles that better match with pathogens of an ecosystem, will have better survival chances (Dawkins *et al*. 1999; Lohm *et al*. 2002; Messaoudi *et al*. 2002). Therefore, a population that is adapted to its niche will contain individuals that have niches specific common patterns in their MH genes (Bernatchez and Landry 2003; Bonneaud *et al*. 2006; Lamont 1998). Given these characteristics, MH genes can be used as molecular markers that show not only the genetic line of the individuals, but also the level of adaptation to a particular niche. Therefore they are ideal indicators of the separation of populations isolated for short periods of time, but that nevertheless have adapted to their ecosystem.

This report describes the four full length cDNA sequences of the two MH receptor classes in Arctic charr. The first is the MH Class I receptor, which is involved in the presentation of intracellular peptides to CD8+ T-lymphocytes. It comprises two subunits encoded by different genes: MH Class I α and β 2 microglobulin (β 2 m), and it is expressed on most somatic cells. Second is the MH Class II receptor, which is involved in the presentation of extracellular

peptides to CD4+ T-lymphocytes. It is also formed by two subunits: MH Class II α and MH Class II β, and is expressed only on specialized immune cells called antigen presenting cells (APC). In the case of the Class I receptor, the PBR is contained only in the α subunit. In the case of Class II, the PBR is made by the joining of both the α and β subunits (Steinmetz and Hood 1983). Consequently, polymorphism studies are usually performed on the Class I alpha, Class II α and Class II β molecules, which our laboratory is currently carrying out for Arctic charr populations (Chapters 3-5). This report describes the general characterization of the four Arctic charr MH genes, especially β 2 microglobulin, which is not described in those publications. The β 2 m molecule binds non-covalently to the MH Class I molecule, and it is essential for the proper folding and cell surface display of the MH Class I receptor (Vitiello *et al*. 1990). Contrary to what has been observed in other vertebrates, in rainbow trout the β2m molecule is encoded by multiple genes (Magor *et al*. 2004). However, sequence divergence among the genes generally occurs only in areas that do not affect the mature protein, so the β 2 m molecule is highly conserved, even between different salmonid species (Shum *et al*. 1996).

In order to obtain the MH genes full length sequences, first, RNA was extracted from Arctic charr tissues using Trizol (Invitrogen, CA) according to manufacturer instructions. For the RT-PCRs, 1 µg of RNA was transcribed into cDNA by Oligo-dT priming using the Advantage RT-for-PCR kit (Clontech, CA). The RNA used for the tissue distribution experiments was extracted from the brain, gill, head kidney, heart, intestine, liver, lateral muscle, posterior kidney and spleen of individuals obtained from the University of Guelph Alma Research Station, Alma, Ontario, Canada. Six µg of RNA were used in the β 2 m Northern blots, performed according to the procedure described by (Fujiki *et al*. 2003) using a Dig-labeled probe (Roche, Germany). The Northern blot probe covered the region between the 3rd and 331st nucleotide of the β 2 m cDNA

and it was amplified using the primers Fwd: GCA CAG GCA GAC GGA TCT TTC and Rev: CAC CAA GAG TGT TGG ATT CAC AC. The same primers were used to amplify a fragment from Arctic charr genomic DNA, that was used to synthesize a DIG-labeled probe (Roche, Germany) for the β 2 m Southern blot. The β 2 m Southern blot procedure was done following the procedure of (Fujiki *et al*. 2003), using the enzymes *Bam* HI, *Hind* III, *Eco* RV, *Eco* RI, *Pst* I and *Taq* I (Promega, WI), independently to digest 5 µg of Arctic charr genomic DNA. Each digestion product was purified and loaded on an agarose gel for subsequent transfer to a nylon membrane, followed by probing.

The β 2 m western blots were performed following the procedure described by (Kales *et al*. 2006), running 100 µg of total protein from the different tissues in 15% SDS-PAGE to be transferred to nitrocellulose membranes. The β 2 m contained in the Arctic charr tissues was detected using antibodies raised against *Oncorhynchus mykiss* β 2 m. The western blot was developed by chemiluminescence reaction against the peroxidase-labeled secondary antibodies using ECL Plus (Amersham, NJ).

The RNA used to obtain the full length cDNA sequences was extracted from Arctic charr spleen. Primers were designed to match conserved areas of the cDNAs of other salmonids (Class I: AAG25206; AAG25205; BAA09553; AAN75116; AAN75113; AAG25197; AAG25199. β 2 m: AAB04654; AAG17525. Class II α: AAL40122; CAB96452; CAB96451; CAB96450. Class II β: X70167; X70166; U20944; U20943). For Class I α they amplified a section between the α 3 domain and the cytoplasmic domain (Fwd: CCS TCA GTG TCT CTG CTC CAG and Rev: GAG TTS TCA GAG TCA GTG TCG G). For Class II α and β the primers amplified a fragment that encodes the α 2 or β 2 domain and the transmembrane domain (Class II α: Fwd: GTG GAC CAG GAA CAA TCA GAA TG and Rev: GCT GGC TCA CCT CAG GTT C; Class II β: Fwd:

CAG AGT GAC CTG GCT GAG and Rev: CKA TCT TAT TCC TCT CAG MCT CAG). For β 2 m, the same primers used to amplify the Southern blot probe were used to obtain a 328 bp cDNA amplicon.

The primers yielded fragments of 446 bp, 193 bp and 216 bp for Class I α, Class II α and Class II β respectively. The amplicons were cloned in pGEM-T easy vectors (Promega, WI) according to manufacturer instructions and transformed in XL-1 bacteria made competent by the Inoue procedure (Sambrook *et al*. 2001). A bacterial colony was grown in LB media and the containing plasmid extracted using the GenElute Plasmid Miniprep Kit (Sigma-Aldrich, MA). The clones were sequenced using T7 and Sp6 primers with a big dye terminator v3.1 (Applied Biosystems, Foster City, CA) using a 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA). Sequencing of the conserved areas confirmed that they were the correct regions of the Class I α, Class II α, Class II β and β 2 m genes respectively. These sequences were used to design primers to pair with RACE-PCR primers that anchored the beginning and the end of each respective cDNA. RACE-PCR was performed according to manufacturer instruction using the Smart RACE cDNA amplification kit (Clontech, CA).

The forward primers paired to the RACE primers that allowed us to obtain the 3' end of the cDNAs were GGT CTC ATC ATT GGA GGG GTG ATA GC for Class I α, GTG GGT CTG ACT CTG GGG CTG CTG GGA G for Class II α, CTG ATG TGA CCT CCA CTG AGG AGC TG for Class II β and CAG AGT CCG CCA CCT GAA GAA CCT G for β 2 m. The reverse primers that allowed us to obtain the 5' ends were CTG TTC TCG CCC GTC TTT CTG CCA G for Class I α, GTG CAG GGT GGA ACC CAC TGA CGT GGC AG for Class II α, GCA CAT CAG CAT GGC AGG GTG TCT G for Class II β, and GTT CTT CAG GAG CTG GAT GCT GAT GTC for β 2 m. To obtain the 5' and 3' end of Class I α cDNA, it was necessary to perform

a nested PCR over the first RACE-PCR product using the primers GTC GCG TGG CAG GTC ACT GGA G for the 5' end and CCG GCC AGC ACT TCC GAC ACG G for the 3'end. Nested PCR was also necessary to obtain the 3' end of Class II β using the primer CCC GTC CCT GCC TGA GTC TGA GAG. The RACE-PCR yielded different band sizes, but only the brightest bands in the gel were used in the subsequent experiments. For Class I α the 5' end RACE-PCR product was 0.7 kb and the 3' end was 1.5 kb approximately. For Class II α the products were around 0.7 kb and 0.9 kb, for 3' and 5' respectively. For Class II β were around 0.5 kb and 0.7 kb, for 3' and 5' respectively. β 2 m 5' size was around 0.4 kb and the 3' end size was 0.9 kb.

The fragments were cloned, and transformed using the above mentioned procedures and the plasmids were extracted from the bacteria and sequenced as described above. Based on the 5' and 3' ends of the cDNAs, primers were designed that amplified the coding region of each gene. The primer pairs that amplify the full length coding regions were Fwd: GAT GAA CGT CAA CAT GAA GGG and Rev: GCT TAA CAT TTT GTA GAC TTT ATT GAA AGT G for Class I α; Fwd: GTA CAG TAT AGT CCT CCT GTA TGT CTC TCT G and Rev: GCT GTA GTC TCA GTT ACA CTG GTT for Class II α; Fwd: TGT CKA WGY CAA TTS SCT TC and Rev: CTA GAG CAC CCC AGA AGA C for Class II β; and Fwd: GCA CAG GCA GAC GGA TCT TTC and Rev: CAT ATC TGC CTC CCA GGT GTA G for β 2 m. The full length coding regions were 1074 bp for Class I alpha, 711 bp for Class II α, 741 bp for Class II β, and 354 bp for β 2m.

The full length coding sequences from Class I α, β 2 m, Class II α and Class II β were aligned with other salmonid counterparts using the software Muscle (Edgar 2004), translated to amino acid sequences using the software BioEdit (Hall 1999) and their divergence from the other salmonid sequences was analyzed with the software MEGA 4 (Tamura *et al*. 2007).

We found Arctic charr Class I α (SAAL UBA-Figure 2-1) to resemble the characteristics of classical Class I molecules and therefore it was assumed to correspond to the UBA gene described in other salmonids (Grimholt *et al*. 2003): It possessed twelve conserved residues that are known to be essential to bind antigenic peptides: L5, Y7, Y9, F21, G25, Y57, R82, T140, K143, W144, Y157, Y169 (Kaufman *et al*. 1994). Acidic residues located approximately between the positions 215 and 227 are known to be critical for the contact between the α 3 domain and the CD8 receptor of lymphocytes (van Erp *et al*. 1996) and the Arctic charr MH Class I molecule possessed such residues: D215, D218, E221, D222, E224, E227. Also conserved in Arctic charr UBA is a set of residues critical for the association with β2m: T10, Q93, Q112, G117, D119, Q237 (Hansen *et al*. 1996). The molecule also showed a potential glycosylation site at N84. Most of the residues that are related to the structure of the α 3 domain are also conserved: C198, F203, Y204, P205, W212, G234, Y254, C256, V258, as well as other structural residues: F203, Y204, P205, W212, G226, Y254, V258 (Grimholt *et al*. 2002). The Arctic charr MH Class I α molecule contained the usual four conserved cysteines in salmonids (C98, C162, C198, C256) (Grimholt *et al*. 2002). Two potential phosphorylation sites conserved through salmonids were also present: S329, S332 and three of the four residues that are believed to form salt bridges in zebrafish class I molecules are present in Arctic charr (H3, H90, D116) (Hansen *et al*. 1996). A fourth residue, N28 was not conserved although it preserved the polarity of the site.

Figure 2-1. MH cDNA sequences.

Class I α (SAAL-UBA), Class II α (SAAL-DAA) and Class II β (SAAL-DAB) amino acid sequences are shown. Over the sequences is indicated the first amino acid of each domain of the molecule. Tm indicates the area that covers the connective peptide, transmembrane and cytoplasmic domain.


LP 10 α1 30 40 50 60 70 80 90 100
SAAL-UBA MKCVILLLLGMALHSSSAATHSLRYVYTATSGIPDFPDFVTVGLVNGEPISYYDSIIRRETPRQDWMAKTEGSDYWESQTQVSIGAEQSFKANIDIAKQRFNQTGGVHI
α2 120 130 140 150 160 170 180 190 200
SAAL-UBA SQRMYGCEWNDETGATGGYFQDGYDGEDFIAFDLKTKTWIAPKPQAVITKHKWDSNIGQTEQQKNYLTQECIEWLKKYLDYGKSTLMRTVRPS
α3 210 220 230 240 250 260 270 280 290 300 310
SAAL-UBA VSLLQKTPSSPVTCHATGFYPSGVMVFWQKDGQDHHEDVEHGETLQNDDGTFQKSTHLTVTPEEWKKTQYQCVVQVTGIKEDIITVLTESEIKTNWNASNFVLIAVVVALL
Tm 320 330 340 350
SAAL-UBA VIVAVAVAVAVGVVIWKKKSKKGFVPASTSDTDSENSGKGAQKI
LP 10 α1 30 40 50 60 70 80 90 100 110 120
SAAL-DAA MKTSVIVLILCCQVYAEDKLLHIDLYIAGCSDSDGVDMYGLDGEELWYADFKKGEVVMPLPPFADPFTYPGAYEQAVGNQGACKANLATSIKAYKNPPETIAPPHSSIYPRDDVELGVEN
α2 130 140 150 160 170 180 190 200 210 Tm 230
SAAL-DAA TLICHVSGFHPAPVRVRWTRNNQNVTEGVRTSTPYPNTDFTLNQFSSLTFTPEEEDIYGCTVEHKGHIHEPLTRIWEPEVSQPSVGPAVFCGVGLTLGLLGVATGTFFLIKGNQCN
LP 10 β1 30 40 50 60 70 80 90 100 110 120
SAAL-DAB MSMPIGFYICLTLFLSIFSGTDGYFLQMVRQCQYSSKDLHGIESIDSYVFNQVENIRFNSTVGKYVGYTEVGLKNAETWNKGSELARELGQLERFCKHNADIHYSAVLDKTVEPHVRLSSVT
β2 130 140 150 160 170 180 190 200 210 Tm 230 240
SAAL-DAB PPSGRHPAMLMCSAYDFYPKPIRVTWLRDGHEVKSDVTSTEELANGDWYYQIHSHLEYTPKSGEKISCMVEHNSLTEPMVYHWDPSLPESERNKIAIGASGLVLGAVLALAGLIYYKKKSSGVL

The Class I α transcription pattern (Figure 2-2) showed potential expression in all tissues studied, except for brain tissue where immunological activity is generally low (Steinmetz and Hood 1983).

We compared Arctic charr Class I α with classical full length translated sequences from other salmonids published in Genbank (AAG25206; AAG25205; BAA09553; AAN75116; AAN75113; AAG25197; AAG25199). The average p-distance of the amino acid sequence from the other salmonids was 0.24, with a maximum p-distance from the *Oncorhynchus gorbuscha* sequence of 0.39, and the closest similarity was with *Oncorhynchus mykiss* allele Onmy-UBA201 at 0.18. The divergence among salmonids in the leader peptide was 0.18, 0.24 in the α 1 domain, 0.29 in the α 2 domain, 0.16 in the α 3 domain, and the area of the connecting peptide, transmembrane and cytoplasmic domain had a surprising divergence of 0.38. The higher divergence of α 1 and α 2 domain is expected in classical Class I molecules since they hold the PBR. The divergence on the three last domains of the molecule is caused by the low conservation of the connecting peptide and cytoplasmic domains, while the transmembrane region appears highly conserved among *Oncorhynchus mykiss*, *Salmo salar* and Arctic charr. The polymorphism in the last regions of the cDNA involves synonymous and non-synonymous substitutions so it is probably a consequence of the relaxation of the selective pressure over connecting peptide and cytoplasmic domain. High polymorphism in the cytoplasmic domain has been reported, and in some cases the cytoplasmic domain is no longer present in the molecule (Obuchi *et al*. 2001). Class I α, α 1 and α 2 domains combined had a closer similarity with the allele Sasa-UBA0301 of S*almo salar*, and they were most distant from the allele of *Oncorhynchus mykiss*, Onmy-UBA0701. The discordance between the distances of these domains versus the distances found using the whole molecule are probably due to the fact that

the α 1 and the α 2 domain polymorphisms are more related to specific selective pressure of pathogens while the variation in the rest of the molecule tends to follow the genetic line of the species nonselectively (Vogel *et al*. 1999).

The translated β 2 m cDNA (Figure 2-3) showed common characteristics for a member of the Ig superfamily, with two cysteines in positions 44 and 100 that likely join two sides of the molecule in a disulfide bridge. Also, the residues around one of the cysteines presented the motif YTCRVRH, a pattern commonly found in the Ig family (Dixon *et al*. 1993). Although we only sequenced one β 2 microglobulin cDNA, the Southern blot showed the presence of at least six different loci in the genome (Figure 2-4). This is in agreement with the studies of (Magor *et al*. 2004; Shum *et al*. 1996) who found 10-12 different allelic sequences in *Oncorhynchus mykiss* that varied in the untranslated and leader peptide sequences, not the mature peptide sequence. The northern blot of β 2 m (Figure 2-2) showed higher mRNA levels in gill, head kidney, intestine, posterior kidney and spleen, all organs where high Class I activity normally occurs. The mRNA levels also agreed with the expression seen in the western blot analysis where the highest expression levels also corresponded to the same tissues. However, due to the higher sensitivity of the Western blot, β 2 m is detected at lower levels in all tissues analyzed except muscle (Figure 2-2). This is expected for a molecule that pairs with MH Class I α to present peptides in most somatic cells (Steinmetz and Hood 1983). The β 2 m amino acid sequence had a total divergence of 3% when compared with salmonid counterparts (AAB04654; AAG17525), and a 42% when comparing with several other teleosts (AAC67230; CAD44965; ABB60037; AAH62841; BAD22758; AAW65850; BAB60851; CAA10761; AAN40738) (Figure 2-3 A). The phylogenetic tree constructed (Figure 2-3 B) shows Arctic charr β 2m clustering with salmonids, closer to the *Oncorhynchus mykiss* molecule than from the *Salmo salar*.

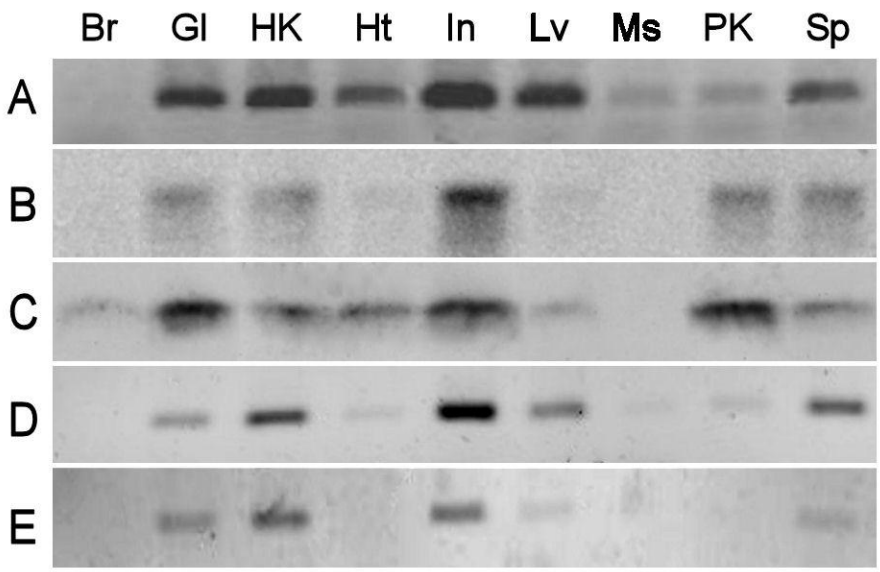


Figure 2-2. Tissue distribution of the MH transcription.

A: RT-PCR for MH Class I; B: Northern blot for β 2 m; C: Western blot for β2 m; D: RT-PCR for MH Class II α; E: RT-PCR for MH Class II β. The RNA or protein extracts were obtained from Br: Brain; Gl: Gill; HK: Head kidney; Ht: Heart; In: Intestine; Lv: Liver; Ms: Muscle; PK: Posterior kidney and Sp: Spleen

A

```
          10        20        30        40        50        60        70        80        90        100       110
     ....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|....|..
Saal  MKYILSIVVLGLIYSAVESKESSPKVQVYSRNPGNFGDKNTLICHVSGFHPPDISIQLLKNGVEIPDAKQTDLAFERQGWQFHLTKSVGFTPASGEEYTCRVRHLKNLKTYTWEADM
Onmy  ..................A...P.....................................................-.....................................
Sasa  ..S.......V...........P.....................................................-................D.....................
Icpu  ..FL..F...AVFSASAFA...P..I........E..KE........D......T.D.....EV..N.E.......-K..K.......S...T.NDKF........ET.NIS..P..
Bain  .RA.ITFALFCVL.IT.QA.T.........HF..EY.KE...............T.E..R..EIL.NTQ.......-K..........S...ER.QD.A.K...MS.TNA.S..PN.
Dare  .RALITFAL.C.L.IT.QG.V.T...H...HF..EY.KP.....Y..S........E.....QVMS.T........-K..........A...EK.D....S...M.ET.KFS..PN.
Ctid  .RALV.FALFCVL.IS.QG.V....I....HY..EY.KE.....Y..N........E.....EV.A..Q.......-K..........S.K.EKSD..S...K.MSDN..IV..SN.
Stvi  .MLAWCLAA.VAVSF.AD..HT........NH..EY.KD.......T.......T..VM.....L.E.........K-.D.H.....T.P...MD.DK...K.T.GNTV.Y.A..PN.
Pscr  ..FV.CLAA.VAVCYSQD..HTR........D..K..S..V.............T...M.DEA.L.N.N......K-...R.....N.A...LD..K.S...T.GQKVSN.A..PN.
Orla  ..ELFF.AA.AAFCC.AF..T.P.T......D..QY.KE........N......T.R.MEDDQ.L.E........K-.N.H......AP...KQ.AR...K.T.SGLA.D....SN.
Gamo  ..ITFCT.AVA.LLCT......V.......SK.AEL.EA.....L.QA....E.T.E...D.AVL.N.N....S.D-KT.H...S.FADI..SKE.S.A.Q.T.IGKRTVHA..PN.
Paol  .GLLICSLL...LCCSM-A.DTP...E..T.E.EE..KP.SF.......Y..Q.N.T...D.K...NTQ.......-AN.YYY...H.S...KEDD.FI...T.MGKS.DHFLMIGL
```

B

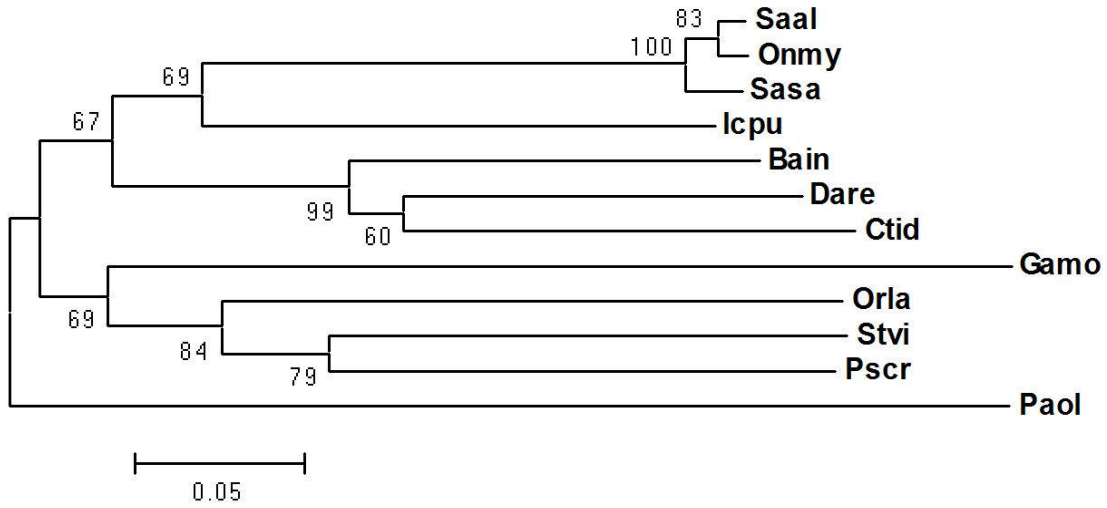


Figure 2-3. The β 2 m amino acid sequence. Figure A correspond to the alignment of the translated sequence of Arctic charr β 2 m with other teleosts β 2 m amino acid sequences: Onmy (*Oncorhynchus mykiss* AAB04654), Sasa (*Salmo salar* AAG17525), Icpu (*Ictalurus punctatus* AAC67230), Bain (*Barbus intermedius* CAD44965), Pscr (*Pseudosciaena crocea* ABB60037), Dare (*Danio rerio* AAH62841), Ctid (*Ctenopharyngodon idella* BAD22758), Stvi (*Stizostedion vitreum* AAW65850), Orla (*Oryzias latipes* BAB60851) Gamo (*Gadus morhua* CAA10761), Paol (*Paralichthys olivaceous* AAN40738). Figure B is the phylogenetic reconstruction using the p-distances among the amino acid sequences, the tree was bootstrapped 1000 times, and only bootstrap values over 50% are shown.

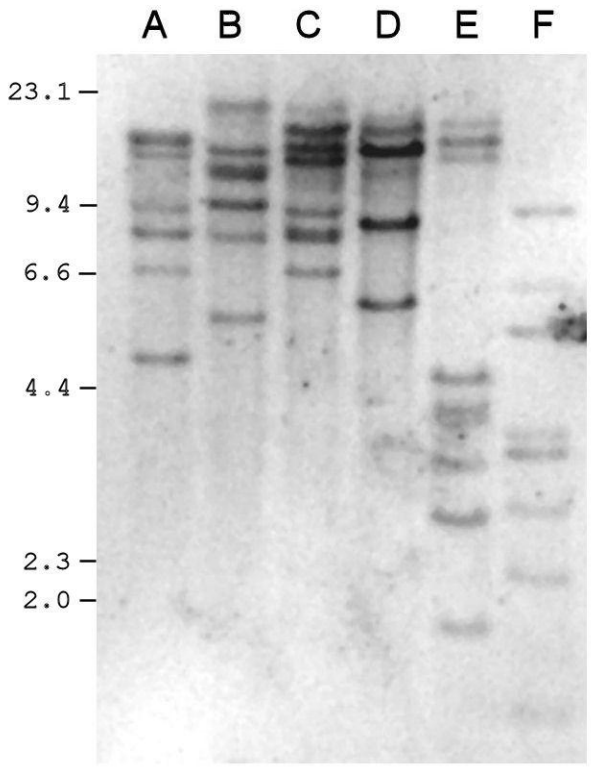


Figure 2-4. β 2 m Southern blot.

The genomic DNA extract was digested with A: *Bam* HI; B: *Hind* III; C: *Eco* RI; D: *Pst* I; E: *Taq* I. Sizes positions in kilobases are indicated in the left.

The Arctic charr MH Class II α gene (Figure 2-1) had characteristics similar to other Class II α molecules described in salmonids: It possessed the usual four cysteines, two of them close enough to form an intradomain disulfide bond (C13, C66) (Sultmann *et al*. 1993). It also presented a potential N-glycosylation signal, NVT between the sites 127-129, typical of teleost class II α. In the transmembrane domain, Arctic charr Class II α presented a typical motif of Class II α chains, GXXXGXXGXXXG (X being a hydrophobic residue), that in Arctic charr is shown as GLTLGLLGVATG. The G amino acids of the motif are believed to shape the area that contacts the Class II β chain to form the Class II receptor (Cosson and Bonifacino 1992). The transcription pattern of Class II α was as described for other classical Class II molecules, with high expression in gill, head kidney, intestine, liver and spleen (Figure 2-2). Excepting liver and gill, all other tissues are known secondary lymphoid organs in fish, so antigen presentation is expected in them. The expression in liver might be due to high blood content and APC transit. The expression in gill might be related to the exposure of this organ to external contamination in an analogous function to what is observed in lungs on terrestrial animals (Steinmetz and Hood 1983).

We compared the Arctic charr Class II α molecule with other salmonid Class II α amino acid sequences (AAL40122; CAB96452; CAB96451; CAB96450). The salmonid Class II α molecules had a mean p-distance of 0.14, with Arctic charr having the closest similarity with the *Oncorhynchus mykiss* allele Onmy-DAA03 and the highest difference with the allele Onmy-DAA02. The divergence in the leader peptide was 0.06, 0.21 in the α 1 domain, 0.11 in the α 2 domain, and the area of the connecting peptide, transmembrane and cytoplasmic diverged only 0.06. The high divergence found in the α 1 domain is a consequence of the PBR. It is interesting to note that the highest and the lowest similarities of the Class II α molecule are both with

Oncorhynchus mykiss alleles, which is an effect of the trans-species polymorphism phenomenon that characterize MH receptors (Klein 1987). According to this evolutionary process, the polymorphism of MH molecules predates speciation and therefore in this case, the Onmy-DAA03 allele has higher similarity with the Arctic charr Class II α allele described, than with the allele Onmy-DAA02 that belongs to the same species.

The Arctic charr MH Class II β gene (Figure 2-1) also showed qualities that are observed in salmonid classical Class II β molecules. It contained four cysteines, whose surrounding amino acids are fairly conserved among teleosts (Hordvik *et al*. 1993). It also contained residues conserved among salmonids between the sites 39-42 and 132-135 that are thought to interact with the CD4 receptor in T-lymphocytes (Cammarota *et al*. 1992). Glycine residues in the positions 203 and 207 that are necessary for folding and transporting in mouse were also present in Arctic charr (Wade *et al*, 1993). Also, the sequence VGYT in the position 44-47, which is a conserved area in most teleosts as a contact for the Class II α molecule. Sites 35-41, that are thought to be involved in invariant chain contact, were also conserved in Arctic charr Class II β, as well as the conserved motif NGDW in the sites 145-148 (Glamann 1995).

Arctic charr Class II β transcription indicated potential expression in the gill, head kidney, intestine, liver and spleen, a pattern expected for a MH Class II gene (Figure 2-2). Compared with other salmonid Class II β genes (X70167; X70166; U20944; U20943), the mean p-distance of the Arctic charr molecule was 0.12. The Arctic charr Class II β molecule had the closest similarity with the *Salmo salar* allele c22. The leader peptide diverged 0.2, the β 1 domain 0.23, the β 2 domain 0.03, and the area of the connecting peptide, transmembrane and cytoplasmic domain diverged 0.03. The high divergence of the β 1 domain was expected since it contains the PBR. The unexpected high divergence of the leader peptide is caused by a 10-13 amino acid

motif, conserved within species but not between species. Interestingly when comparing this motif among salmonids, the non-synonymous substitutions outnumber the synonymous by 2.25, a characteristic generally reserved only for the PBR in MH molecules. This might indicate that natural selection has favored the differentiation of the Class II β leader peptide in the different species of the salmonidae family. The leader peptide of Class II β is fully conserved in all *Salmo salar* sequences published in Genbank (AM259944-AM259955; AM229714-AM229721; AJ438971-AJ438975; AJ439067-AJ439069; X70165-X70167), and the variation within *Oncorhynchus mykiss* DAB gene (AF115532-AF115535; U20943-U20945) occurs only in 3 residues. However, the *Oncorhynchus mykiss* DBB leader peptide (AF115529; AF115531) is highly divergent from the DAB counterpart. Since leader peptides are involved in the regulation of protein expression, its divergence might be related with variation on the expression levels of different MH Class II β loci. In fact the expression of DBB genes is known to be very low compared to DAB in Atlantic salmon (Grimholt *et al*. 2003). Given that the duplication of MH loci in the salmonidae and other vertebrates, transcends the speciation events (Bontrop *et al*. 1990; Miller *et al*. 1997), the high divergence in the leader peptides might be a way of prioritizing the expression of one locus over the other, in a family where the excess of gene duplications might have a counter-productive effect on the negative selection of T-cells during maturation (Walser-Kuntz *et al*. 1995). Variation in the expression of different MH loci has been shown in humans to be related with differential promoter sequences (Louis *et al*. 1993), but this is the first time that this potential type of regulation is reported for coding regions. The fact that the polymorphism occurred in coding regions allowed us to calculate the synonymous versus non-synonymous substitutions, and to conclude that the differential expression might be a

consequence of evolutionary processes that overcome simple genetic drift, and that are controlled by selective forces.

Most Arctic charr fisheries are overexploited and management strategies are difficult to design since the geographical and ecological limits of the stocks are generally obscure (Johnson 1980). MH polymorphism data represents the genetic lineage of the populations but also their adaptation to their ecosystems. Populations that have distinct MH patterns are specialized to confront niche-specific pathogens, and therefore comprise independent conservation units that require separate management strategies. This paper sets the basis for MH polymorphism studies on Arctic charr, presenting the classical MH genes on which future studies can be built. All of the Arctic charr MH genes have similar characteristics to their known salmonid counterparts. As shown here, although polymorphism of MH receptors involves mostly the PBRs, it can also occur in other areas of the molecule that might play a role in the defense against pathogens and therefore requires further analysis.

References

Bernatchez, L. and Landry, C.: MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? Journal of Evolutionary Biology 16: 363-77, 2003

Bernatchez, L., Rhydderch J, and F, Kircheis.: Microsatellite gene diversity analysis in landlocked Arctic Charr from Maine. Transactions of the American Fisheries Society 131: 1106-1118, 2002

Bonneaud, C., Chastel, O., Federici, P., Westerdahl, H., and Sorci, G.: Complex Mhc-based mate choice in a wild passerine. Proceedings of the Royal Society of London. Series B. 273: 1111-6, 2006

Bontrop, R. E., Elferink, D. G., Otting, N., Jonker, M., and de Vries, R. R.: Major histocompatibility complex class II-restricted antigen presentation across a species barrier: conservation of restriction determinants in evolution. Journal of Experimental Medicine 172: 53-9, 1990

Brunner, P. C., Douglas, M. R., Osinov, A., Wilson, C. C., and Bernatchez, L.: Holarctic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. Evolution 55: 573-86, 2001

Cammarota, G., Scheirle, A., Takacs, B., Doran, D. M., Knorr, R., Bannwarth, W., Guardiola, J., and Sinigaglia, F.: Identification of a CD4 binding site on the beta 2 domain of HLA-DR molecules. Nature 356: 799-801, 1992

Cosson, P. and Bonifacino, J. S.: Role of transmembrane domain interactions in the assembly of class II MHC molecules. Science 258: 659-62, 1992

Dawkins, R., Leelayuwat, C., Gaudieri, S., Tay, G., Hui, J., Cattley, S., Martinez, P., and Kulski, J.: Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. Immunological Reviews 167: 275-304, 1999

Dixon, B., Stet, R. J., van Erp, S. H., and Pohajdak, B.: Characterization of beta 2-microglobulin transcripts from two teleost species. Immunogenetics 38: 27-34, 1993

Edgar, R. C.: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-7, 2004

Fujiki, K., Gerwick, L., Bayne, C. J., Mitchell, L., Gauley, J., Bols, N., and Dixon, B.: Molecular cloning and characterization of rainbow trout (*Oncorhynchus mykiss*) CCAAT/enhancer binding protein beta. Immunogenetics 55: 253-61, 2003

Glamann, J.: Complete coding sequence of rainbow trout Mhc II beta chain. Scand Journal of Immunology 41: 365-72, 1995

Grimholt, U., Drablos, F., Jorgensen, S. M., Hoyheim, B., and Stet, R. J.: The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. Immunogenetics 54: 570-81, 2002

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebo, S., and Stet, R. J.: MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. Immunogenetics 55: 210-9, 2003

Hall, T. A.: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. 41: 95-98, 1999

Hansen, J. D., Strassburger, P., and Du Pasquier, L.: Conservation of an alpha 2 domain within the teleostean world, MHC class I from the rainbow trout *Oncorhynchus mykiss*. Developmental and Comparative Immunology 20: 417-25, 1996

Hordvik, I., Grimholt, U., Fosse, V. M., Lie, O., and Endresen, C.: Cloning and sequence analysis of cDNAs encoding the MHC class II beta chain in Atlantic salmon (*Salmo salar*). Immunogenetics 37: 437-41, 1993

Johnson, L.: The Arctic charr, *Salvelinus alpinus*. In e. E. K. Balon (ed.): Charrs: salmonid fishes of the genus Salvelinus, pp. 19-98, Dr. W. Junk, The Hague, 1980

Jonsson, B. and Jonsson, N.: Polymorphism and speciation in Arctic charr. Journal of Fish Biology 58: 605-638, 2001

Kales, S., Parks-Dely, J., Schulte, P., and Dixon, B.: Beta-2-microglobulin gene expression is maintained in rainbow trout and Atlantic salmon kept at low temperatures. Fish and Shellfish Immunology 21: 176-86, 2006

Kaufman, J., Salomonsen, J., and Flajnik, M.: Evolutionary conservation of MHC class I and class II molecules-different yet the same. Seminars in Immunology 6: 411-24, 1994

Klein, J.: Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Human Immunology 19: 155-62, 1987

Klemetsen, A., Amundsen, P., Dempson, B., Jonsson, B., Jonsson, N., O'Connell, F., and Mortensen, E.: Atlantic salmon *Salmo salar* L., brown trout *Salmo trutta* L. and Arctic charr *Salvelinus alpinus* (L.): a review of aspects of their life histories. Ecology of Freshwater Fish 12: 1-59, 2003

Lamont, S. J.: Impact of genetics on disease resistance. Poultry Science 77: 1111-8, 1998

Lohm, J., Grahn, M., Langefors, A., Andersen, O., Storset, A., and von Schantz, T.: Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. Proceedings of the Royal Society of London. Series B. 269: 2029-33, 2002

Louis, P., Eliaou, J. F., Kerlan-Candon, S., Pinet, V., Vincent, R., and Clot, J.: Polymorphism in the regulatory region of HLA-DRB genes correlating with haplotype evolution. Immunogenetics 38: 21-6, 1993

Magor, K. E., Shum, B. P., and Parham, P.: The beta 2-microglobulin locus of rainbow trout (*Oncorhynchus mykiss*) contains three polymorphic genes. Journal of Immunology 172: 3635-43, 2004

Messaoudi, I., Guevara Patino, J. A., Dyall, R., LeMaoult, J., and Nikolich-Zugich, J.: Direct link between mhc polymorphism, T cell avidity, and diversity in immune defense. Science 298: 1797-800, 2002

Miller, K. M., Withler, R. E., and Beacham, T. D.: Molecular evolution at Mhc genes in two populations of chinook salmon *Oncorhynchus tshawytscha*. Molecular Ecology 6: 937-54, 1997

Obuchi, N., Takahashi, M., Nouchi, T., Satoh, M., Arimura, T., Ueda, K., Akai, J., Ota, M., Naruse, T., Inoko, H., Numano, F., and Kimura, A.: Identification of MICA alleles with a long Leu-repeat in the transmembrane region and no cytoplasmic tail due to a frameshift-deletion in exon 4. Tissue Antigens 57: 520-35, 2001

Sambrook, J., Russell D, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, 2001

Savvaitova, K.: Patterns of diversity and processes of speciation in Arctic charr. Nordic Journal of Freshwater Research 71: 81-91, 1995

Shum, B. P., Azumi, K., Zhang, S., Kehrer, S. R., Raison, R. L., Detrich, H. W., and Parham, P.: Unexpected beta2-microglobulin sequence diversity in individual rainbow trout. Proceedings of the National Academy of Science 93: 2779-84, 1996

Steinmetz, M. and Hood, L.: Genes of the major histocompatibility complex in mouse and man. Science 222: 727-33, 1983

Sultmann, H., Mayer, W. E., Figueroa, F., O'HUigin, C., and Klein, J.: Zebrafish Mhc class II alpha chain-encoding genes: polymorphism, expression, and function. Immunogenetics 38: 408-20, 1993

Tamura, K., Dudley, J., Nei, M., and Kumar, S.: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596-9, 2007

van Erp, S. H., Dixon, B., Figueroa, F., Egberts, E., and Stet, R. J.: Identification and characterization of a new major histocompatibility complex class I gene in carp (*Cyprinus carpio* L.). Immunogenetics 44: 49-61, 1996

Vitiello, A., Potter, T. A., and Sherman, L. A.: The role of beta 2-microglobulin in peptide binding by class I molecules. Science 250: 1423-6, 1990

Vogel, T. U., Evans, D. T., Urvater, J. A., O'Connor, D. H., Hughes, A. L., and Watkins, D. I.: Major histocompatibility complex class I genes in primates: co-evolution with pathogens. Immunological Reviews 167: 327-37, 1999

Walser-Kuntz, D. R., Weyand, C. M., Fulbright, J. W., Moore, S. B., and Goronzy, J. J.: HLA-DRB1 molecules and antigenic experience shape the repertoire of CD4 T cells. Human Immunology 44: 203-9, 1995

Chapter 3. Classical and non-classical Major Histocompatibility Class I gene polymorphism in Arctic charr (*Salvelinus alpinus*)

Introduction

Classical Major Histocompatibility (MH) genes encode membrane receptors involved in the distinction between self and non-self peptides. These receptors bind peptides and carry them to the cell surface for presentation to T-lymphocytes. T-cell receptors bind the MH-peptide complex and if the peptide is foreign, the T-lymphocyte initiates an immune response that ultimately leads to the elimination of the pathogen that was the source of the foreign peptide (Steinmetz and Hood 1983). Since peptide binding is a key step in the immune response, pathogens have evolved to develop proteins with peptides that are hard for MH receptors to bind. As a result, vertebrate evolution has turned the peptide binding region of MH genes into the most polymorphic coding region known to date (Grimholt *et al*. 2003; Reche and Reinherz 2003). The high polymorphism found in MH genes is meant to increase the possibilities of survival of the species in the face of constantly changing parasites and pathogens (Bernatchez and Landry 2003; Bonneaud *et al*. 2006; Lamont 1998).

Since polymorphism is crucial to disease survival, MH gene evolution has undergone multiple rounds of gene duplication that have allowed rapid colonization of different available niches (Klein 1991). After colonization and settling the environmental conditions become less variable and many copies of MH genes lose relevance in the presentation of peptides (Miller *et al*. 2002). Some of these extra MH genes decay and finally convert to pseudogenes that do not express a functional molecule (Shiina *et al*. 2005). Other MH genes modify their original function and become part of different biochemical pathways that may or may not be related to peptide

presentation. These are called non-classical MH genes, and they are generally non-polymorphic and have different patterns of expression than those of classical MH genes (Shum *et al*. 1999).

There are two types of MH receptors in most vertebrates, MH Class I and MH Class II. MH Class II receptors present peptides from extracellular pathogens and are expressed only on specialized immune cells called Antigen Presenting Cells (APC). MH Class I receptors present peptides from intracellular or vesicular pathogens and are expressed in all somatic cells. MH Class I genes are composed of 6 exons. The second and third exons code for the peptide binding region in classical class I molecules and form the most polymorphic region of the gene (Reche and Reinherz 2003). In salmonids, of ten MH Class I genes described so far, only one, UBA, is considered to be a classical MH gene (Miller *et al*. 2006). The assertion is based on the expression pattern and high degree of polymorphism of UBA, since direct functional analyses have not yet been performed in fish. The other nine MH Class I genes are considered non-classical, but their specific functions are still unknown. Studies on polymorphism of MH Class I genes in global populations have only been reported for UBA genes of salmonids. The resulting lack of knowledge obscures the physiological function of non-classical genes and complicates the elucidation of their evolutionary importance.

This report describes the polymorphism of four MH Class I genes, UBA, UCA, UGA and UEA, in several populations of Arctic charr (*Salvelinus alpinus*). Arctic charr is the most widestly distributed northern salmonid and shows unprecedented levels of divergence and adaptation (Johnson 1980). Arctic charr has adapted to the environment of many postglacial lakes to the extent of generating different morphotypes to exploit different niches (Alekseyev *et al*. 2002; Allendorf and Thorgaard 1984; Klemetsen *et al*. 2003). Adaptation to different environments or habitats is always associated with exposure to different types of parasites (Dick and Belosevic

1980; Knudsen 1995; Knudsen *et al*. 1997), which select for different and novel MH alleles in the resident species.

The MH Class I polymorphism reported here is discussed in the light of studies that used mitochondrial DNA data to divide the global Arctic charr species into four relatively homogeneous geographic lineages: an Atlantic lineage including southeastern Canadian and European Arctic charr; an Arctic lineage for northern North America including western Greenland; a Siberian lineage for northern Russia, and an Acadian group for landlocked populations found largely in southern Québec and Maine (Brunner *et al*. 2001). Studying the MH Class I polymorphism in Arctic charr provided a more global perspective of the nature of the different genes, and information on the adaptation of populations to different environments. Working with Arctic charr allowed us to analyze fish samples from very different parts of the world and resulted in the discovery of a wider range of variation in their MH Class I molecules.

Materials and Methods

Fish samples

Muscle samples of fresh caught Arctic charr were immersed in 70% ethanol or RNA preserving solution (25 mM $C_6H_5NaO_7$, 10 mM EDTA and 4 M $((NH_4)_2SO_4)$ and shipped to our laboratory from the sampling areas: Lake Kiryalta (57°08'N; 119°27'E), in the Transbaikal region of eastern Russia, part of the Siberian lineage described by Brunner *et al*. (2001). Lake Kiryalta is geographically divided into two water bodies (Lake Kiryalta -3 and -4), but connected by a 50 m canal, so considered as only one population for this study; Lake Sitasjaure (68°00'N; 17°25'E) in Sweden and Gander Lake, Newfoundland (48°55'N; 54°35'W) both part of the Atlantic lineage; Lake Aigneau (57°14'N; 70°07'W) in Québec and Resolute Lake (74°41'N; 94°57'W) in Nunavut, Canada, part of the Arctic lineage; and the de la Trinité River (49°25'N; 67°18'W) in Québec, part of the Acadian lineage.

DNA extraction

Samples in ethanol and RNA preserving solution were placed in TE buffer (100 mM Tris, 1mM EDTA) for 2 hrs, and then incubated for 3 hrs in 100 mM Tris, 10 mM EDTA, 240 mM NaCl, 1% SDS and 300 ug/ml Proteinase K at 55 ºC. This was followed by two extractions, each with one volume of phenol and one extraction with one volume of chloroform. The DNA was precipitated with 1/10 volumes of 3 M CH_2COONa pH 5.2 and 2 volumes of ethanol, and then

washed with 70% ethanol. The final pellet was then resuspended in water and stored at 4ºC until used for PCR.

PCR and cloning

PCR forward (SAALUBAF: 5' – GTA TGG HTG TGA GTK GDR TGA TGA GAC– 3') and PCR reverse (SAALUBAR: 5'- GTC GCG TGG CAG GTC ACT GGA G – 3') primers for UBA were designed to match the beginning of the third exon, and the beginning of the fourth exon, respectively. The primers were based on the full length sequences of Arctic charr (Chapter 2).

The conditions for the PCR were 2 min at 95 ºC for initial denaturation, plus 35 cycles of: 95 °C for 30 seconds for denaturation, 40 sec at 60 °C for primer annealing and 1.5 min at 72 ºC for extension. A final extension at 72 ºC for 8 min was done to ensure the addition of an A overhang at the end of the complementary strand.

The PCR amplicons were separated by agarose electrophoresis in a 1% gel to verify size and then extracted directly from the gel using GenElute Agarose Spin Columns (Sigma-Aldrich, St Louis, MO). The fragments were inserted in pGEM-T vectors as per manufacturer instructions (Promega Corporation, Madison, WI). XL-blue strains of *E. coli* bacteria were made competent by washing them twice and then re-suspending them with Thawing buffer (10mM PIPES, 15mM $CaCl\text{-}_2H2O$, 250 mM KCl, 55mM $MnCl_2\text{-}4H_2O$, pH 6.7). The bacteria were transformed by the Inoue procedure (Sambrook *et al.*, 2001) and grown in LB plates with ampicilin (100 ug/ml) plus 1 uM IPTG and 100 ug/ml X-gal. Eight to twelve white colonies were selected and grown in LB media. The bacteria were used for PCR with SAALUBAF primer and a Reverse primer for

Denaturant Gradient Gel Electrophoresis (DGGE), designed to match the end of the third exon (SAALUBADGGER: CGG GCG GGG GCG GCG GGC CGG GCG CGG GGC GCG GCG GGC GCT TCT TCA RCC ACT CAA TGC). The PCR conditions were the same as described above, but the annealing temperature was lowered to 56 °C. The amplicons were separated in a denaturing gradient of 25-60% for 15.5 hrs at 60 °C (Bio-Rad laboratories, Richmond, CA). Only the alleles that differed between samples were selected for DNA sequencing in order to avoid redundancies. The plasmids selected for sequencing were extracted from the grown bacteria with GenElute Plasmid Miniprep Kit (Sigma-Aldrich, St Louis, MA) according to manufacturer instructions, and sequenced using the T7 and Sp6 primers with a big dye terminator v3.1 (Applied Biosystems, Foster City, CA) in a 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA).

Since the PCR yielded four clearly different lineages of sequences, later identified as third exons from different loci (UBA, UCA, UGA and UEA), specific reverse primers for each gene were designed to match the third intron, about 200 bp upstream from the beginning of the fourth exon. The designed primers were SAALUBA2R: GAA TAA TCA GAC ACA CGG AGA TG; SAALUCAR: GAT TTC TGT AGA TAA CAC AGA ATC AG; SAALUGAR: GAG AGA CAT ACA AGC ATA GTG ATG and SAALUEAR: GAT TGA ATA GGA CGC CAT GTG, for UBA, UCA, UGA and UEA respectively. The experimental conditions for all the primers were kept the same as those described for the SAALUBAR primer.

Sequencing and data analysis

The sequences obtained were aligned using the software Muscle (Edgar 2004). An alignment of the previously sequenced full length Class I cDNA sequence (Chapter 2), was used as a guide for the definition of the end of the second exon and the beginning of the intron. Only the third exons of the genes were used for the analysis. BioEdit (Hall 1999) was used to define the open reading frames and to translate them into protein sequences. The software was also used to obtain entropy values (H) that measure the variability of each amino acid residue position (Shannon, 1948 *fide in* Reche and Reinherz, 2003). Following similar criteria used by other authors (Consuegra *et al*. 2005a; Consuegra *et al*. 2005b; Reche and Reinherz 2003), we considered highly entropic sites as those having values approximately 25% of the maximum observed entropy in each molecule. MEGA 4 (Tamura *et al*. 2007) was used to find and correct sequences containing singleton errors and to calculate the diversity of the sequences among and within populations, the rates of non-synonymous/synonymous substitutions, and generate the phylogenetic trees using the Nei-Gojobori model (Nei and Gojobori 1986) for non-synonymous substitutions in the third exon sequences of each gene, and the Kimura-2 parameter model (Kimura, 1980) for all four genes together.

Alrequin software ver 3.01 (Excoffier *et al*. 2005) was used to calculate the allelic frequencies and the pairwise F-statistics (F_{ST}) for population differentiation.

Results

A total of 81 fish samples were analyzed, 10 samples from Lake Aigneau, 19 from Gander Lake, 17 from Lake Kiryalta, 12 from Resolute Lake, 12 from Lake Sitasjaure and 11 from the de la Trinité River. A total of 158 different sequences (alleles) were isolated.

Phylogenetic reconstruction and population differentiation

Phylogenetic reconstruction using the third exon of the MH Class I α gene yielded four well defined branches, three of them with very good bootstrap values (Figure 3-1, MH Class I). Members of one well bootstrapped branch had close sequence similarities to a non-classical Class I gene previously described for rainbow trout, *Oncorhynchus mykiss* (Dijkstra *et al*. 2006), named UCA. A second branch corresponded to the previously reported salmonid UGA gene also considered non-classical (Kiryu *et al*. 2005). A third branch corresponded to another non-classical salmonid Class I α gene named UEA in *Oncorhynchus mykiss* (Shiina *et al*. 2005). The rest of the sequences clustered together relatively well in the phylogenetic tree but with a low total bootstrap value, and had close similarities with the classical Class I α described for salmonids, named UBA (Grimholt *et al*. 2002; Shum *et al*. 2001).

The Arctic charr UBA third exon was very polymorphic and the phylogenetic tree constructed using it presented several internal well bootstrapped branches (Figure 3-1, UBA). More than half of the branches were related geographically, grouping a great majority of individuals belonging to the same population. No branch, however, grouped populations of the same phylogeographic

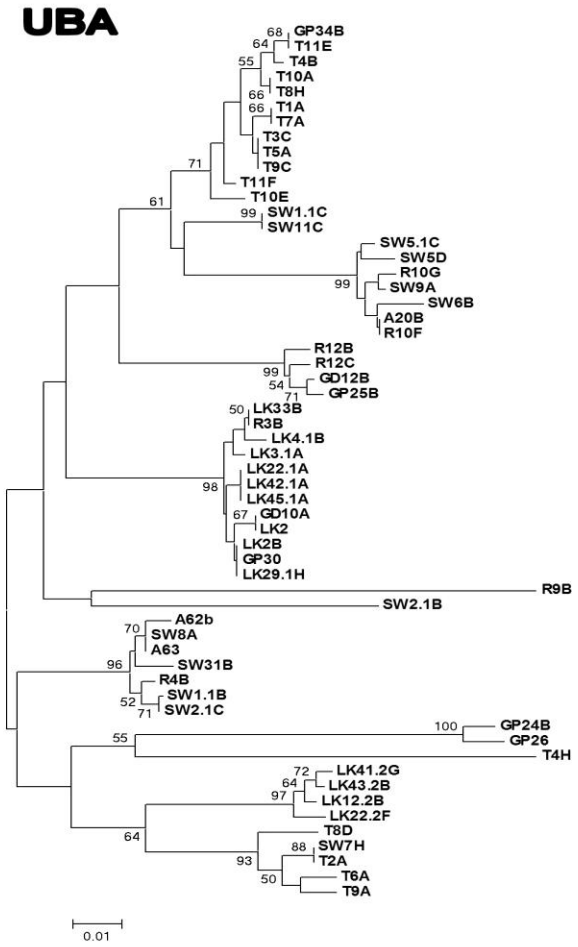
lineage as defined by Brunner (2001). UBA third exon F_{ST} analysis allowed differentiation among some populations and also showed the isolation of all the geographic lineages, excepting the Arctic lineage from the Atlantic lineage (Table 3-1a).

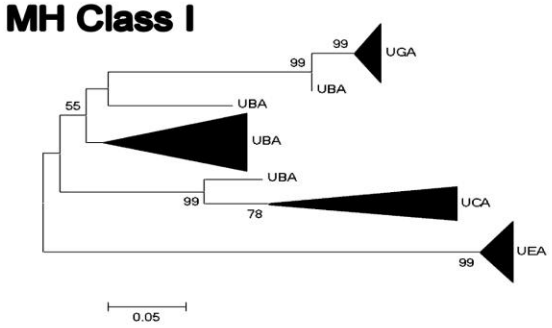
The UCA third exon phylogenetic tree had two separate well bootstrapped main branches, unrelated to geographic distribution (Figure 1, UCA). The tree's branches were separated by long genetic distances. However, it had very few well bootstrapped sub-clusters, and fewer still were geographically related. The homogeneity of the distribution was also reflected by the F_{ST} test for population differentiation, which showed few populations differentiated from each other (Table 3-1b). The UCA third exon F_{ST} analysis was unable to isolate either populations or geographic lineages. The two main branches found in the UCA tree might indicate a loci differentiation between UCA and UDA, usually so similar that they have been impossible to separate by sequence analysis (Dijkstra *et al*. 2006). Nevertheless, independent analysis of either branch separately provided similar results to the ones obtained when the branches were considered as being derived from the same locus.

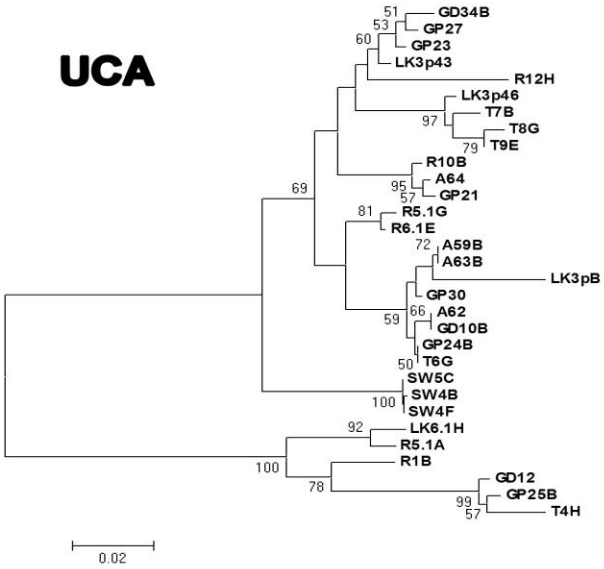
The UGA third exon phylogenetic tree had three main branches that could be related to geographic lineages. Two of the branches comprised only Canadian samples, and the last one contained only fish from Lake Sitasjaure and Lake Kiryalta (Figure 3-1, UGA). Aside from this, the tree did not show any other well bootstrapped group inside the main branches. Nonetheless, the FST analysis using the UGA third exon allowed some differentiation among the individual Arctic charr populations and moreover, it showed the isolation of all three phylogeographic lineages defined by Brunner (Table 3-1c).

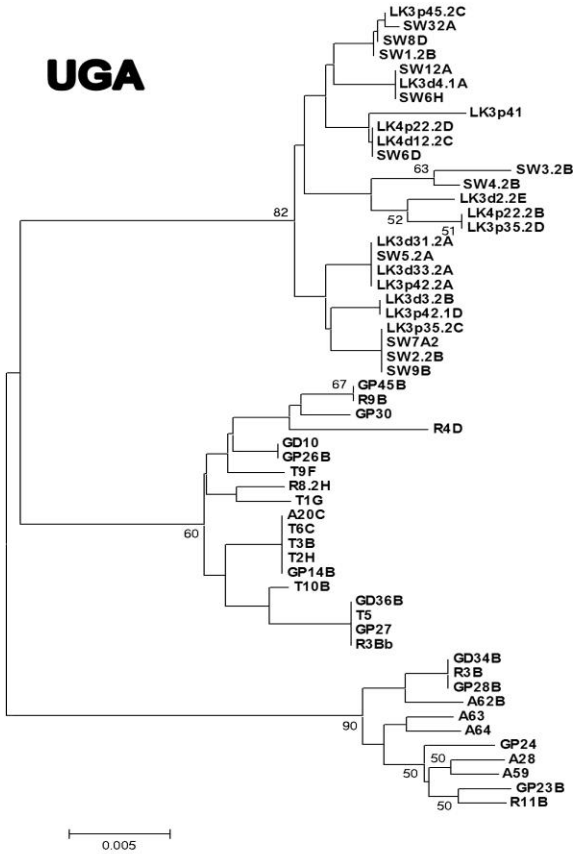
The UEA third exon phylogenetic tree had a very low variability and did not have any well bootstrapped branches. As expected no branch was related to known geographic lineages (Figure 1, UEA). It is surprising then that the obtained FST values did differentiate some of the populations from each other and showed the isolation of the Arctic geographic lineage (Table 3-1d).

Figure 3-1. MH Class I phylogenetic reconstruction using the Nei-Gojobori model for the third exon sequence non-synonymous substitutions, for each of the four genes studied (UBA, UCA, UGA and UEA). The trees were bootstrapped 1000 times and only bootstrap values over 50% are shown. The sequence names are defined for each population and thus start with an A for Lake Aigneau, G for Gander Lake, R for Resolute Lake, SW for Lake Sitasjaure, T for de la Trinité River and LK for Lake Kiryalta. The tree labeled "MH Class I" was constructed using the Kimura-2 parameter model for all the sequences in the four genes. The branches representing each gene are compressed


UBA
GP34B
T11E
T4B
T10A
T8H
T1A
T7A
T3C
T5A
T9C
T11F
T10E
SW1.1C
SW11C
SW5.1C
SW5D
R10G
SW9A
SW6B
A20B
R10F
R12B
R12C
GD12B
GP25B
LK33B
R3B
LK4.1B
LK3.1A
LK22.1A
LK42.1A
LK45.1A
GD10A
LK2
LK2B
GP30
LK29.1H
R9B
SW2.1B
A62b
SW8A
A63
SW31B
R4B
SW1.1B
SW2.1C
GP24B
GP26
T4H
LK41.2G
LK43.2B
LK12.2B
LK22.2F
T8D
SW7H
T2A
T6A
T9A
0.01


MH Class I
UGA
UBA
UBA
UBA
UBA
UCA
UEA
0.05


UCA
GD34B
GP27
GP23
LK3p43
R12H
LK3p46
T7B
T8G
T9E
R10B
A64
GP21
R5.1G
R6.1E
A59B
A63B
LK3pB
GP30
A62
GD10B
GP24B
T6G
SW5C
SW4B
SW4F
LK6.1H
R5.1A
R1B
GD12
GP25B
T4H
0.02


UGA
LK3p45.2C
SW32A
SW8D
SW1.2B
SW12A
LK3d4.1A
SW6H
LK3p41
LK4p22.2D
LK4d12.2C
SW6D
SW3.2B
SW4.2B
LK3d2.2E
LK4p22.2B
LK3p35.2D
LK3d31.2A
SW5.2A
LK3d33.2A
LK3p42.2A
LK3d3.2B
LK3p42.1D
LK3p35.2C
SW7A2
SW2.2B
SW9B
GP45B
R9B
GP30
R4D
GD10
GP26B
T9F
R8.2H
T1G
A20C
T6C
T3B
T2H
GP14B
T10B
GD36B
T5
GP27
R3Bb
GD34B
R3B
GP28B
A62B
A63
A64
GP24
A28
A59
GP23B
R11B
0.005

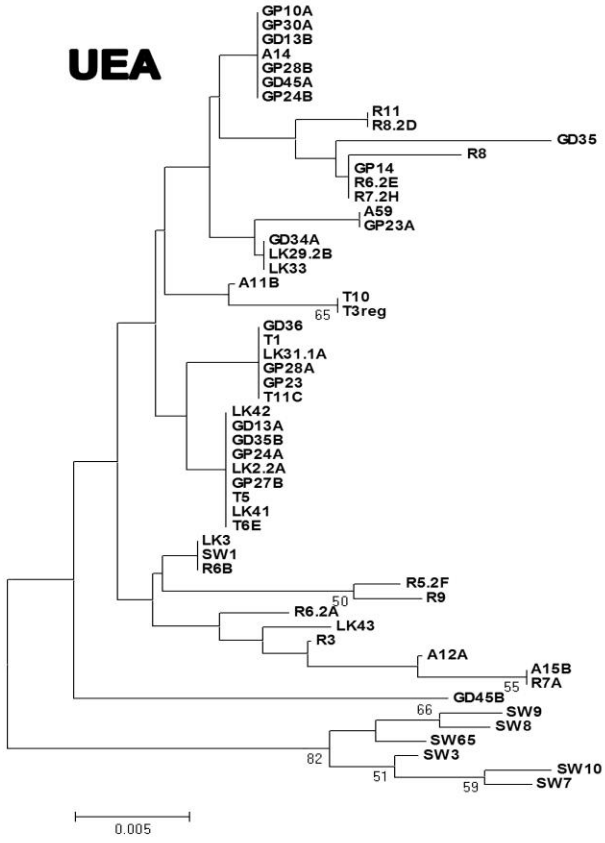

UEA
GP10A
GP30A
GD13B
A14
GP28B
GD45A
GP24B
R11
R8.2D
GD35
R8
GP14
R6.2E
R7.2H
A59
GP23A
GD34A
LK29.2B
LK33
A11B
T10
T3reg
GD36
T1
LK31.1A
GP28A
GP23
T11C
LK42
GD13A
GD35B
GP24A
LK2.2A
GP27B
T5
LK41
T6E
LK3
SW1
R6B
R5.2F
R9
R6.2A
LK43
R3
A12A
A15B
R7A
GD45B
SW9
SW8
SW65
SW3
SW10
SW7
0.005

Table 3-1. Population differentiation according to the third exon of MH Class I genes.

Above the diagonal are the pairwise FST distance values. Below the diagonal are the p-values for the pairwise FST analysis (5000 permutations). P-values <0.05 were considered significant for this study, and are indicative of the isolation of the populations from each other (shaded). Each sub-table (a, b, c, d) shows the differentiation for the UBA, UCA, UGA and UEA gene third exons respectively. Tables numbers ending in .1 correspond to the differentiation of each individual population studied and those ending .2 provide the differentiation of the phylogeographic lineages previously described using mDNA (Brunner *et al*. 2001). The average errors for all the tests were always lower than 0.008. Values under 0.00001 are shown as "0.000".

3-1a.1

UBA	Aigneau	Gander	Kiryalta	Resolute	Sitasjaure	Trinité
Aigneau		0.143	0.380	-0.054	-0.104	0.198
Gander	0.106		0.163	0.046	0.141	0.198
Kiryalta	0.010	0.024		0.270	0.299	0.342
Resolute	0.680	0.245	0.002		-0.033	-0.045
Sitasjaure	0.779	0.017	0.000	0.692		0.120
Trinité	0.051	0.018	0.000	0.019	0.020	

3-1a.2

	Arctic	Atlantic	Siberian	Acadian
Arctic		-0.027	0.280	0.161
Atlantic	0.765		0.185	0.093
Siberia	0.001	0.000		0.342
Acadia	0.004	0.010	0.000	

3-1b.1

UCA	Aigneau	Gander	Kiryalta	Resolute	Sitasjaure	Trinité
Aigneau		0.030	0.103	0.177	0.693	0.126
Gander	0.192		-0.085	-0.052	0.283	-0.036
Kiryalta	0.091	0.594		-0.088	0.340	-0.159
Resolute	0.147	0.520	0.521		0.317	-0.028
Sitasjaure	0.023	0.053	0.027	0.100		0.323
Trinité	0.116	0.553	0.830	0.368	0.016	

3-1b.2

	Arctic	Atlantic	Siberian	Acadian
Arctic		-0.030	-0.066	-0.009
Atlantic	0.628		-0.041	0.001
Siberia	0.592	0.558		-0.159
Acadia	0.532	0.451	0.842	

3-1c.1

UGA	Aigneau	Gander	Kiryalta	Resolute	Sitasjaure	Trinité
Aigneau		0.146	0.662	0.160	0.665	0.574
Gander	0.112		0.521	-0.079	0.526	0.150
Kiryalta	0.000	0.000		0.564	0.042	0.681
Resolute	0.157	0.771	0.000		0.553	0.184
Sitasjaure	0.000	0.000	0.161	0.000		0.690
Trinité	0.005	0.039	0.000	0.009	0.000	

3-1c.2

	Arctic	Atlantic	Siberian	Acadian
Arctic		0.182	0.538	0.317
Atlantic	0.002		0.146	0.234
Siberia	0.000	0.003		0.681
Acadia	0.003	0.002	0.000	

3-1d.1

UEA	Aigneau	Gander	Kiryalta	Resolute	Sitasjaure	Trinité
Aigneau		0.178	0.128	0.059	0.455	0.193
Gander	0.020		0.029	0.306	0.575	0.000
Kiryalta	0.098	0.198		0.228	0.521	-0.023
Resolute	0.218	0.000	0.000		0.441	0.346
Sitasjaure	0.001	0.000	0.001	0.000		0.571
Trinité	0.043	0.431	0.571	0.000	0.001	

3-1d.2

	Arctic	Atlantic	Siberian	Acadian
Arctic		0.155	0.159	0.252
Atlantic	0.000		0.033	0.038
Siberia	0.011	0.192		-0.023
Acadia	0.005	0.202	0.577	

MH Class I diversity

The mean Class I α third exon p-distance was 0.26 with a diversity of 0.07 among the four different genes found. The non-synonymous/synonymous substitution rate was 0.62, very low for what it is expected for MH genes.

The UBA third exon mean p-distance was 0.097. Interpopulation diversity was 0.013 when considering grouping according to populations and 0.014 according to the geographic lineages (Table 3-2). The non-synonymous/synonymous substitution rate was 0.88, a value that comes from the fluctuation around 1 that the different populations had, with values from 1.4 in Lake Aigneau to 0.66 in de la Trinité River populations (Table 3-3).

The UCA third exon mean p-distance was 0.104 with 0.018 of interpopulation diversity. The non-synonymous/synonymous substitution rate was 1.12, indicating marginal balancing selection acting on the α 2 domain of the UCA (Table 3-2). In the Resolute Lake population the level of balancing selection was the highest with a rate of 1.4 and the lowest was held by the Lake Aigneau population with 0.92 (Table 3-3).

The UGA third exon mean p-distance was 0.025 with a 0.01 of interpopulation diversity. The rate of non-synonymous/synonymous substitutions was 1.4. Surprisingly, on the UGA gene the interpopulation rate of non-synonymous/synonymous substitutions was 12.2, indicating that most of the changes that occurred after the separation of the populations might correspond to adaptive changes. Nevertheless, the non-synonymous/synonymous substitution rates in each population were relatively low, with a minimum of 0.45 in de la Trinité River and a maximum of 1.2 in Resolute Lake (Table 3-3). Among the lineages the interpopulation rate of non-synonymous/synonymous substitutions was high (5.6), although each lineage had low mean

values and only the Arctic and Atlantic lineages surpassed the neutral expectations (data not shown).

The UEA third exon mean p-distance was 0.022, with 0.005 of interpopulation diversity (Table 3-2). The non-synonymous/synonymous substitution rate was only 0.38, with the highest rate being 0.6 in de la Trinité River and the lowest (0.2) in Lake Aigneau populations (Table 3-3).

MH Class I entropy

In the UBA gene third exon sequences, 24 residues were highly entropic, eight of which were sites that had high entropy uniquely in the UBA gene (Table 3-4). The remaining sixteen sites also had high entropy in UCA and 4 of them showed high entropy in UGA as well. Only two amino acid sites had high entropy in all four of the genes studied. Seven of the highly entropic sites in the Arctic charr UBA third exon matched with peptide binding residues in the α 2 domain of HLA, out of 14 human peptide binding sites (Reche and Reinherz 2003).

In total UCA had 40 highly entropic residues, 21 that had high entropy only in the UCA gene. Seven sites were also highly entropic in UGA and one in UEA. Nine UCA highly entropic sites matched with the peptide binding residues in the HLA α 2 domain (Reche and Reinherz 2003). UGA had 10 highly entropic sites, three of which had high entropy only in UGA. Four sites matched with binding residues in the HLA counterpart. UEA had four highly entropic sites, one of them unique to UEA. There was not any match between UEA highly entropic sites and the HLA gene (Table 3-4).

Gene	Total	Within	Inter popul.	dn/ds rate	Inter dn/ds
UBA	0.097	0.085	0.013	0.881	0.694
UCA	0.104	0.086	0.018	1.122	1.456
UGA	0.025	0.016	0.010	1.401	12.17
UEA	0.022	0.017	0.005	0.382	0.660

Table 3-2. MH Class I third exon divergence.

The mean distance values in the four genes studied, the diversity within and between populations, non-synonymous/synonymous substitution ratio and the interpopulation non-synonymous/synonymous substitution ratio.

Population	UBA	UCA	UGA	UEA
Aigneau	1.435	0.923	0.981	0.217
Gander	0.931	1.174	1.122	0.478
Kiryalta	1.233	1.109	0.710	0.262
Resolute	0.745	1.365	1.224	0.307
Sitasjaure	1.058	0.000	0.579	0.655
Trinité	0.656	0.990	0.454	0.417

Table 3-3. MH Class I non-synonymous/synonymous substitution ratio in all the populations studied for each of the four genes studied.

Table 3-4. MH Class I α, α 2 domain amino acid entropy.

The entropy values for each amino acid positions, in each of the four genes studied. Entropy values over 25% of the maximum entropy for each individual gene are shaded. Amino acid position is shown in the first column and the superscript indicates peptide binding residues found in any HLA gene in human. Superscript C indicates conserved residues and V indicates variable residues in HLA genes (Reche and Reinherz 2003).

Aa	UBA	UCA	UGA	UEA	Aa	UBA	UCA	UGA	UEA
v 1	0	0	0	0	c 44	0	0.491	0	0
2	0	0	0	0	45	0	0.491	0	0
3	0	0	0	0	46	0	0.318	0	0
4	0	0	0	0	c 47	0.670	0.750	0	0
5	0.765	0.667	0.769	0.761	48	0	0.615	0	0
6	1.603	1.248	1.649	1.524	49	0	0	0	0
7	0	0	0	0	50	0.204	0	0	0
8	0	0	0	0	51	0.150	1.252	0	0.345
9	0	0	0	0	52	0.678	0	0	0
10	0.510	0.143	0	0	53	0.624	1.487	0	0
11	1.048	0.579	0	0	v 54	0.479	0	0	0
12	0.150	0	0	0	55	1.823	0.491	0	0
13	0.510	0.491	0	0	56	0.577	0	0	0
14	0.150	0.318	0	0	57	0	0.997	0	0.345
15	0.401	0.650	0	0	v 58	0.91	0.385	1.051	0
v 16	1.599	0.523	0	0	59	1.795	0.442	0.090	0
17	0	0	0	0.685	60	0.323	0.918	0	0
v 18	1.407	0.385	0	0	c 61	0.510	0.491	0	0.156
19	0	0.318	0	0	62	0	0	0	0
20	0	0	0	0	63	0.997	0.615	0	0
21	0	0.491	0	0	64	0	0	0.641	0
22	0	0	0	0	v 65	0	0	0	0
23	0.150	0	0	0	c 66	0.672	0.318	0	0
c 24	0	0.534	0	0	67	0	0	0	0
c 25	0.150	0	0	0	68	0	0	0	0
26	0.743	0	0	0	c 69	0.553	0.457	0.495	0.156
27	0.150	0.704	0	0	70	0	0.491	0	0
28	0	0	0	0.091	71	0	0	0	0
29	0	0	0	0	72	0	0.318	0	0
30	0.087	0	0.691	0	c 73	0	0.318	0.495	0
31	0	0.442	0.691	0	74	0	0.491	0	0
32	0	0.442	0	0	75	0.251	0.491	0	0
33	1.179	0.571	0	0	76	0	0.688	0.520	0
34	1.027	0.571	0	0	77	0	0.846	0	0
35	0.150	0	0	0.261	78	0.087	0	0	0.091
36	0	0.491	0	0	79	0.087	0	0	0
37	0	0.491	0	0	80	0	0.602	0	0
38	0	0.491	0	0	81	0	0.935	0	0.470
39	0.803	0.143	0	0	82	0	0	0	0
40	0.565	0	0	0	83	0.451	0.794	0	0
41	0	0	0	0	84	0	0	0	0.305
42	0	0	0	0	85	0	0	0	0.091
c 43	0.236	1.059	0.641	0					

Discussion

In most vertebrates there are multiple classical Class I genes involved directly in the presentation of peptide antigens to T-cytotoxic lymphocytes (Stet *et al*. 2003). There are also multiple non-classical genes that share a very similar structure with classical Class I genes, but do not bind and present peptides. Instead, they can be involved in the presentation of lipid antigens to T-cells, as well as several other functions such as regulation of NK cells or iron metabolism (Dijkstra *et al*. 2006).

Classical and non-classical MH Class I genes are generated through evolutionary processes of gene expansion (Klein 1991), followed by stabilization periods in which some MH genes lose the function of peptide presentation and become non-classical (Miller *et al*. 2002). Since the polymorphism of the MH loci predates speciation and they are carried by homologous species separated by millions of years (Ayala *et al*. 1994; Bontrop *et al*. 1990; Miller and Withler 1997), some classical MH Class I molecules have non-classical orthologous counterparts, and vice versa, in close related species (Gyllensten *et al*. 1991; Yeager and Hughes 1999). Thus the classical or non-classical status of certain MH genes may change during evolution.

In the absence of direct functional methods to define classical or non-classical MH Class I genes in fish, it is generally assumed that classical Class I genes should be polymorphic in the population, present evidence of selection in the peptide binding region, be ubiquitously expressed, and tend to conserve certain residues necessary for peptide binding and antigen presentation.

Among the ten MH Class I genes found in salmonids, UBA is the only locus that is indisputably considered a classical MH gene (Grimholt *et al*. 2003; Miller *et al*. 2006). Only recently could the UBA gene's unprecedented polymorphism derived from several deep allelic lineages be assigned to a single locus (Aoyagi *et al*. 2002), which has undergone multiple inter-locus recombinations with non-classical genes (Shum *et al*. 2001).

The recent evolution of the Arctic charr UBA α 2 domain is not guided by balancing selection

When comparing the UBA third exon of Arctic charr with all other salmonids third exon sequences reported in Genebank (data not shown), the Arctic charr sequences clustered with only one of the three lineages found for rainbow trout (Miller *et al*. 2006), and presented a correspondingly lower divergence. On the unique UBA third exon lineage found in Arctic charr, discordant with the findings in rainbow trout (Shum *et al*. 2002), the non-synonymous/synonymous substitution ratios were only positive in some of the populations and not in the global Arctic charr range (Table 3-3).

UBA third exon alleles were the only Class I gene sequences that did not cluster together in a single group with a significant bootstrap in the Arctic charr phylogenetic tree. Indeed, it appeared to be broadly distributed on the margins of the other very well bootstrapped Class I gene clusters (Figure 3-1). The broad distribution is a sign of its high polymorphism, and may also be indicative of the UBA sequences deriving in part from recombination with other Class I non-classical loci, as has been observed for rainbow trout (Miller *et al*. 2006). Although the high divergence of UBA prevents allele sharing among populations, which was also reported in

Atlantic salmon, *Salmo salar*, (Consuegra *et al*. 2005b) in Arctic charr the specific divergence among alleles was not enough to indicate isolation of all populations or phylogeographic lineages (Table 3-1). Although some differentiation could be inferred from the phylogenetic tree and from the F_{ST} analysis (ie. isolation of populations from Lake Sitasjaure, de la Trinité River, Lake Kiryalta), the fact that the isolation was not supported by positive non-synonymous/synonymous ratios, together with a relatively low inter-population diversity indicates that ancestral genetic drift and fixation by positive selection drive UBA third exon polymorphism. Despite the classical nature of UBA in other salmonids, in Arctic charr its recent evolution does not seem to be guided by balancing selection or adaptation to local environments.

The UCA α 2 domain might retain peptide binding function

The UCA gene expression patterns in rainbow trout are consistent with those of classical Class I genes, but its sequence encodes divergent residues that would interfere with the usual structure of classic molecules (Miller *et al*. 2006). In Arctic charr, the total divergence of UCA was higher than UBA and it presented more highly entropic sites (Table 3-4). However, this might be due to the fact that UCA is hardly distinguishable from UDA in salmonids (Dijkstra *et al*. 2006), which might explain the two constituent clusters in the phylogenetic tree (Figure 3-1). The high polymorphism which in only one case clustered together geographically related individuals, only isolated a few populations and lineages in the F_{ST} test (Table 3-1), indicating that the polymorphism is ancestral and not specific for each local population. In Arctic charr, the UCA third exon had a slightly positive ratio of non-synonymous/synonymous substitutions (higher when considering the interpopulation diversity) which indicates balancing selection influencing

its high polymorphism (Table 3-2) (Potts and Slev 1995). When considering the ubiquitous expression patterns found in rainbow trout, the UCA third exon might still retain an important function, that judging by the non-synonymous/synonymous substitutions ratio might be related to the binding of certain varieties of peptides.

The non classical nature of salmonid UEA (Shiina *et al*. 2005) was confirmed in Arctic charr, with extremely low non-synonymous/synonymous substitution ratios, indicative of positive selection and consequent strong sequence conservation (Table 3-2). Nevertheless, it is surprising to note that with a relatively high proportion of interpopulation variation, UEA third exons sequence analysis supported considerable population and lineage differentiation (Table 3-1). The phenomenon is probably a consequence of genetic drift followed by allelic fixation, and is probably not a result of adaptation to local environments.

The UGA gene might be involved in the adaptation to local pathogens in Arctic charr

The UGA gene is expressed in most tissues in rainbow trout and is was considered a non-classical Class I gene because its expression levels vary significantly among different tissues and because it is monomorphic (Miller *et al*. 2006). In Arctic charr, however, the UGA third exon is polymorphic (Table 3-2), and although it was less divergent than the UBA classical gene, UGA fell in the range of classical MH genes polymorphism (O'Brien and Yuhki 1999). UGA also contained several highly entropic sites, and a relatively high proportion of them matched with HLA peptide binding residues (Table 3-4). UGA had the highest ratio of non-synonymous/synonymous substitutions among all the Class I genes studied in Arctic charr,

which indicates balancing selection acting on the UGA evolution. More interestingly, UGA had a very high proportion of interpopulation diversity and the interpopulation non-synonymous/synonymous substitution ratio was 12.2, a very high value even for genes under balancing selection. The interpopulation diversity indicated that although the variation on UGA has not been high for classical class I genes, a lot of it has occurred independently for each population, and much of the independent variation is related to the adaptation to new niches. Consequently, out of the four Class I genes studied, the UGA gene was the most successful at isolating Arctic charr populations (Table 3-3), and it was the only gene that supported the isolation of all the geographic lineages previously defined using mDNA (Brunner *et al*. 2001). The FST levels of differentiation of the lineages and populations were similar to those found for the same populations using Class II α (Conejeros *et al*, submitted), which in teleosts is thought to have limited trans-species polymorphism levels compared with Class I (Shum *et al*. 2001).

The high ratios of non-synonymous/synonymous substitutions and the high interpopulation differentiation suggests that the UGA molecule might be involved in binding a wide variety of peptides that are specific for local geographic areas. The phenomenon is usually seen in mammalian MHC Class I (Kaufman *et al*. 1994) and in teleost MH Class II genes when using them as population markers, as both are related more recent than ancestral polymorphism (Consuegra *et al*. 2005a). Arctic charr, with a broader range of colonization compared to other salmonids, and with superior levels of adaptation to different niches (Jonsson and Jonsson 2001), may require a Class I gene to confront local pathogens. In absence of the wide variety of UBA lineages present in other salmonids, the Arctic charr may be using the UGA gene as a complementary classic Class I, specific to local environments. More studies are required to

confirm the assertion for Arctic charr and to determine whether UGA is best treated as a classical or non-classical gene in salmonids.

Multiple genetic expansion and contraction events have left salmonids with many MH Class I genes, most of them with non-classical functions. MH Class I genes are not used in population studies in salmonids because in contrast to mammals, they show ancestral polymorphism. However, the classical and non-classical status of MH genes changes even within the same taxonomic family, and new adaptation necessities may have lead Arctic charr to recruit non-classical genes to present peptides. The novel used non-classical genes is yet another sign of the outstanding genetic and phenotypic plasticity of Arctic charr and provides researchers with a promising new tool to study differences within and among its populations.

References

Alekseyev, S. S., Samusenok, V. P., Matveev, A. N., and Pichugin, M. Y.: Diversification, Sympatric Speciation, and Trophic Polymorphism of Arctic Charr, *Salvelinus alpinus* Complex, in Transbaikalia. Environmental Biology of Fishes 64: 97-114, 2002

Allendorf, F. and Thorgaard, G.: Tetraploidy and. the evolution of salmonid fishes. Turner BJ, ed, New York, 1984

Aoyagi, K., Dijkstra, J. M., Xia, C., Denda, I., Ototake, M., Hashimoto, K., and Nakanishi, T.: Classical MHC class I genes composed of highly divergent sequence lineages share a single locus in rainbow trout (*Oncorhynchus mykiss*). Journal of Immunology 168: 260-73, 2002

Ayala, F. J., Escalante, A., O'Huigin, C., and Klein, J.: Molecular genetics of speciation and human origins. Proceedings of the National Academy of Science 91: 6787-94, 1994

Bernatchez, L. and Landry, C.: MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? Journal of Evolutionary Biology 16: 363-77, 2003

Bonneaud, C., Chastel, O., Federici, P., Westerdahl, H., and Sorci, G.: Complex Mhc-based mate choice in a wild passerine. Proceedings of the Royal Society of London. Series B. 273: 1111-6, 2006

Bontrop, R. E., Elferink, D. G., Otting, N., Jonker, M., and de Vries, R. R.: Major histocompatibility complex class II-restricted antigen presentation across a species barrier: conservation of restriction determinants in evolution. Journal of Experimental Medicine 172: 53-9, 1990

Brunner, P. C., Douglas, M. R., Osinov, A., Wilson, C. C., and Bernatchez, L.: Holarctic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. Evolution 55: 573-86, 2001

Consuegra, S., Megens, H. J., Leon, K., Stet, R. J., and Jordan, W. C.: Patterns of variability at the major histocompatibility class II alpha locus in Atlantic salmon contrast with those at the class I locus. Immunogenetics 57: 16-24, 2005a

Consuegra, S., Megens, H. J., Schaschl, H., Leon, K., Stet, R. J., and Jordan, W. C.: Rapid evolution of the MH class I locus results in different allelic compositions in recently diverged populations of Atlantic salmon. Molecular Biology and Evolution 22: 1095-106, 2005b

Dick, T. and Belosevic, M.: Parasites of Arctic charr *Salvelinus alpinus* (Linnaeus) and their use in separating sea-run and non-migrating charr. Journal of Fish Biology 18: 339-347, 1980

Dijkstra, J. M., Kiryu, I., Yoshiura, Y., Kumanovics, A., Kohara, M., Hayashi, N., and Ototake, M.: Polymorphism of two very similar MHC class Ib loci in rainbow trout (*Oncorhynchus mykiss*). Immunogenetics 58: 152-67, 2006

Edgar, R. C.: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-7, 2004

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

Goldsby, R. A., Kindt, T. J., Osborne, B. A., and Kuby, J.: Kuby immunology. W.H. Freeman, New York, 2000

Grimholt, U., Drablos, F., Jorgensen, S. M., Hoyheim, B., and Stet, R. J.: The major histocompatibility class I locus in Atlantic salmon (*Salmo salar* L.): polymorphism, linkage analysis and protein modelling. Immunogenetics 54: 570-81, 2002

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebo, S., and Stet, R. J.: MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. Immunogenetics 55: 210-9, 2003

Gyllensten, U., Sundvall, M., Ezcurra, I., and Erlich, H. A.: Genetic diversity at class II DRB loci of the primate MHC. Journal of Immunology 146: 4368-76, 1991

Hall, T. A.: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. 41: 95-98, 1999

Johnson, L.: The Arctic charr, *Salvelinus alpinus*. In e. E. K. Balon (ed.): Charrs: salmonid fishes of the genus Salvelinus, pp. 19-98, Dr. W. Junk, The Hague, 1980

Jonsson, B. and Jonsson, N.: Polymorphism and speciation in Arctic charr. Journal of Fish Biology 58: 605-638, 2001

Kaufman, J., Salomonsen, J., and Flajnik, M.: Evolutionary conservation of MHC class I and class II molecules--different yet the same. Seminars in Immunology 6: 411-24, 1994

Kiryu, I., Dijkstra, J. M., Sarder, R. I., Fujiwara, A., Yoshiura, Y., and Ototake, M.: New MHC class Ia domain lineages in rainbow trout (*Oncorhynchus mykiss*) which are shared with other fish species. Fish and Shellfish Immunology 18: 243-54, 2005

Klein, J.: Of HLA, tryps, and selection: an essay on coevolution of MHC and parasites. Human Immunology 30: 247-58, 1991

Klemetsen, A., Amundsen, P., Dempson, B., Jonsson, B., Jonsson, N., O'Connell, F., and Mortensen, E.: Atlantic salmon Salmo salar L., brown trout *Salmo trutta* L. and Arctic charr *Salvelinus alpinus* (L.): a review of aspects of their life histories. Ecology of Freshwater Fish 12: 1-59, 2003

Knudsen, R.: Relationships between Habitat, Prey Selection and Parasite Infection in Arctic Charr (*Salvelinus alpinus*). Nordic Journal of Freshwater Research 71: 333-344, 1995

Knudsen, R., Kristoffersen, R., and Amundsen, P. A.: Parasite communities in two sympatric morphotypes of Arctic charr, *Salvelinus alpinus* (L.), in northern Norway. Canadian Journal of Zoology 75: 2003-2009, 1997

Lamont, S. J.: Impact of genetics on disease resistance. Poultry Science 77: 1111-8, 1998

Miller, K. M., Kaukinen, K. H., and Schulze, A. D.: Expansion and contraction of major histocompatibility complex genes: a teleostean example. Immunogenetics 53: 941-63, 2002

Miller, K. M., Li, S., Ming, T. J., Kaukinen, K. H., and Schulze, A. D.: The salmonid MHC class I: more ancient loci uncovered. Immunogenetics 58: 571-89, 2006

Miller, K. M. and Withler, R. E.: Mhc diversity in Pacific salmon: population structure and trans-species allelism. Hereditas 127: 83-95, 1997

Nei, M. and Gojobori, T.: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 3: 418-26, 1986

O'Brien, S. J. and Yuhki, N.: Comparative genome organization of the major histocompatibility complex: lessons from the Felidae. Immunological Reviews 167: 133-44, 1999

Reche, P. A. and Reinherz, E. L.: Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. Journal of Molecular Biology 331: 623-41, 2003

Shiina, T., Dijkstra, J. M., Shimizu, S., Watanabe, A., Yanagiya, K., Kiryu, I., Fujiwara, A., Nishida-Umehara, C., Kaba, Y., Hirono, I., Yoshiura, Y., Aoki, T., Inoko, H., Kulski, J. K., and Ototake, M.: Interchromosomal duplication of major histocompatibility complex class I regions in rainbow trout (*Oncorhynchus mykiss*), a species with a presumably recent tetraploid ancestry. Immunogenetics 56: 878-93, 2005

Shum, B. P., Guethlein, L., Flodin, L. R., Adkison, M. A., Hedrick, R. P., Nehring, R. B., Stet, R. J., Secombes, C., and Parham, P.: Modes of salmonid MHC class I and II evolution differ from the primate paradigm. Journal of Immunology 166: 3297-308, 2001

Shum, B. P., Mason, P. M., Magor, K. E., Flodin, L. R., Stet, R. J., and Parham, P.: Structures of two major histocompatibility complex class I genes of the rainbow trout (*Oncorhynchus mykiss*). Immunogenetics 54: 193-9, 2002

Shum, B. P., Rajalingam, R., Magor, K. E., Azumi, K., Carr, W. H., Dixon, B., Stet, R. J., Adkison, M. A., Hedrick, R. P., and Parham, P.: A divergent non-classical class I gene conserved in salmonids. Immunogenetics 49: 479-90, 1999

Stet, R. J., Kruiswijk, C. P., and Dixon, B.: Major histocompatibility lineages and immune gene function in teleost fishes: the road not taken. Critical Reviews in Immunology 23: 441-71, 2003

Potts, W. K. and Slev, P. R.: Pathogen-based models favoring MHC genetic diversity. Immunological Reviews 143: 181-97, 1995

Tamura, K., Dudley, J., Nei, M., and Kumar, S.: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596-9, 2007

Wilson, A. J., Gislason, D., Skulason, S., Snorrason, S. S., Adams, C. E., Alexander, G., Danzmann, R. G., and Ferguson, M. M.: Population genetic structure of Arctic charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales. Molecular Ecology 13: 1129-42, 2004

Yeager, M. and Hughes, A. L.: Evolution of the mammalian MHC: natural selection, recombination, and convergent evolution. Immunological Reviews 167: 45-58, 1999

Chapter 4. MH Class II α polymorphism in local and global adaptation of Arctic charr (*Salvelinus alpinus* L.)

Nucleotide sequence data reported are available in the GenBank databases under the accession numbers:

EF450325 to EF450340 and EF450348 to EF450455

Introduction

Arctic charr (*Salvelinus alpinus* L.) is the most northerly distributed freshwater fish (Johnson 1980). In many aspects of its distribution and life-history, Arctic charr are extremely complex, displaying remarkable ecological, morphological and genetic variation that locally adapt populations to the geographic areas or the environmental constraints of a particular niche (Alekseyev *et al*. 2002; Allendorf and Thorgaard 1984; Klemetsen *et al*. 2003). In some instances sympatric populations exist that may be reproductively isolated (Elliott and Baroudy 1995; Frost 1965; Hesthagen *et al*. 1995; Nordeng 1993; Skulason *et al*. 1989), giving raise to further identifiable variability. Given such variation, some authors (e.g. (Savvaitova 1995)) have reported a total of 56 sub-species, conventionally reduced to 4 major lineages (Behnke 1984). According to Behnke (1984), lineages include a group covering European and Greenland Arctic charr; a group for Northern Canada and Siberia, a group for Alaska and Western Russia, and a group for Eastern North America (Figure 4-1).

One particular point that makes difficult the definition of northern fish populations is that the existing niche partitioning has occurred in the most recent postglacial period, 10000-15000 years ago (Bernatchez *et al*. 2002). The short time interval poses a particular problem for meristic traits that diverge as consequence of adaptation to the environment, and often lead to conflicting results as molecular techniques for the identification of fish populations assume divergence with the differential accumulation of DNA mutations (Goldstein and Pollock 1997). Nevertheless, attempts to genetically characterize Arctic charr populations have been made (Gladden *et al*. 1995; Volpe and Ferguson 1996; Wilson *et al*. 2004), and most consent to the division of the

complex into 5 geographic lineages: An Atlantic lineage including Southeastern Canadian and European Arctic charr; an Arctic lineage for northern North America including western Greenland; a Beringian lineage for North Pacific Dolly Varden charr (*Salvelinus malma*); A Siberian lineage for Northern Russia, and an Acadian group for landlocked populations in south Québec and Maine (Brunner *et al*. 2001) (Figure 4-1).

Here we characterized the Arctic charr complex using a molecular marker that diverges temporarily between populations by mutation accumulation, and by the adaptation of populations when settling in particular niches. The Major Histocompatibility (MH) genes encode for membrane receptors, involved in the distinction among self and non-self particles. MH receptors present fragments of pathogens to T-lymphocytes in order to initiate immune responses (Steinmetz and Hood 1983). Because there are thousands of potential pathogens, MH genes are the most polymorphic coding regions known in vertebrates (Grimholt *et al*. 2003; Reche and Reinherz 2003). It has been observed that some alleles of MH proteins are more effective than others in presenting particular antigens and, therefore, they provide better resistance to the individuals that carry them (Dawkins *et al*. 1999; Lohm *et al*. 2002; Messaoudi *et al*. 2002). The high polymorphism of MH molecules in a population is meant to increase species survival probabilities to an eventual disease outbreak and, consequently, the individuals carrying the right MH alleles will persist in time (Bernatchez and Landry 2003; Bonneaud *et al*. 2006; Lamont 1998). Therefore, the allele composition of a particular group of individuals is based on the genetic line of the group as modified by the selective pressure exerted by the pathogens in their environment (Bryja *et al*. 2006; de Eyto *et al*. 2007).

There are two types of MH receptors present in most jawed vertebrates, MH Class I and MH Class II proteins. The later are composed of two subunits: Class II α and Class II β. Each one has

5 domains, normally codified by 5 separated exons. The second exon encodes the α 1 and β 1 domains, respectively, that bind the antigen to be presented to T-lymphocytes (Peptide Binding Region) and, therefore, contain the highly variable regions of the molecule (Steinmetz and Hood 1983).

For the present work, we used the Class II α gene as a molecular marker, as an initial approach for characterizing the diversity in Arctic charr populations. The Class II α gene has been reported to be highly divergent among populations, but is still strongly selected by nature (Bryja *et al*. 2006; Hardee *et al*. 1995; Piertney and Oliver 2006; Stet *et al*. 2002), thus we expected the Class II α gene to be capable of differentiating populations with recent coancestry, as is likely the case for Arctic charr populations that have recently colonized the greater part of their current distribution (Johnson 1980).

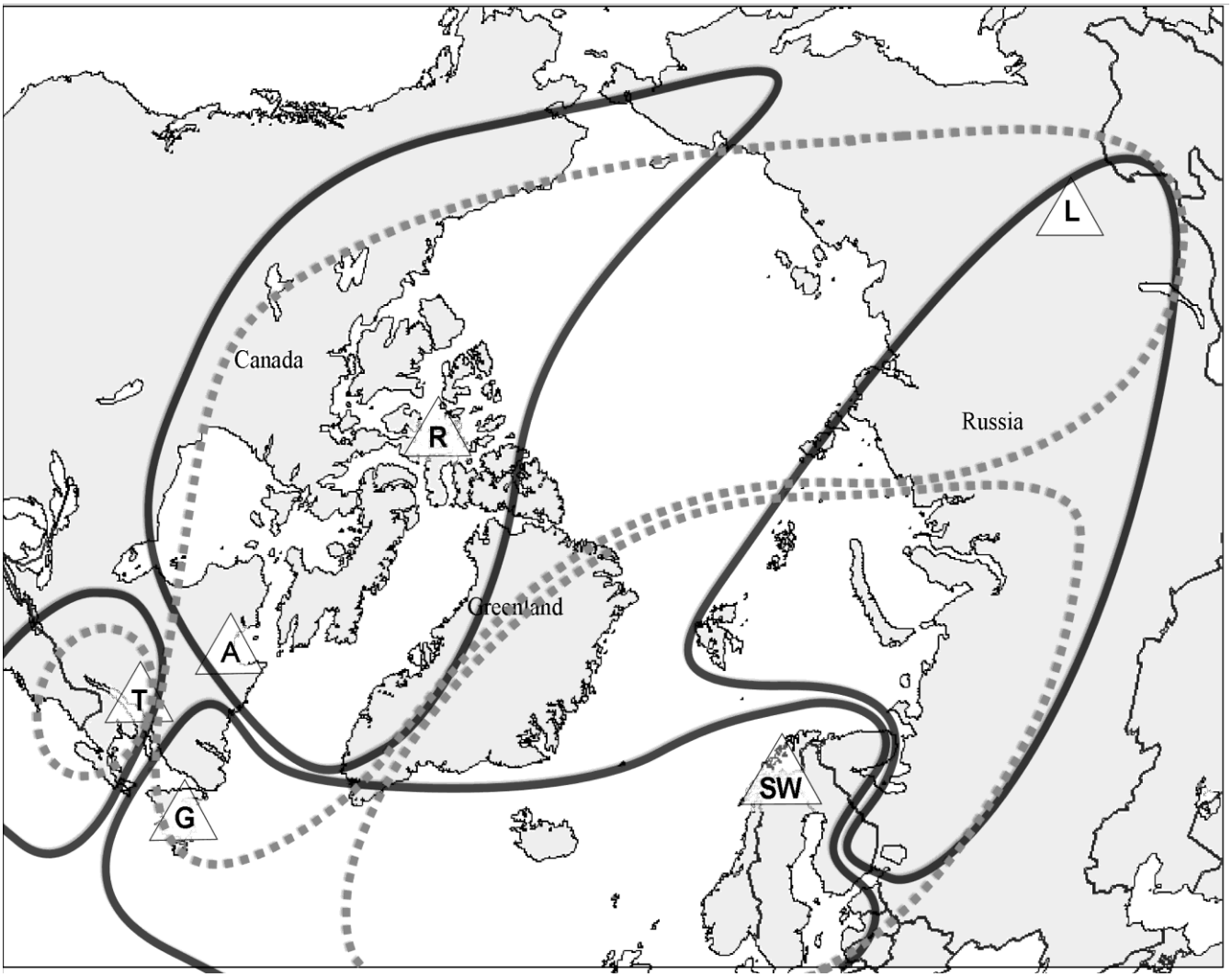


Figure 4-1. Arctic charr phylogeographic lineages. The map shows with dashed lines the limits of the geographic lineages classified using meristic characters (Behnke 1984) and in solid lines areas the geographic extent of lineages according to mDNA sequencing (Brunner *et al*. 2001), with the exception of the Beringian lineage for *Salvelinus malma* which was not included in this study. The populations sampled for this study are indicated within triangles. The three Russian populations are separated by approximately 30 km and therefore cannot be resolved on this map. The names represent the location of each population: A for Lake Aigneau, G for Gander Lake, R for Resolute Lake, SW for Lake Sitasjaure, T for de la Trinité River, L for Lake Kiryalta-3, Lake Kiryalta-4 and Lake Kamkanda.

Materials and Methods

Fish samples

Muscle samples of fresh caught Arctic charr were immersed in 70% ethanol or RNA preserving solution (25 mM $C_6H_5NaO_7$, 10 mM EDTA and 4 M $((NH4)_2SO_4)$ and shipped to our laboratory from various sampling sites. Sites included Lake's Kiryalta-3 (57°08`N; 119°27`E), Lake Kiryalta-4 (57°07`; 119°28`E) and Lake Kamkanda (57°06`N; 119°48`E) in the Transbaikal region of eastern Russia; Lake Sitasjaure (68°00`N; 17°25`E) in Sweden; and Lake Aigneau (57°14`N; 70°07`W) and the de la Trinité River (49°25`N; 67°18`W) in Québec, Gander Lake, Newfoundland (48°55`N; 54°35`W); and Resolute Lake (74°41`N; 94°57`W) in Nunavut, Canada (Figure 4-1). Sample numbers for each population are shown in Table 4-1.

DNA extraction

Samples in ethanol and RNA preserving solution were placed in TE buffer (100 mM Tris, 1mM EDTA) for 2 hrs, and then incubated for 3 hrs in 100 mM Tris, 10 mM EDTA, 240 mM NaCl, 1% SDS and 300 ug/ml Proteinase K at 55 ºC. This was followed by two extractions with one volume of phenol and one extraction with one volume of chloroform. The DNA was precipitated with 1/10 volumes of 3 M CH_2COONa pH 5.2 and 2 volumes of Ethanol, and then washed with 70% ethanol.

PCR and cloning

PCR forward (SAALDAAF: 5' - CTG GAT GCA GTG ATT CAG ATG – 3') and PCR reverse (SAALDAAR: 5'- GAC GTG GCA GAT GAG AGT G – 3') primers were designed to match the beginning of the second exon, and the beginning of the third exon, respectively. These primers were based on conserved regions of the MH genes Class II α published for *Salmo salar* (L77090.1; AY780912.1) and *Oncorhynchus mykiss* (AJ251432.1; AY305398.1), as well as full length sequences of Arctic charr (Chapter 2).

The conditions for the PCR were at 95 °C of initial denaturation for 2 min, plus 35 cycles of 95 °C of denaturation for 30 sec, 58 °C of primer annealing for 40 sec and 72 °C of extension for 1 min. A final extension at 72 °C for 8 min was done to ensure the adding of an A overhang at the end of the complementary strand.

The PCR amplicons were separated by agarose electrophoresis in a 1% gel to verify size and then extracted directly from the gel using GenElute Agarose Spin Columns (Sigma-Aldrich, St Louis, MO). The fragments were inserted in pGEM-T vectors as per manufacturer instructions (Promega Corporation, Madison, WI). XL-blue strains of *E. coli* bacteria were made competent by washing them twice and then re-suspending them with, Thawing buffer (10mM PIPES, 15mM CaCl-$2H_2O$, 250 mM KCl, 55mM $MnCl_2$-$4H_2O$, pH 6.7). The bacteria were transformed by the Inoue procedure (Sambrook *et al*. 2001) and grown in LB plates with ampicillin (100 ug/ml) plus 1 uM IPTG and 100 ug/ml X-gal. Eight to twelve white colonies from each amplicon were selected and grown in LB media. The bacteria were used for PCR with SAALDAAF primer and a new Reverse primer for Denaturant Gradient Gel Electrophoresis (DGGE), designed to match the end of the second exon (SAALDAADGGER: CGG GCG GGG GCG GCG GGC

CGG GCG CGG GGC GCG GCG GGC GCT ATT GTC TCT GGT GGG TYC). The amplicons were separated in a denaturing gradient of 25-50% for 15.5 hrs at 60 °C (Bio-Rad laboratories, Richmond, CA). Only the alleles that differed between samples were selected for DNA sequencing in order to avoid redundancies. Those plasmids selected for sequencing were extracted from the grown bacteria using GenElute Plasmid Miniprep Kit (Sigma-Aldrich, St Louis, MA) according to manufacturer instructions, and sequenced using the T7 and Sp6 primers with a big dye terminator v3.1 (Applied Biosystems, Foster City, CA) in a 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA).

Sequencing and data analysis

The sequences obtained were aligned using the software Muscle (Edgar 2004). An alignment of the previously sequenced full length Class II α cDNA (Chapter 2), aided in the definition of the end of the second exon and the beginning of the intron. The intron and exon sequences were analyzed separately. BioEdit (Hall 1999) was used to define the open reading frames and to translate them into protein sequences. This software was also used to obtain entropy values that measure the variability of each amino acid residue position (Shannon, 1948 *fide in* Reche and Reinherz 2003). We considered highly entropic sites, having values ≥0.5, approximately 35% of the maximum observed entropy in the molecule, according to similar criteria used by other authors (Consuegra *et al*. 2005a; Consuegra *et al*. 2005b; Reche and Reinherz 2003). To avoid errors caused by *Taq* polymerase during the PCR reaction, MEGA 3 (Kumar *et al*. 2004) was used to find and correct sequences containing singletons. MEGA 3 was also used to calculate the diversity of the sequences between and within populations, the rates of non-

synonymous/synonymous substitutions, and to draw the phylogenetic trees, using the Kimura-2 parameter model (Kimura 1980) for the intron sequences and the Nei-Gojorobi model (Nei and Gojobori 1986) Jukes-Cantor corrected (Jukes and Cantor 1969) for non-synonymous substitutions in the exon sequences.

Alrequin 3.01 software (Excoffier *et al*. 2005) was used to calculate the allele frequencies which in this case are defined as MH Class II α allelic frequencies, the F-statistics (F_{ST}), the exact test for population differentiation and to obtain the values for the Watterson (1978) and Slatkin and Excoffier (1996) tests of deviations from the expected homogeneity of the haplotype distribution under neutral conditions.

Results

Arctic charr MH Class II α genes contain an Hpa retrotransposon

The primers used for amplifying the second exon and intron of the gene yielded two characteristic fragment sizes, with little variation among all the samples studied: a fragment of approximately 800 bp, and a smaller fragment of 600 bp. Fragment sizes were found to be related to alleles rather than loci related, as were not always present simultaneously in every fish. The difference in length was due to a putative retrotransposon present in the second intron that has been found to be highly conserved in other genes from several other salmonid species (Kido *et al*. 1995; Takasaki *et al*. 1994). The transposon, an Hpa repeat, is far less variable than the rest of intron (0.007 versus 0.037) and has not previously been found in the Class II α gene of any other vertebrate. The transposon was present in 28% of the samples, and was absent from all but one Russian sample and from all the de la Trinité River fish.

Class II α haplotypes and Charr population diversity

From the 74 samples studied (Table 4-1) we found 33 different exon alleles (Figure 4-2) with an average p-distance of 0.082, 0.069 corresponding to intrapopulation diversity and 0.013 to diversity between populations. The most common allele had a frequency of 0.13 and there were 14 alleles that were present only in one sample. Some haplotypes were shared among the geographic lineages (Table 4-2) but most were unique for each group.

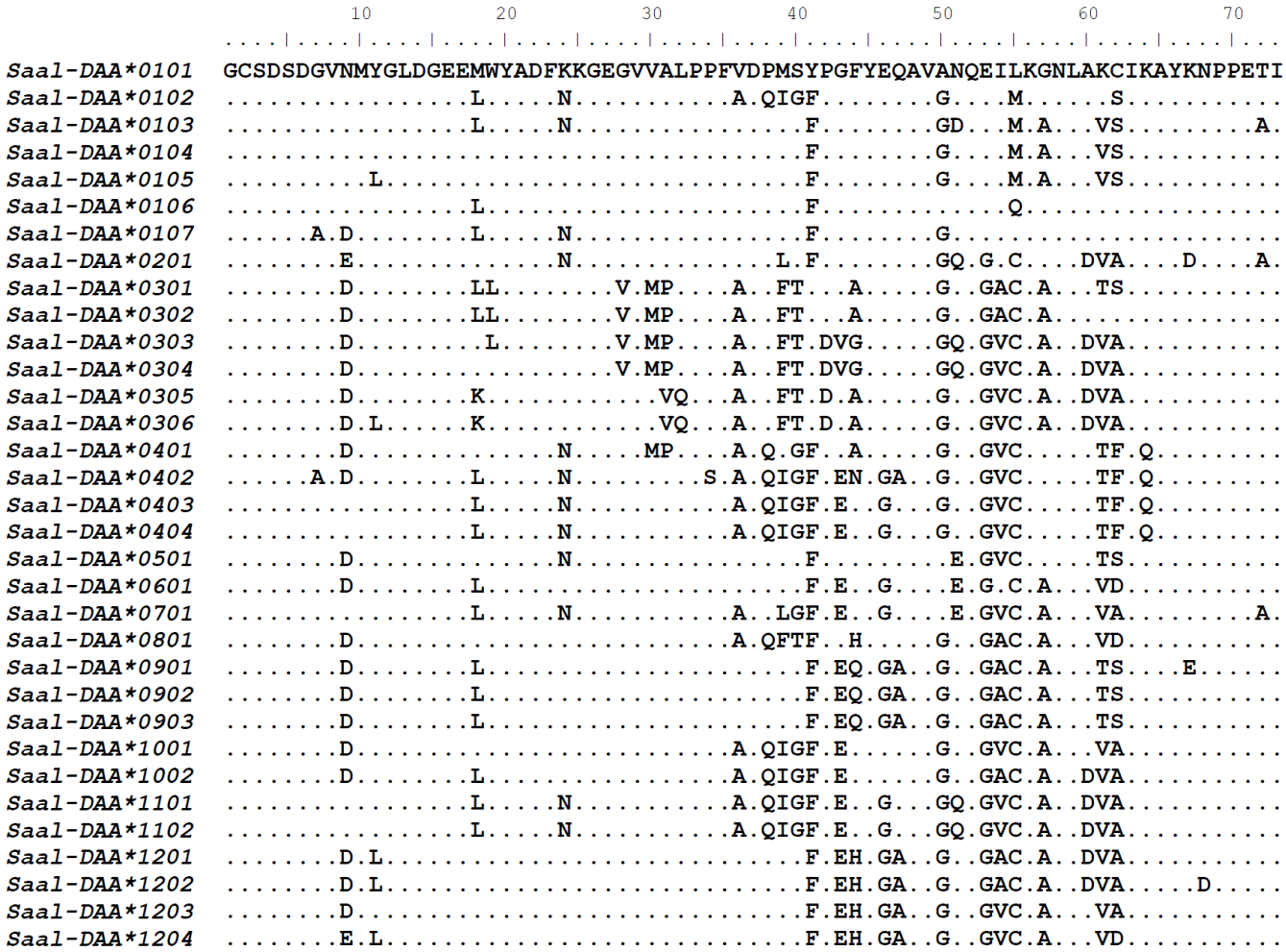


Figure 4-2. MH Class II α alleles.

The translated sequences of the alleles found in this study for the second exon of the MH Class II α gene in Arctic charr.

Population	Samples	Exon variation	Intron variation	Subst: Non-synonymous	Subst: Synonymous	PBR variation	Watterson p values	Slatkin p values
Aigneau	9	0.081	0.037	0.096	0.032	0.216	0.097	0.097
Gander	13	0.078	0.030	0.092	0.030	0.216	0.033	0.034
Kamkanda	3	0.066	0.029	0.073	0.042	0.199	1.000	1.000
Kiryalta-3	14	0.051	0.040	0.061	0.018	0.155	0.360	0.500
Kiryalta-4	5	0.052	0.022	0.064	0.011	0.161	0.530	0.530
Resolute	10	0.086	0.033	0.101	0.037	0.235	0.570	0.570
Sitasjaure	11	0.078	0.028	0.090	0.034	0.203	0.510	0.600
Trinité	9	0.066	0.009	0.074	0.041	0.197	0.110	0.110

Table 4-1. MH Class II α second exon divergence in the populations.

The table shows the sample number for each population studied. It also shows the diversity for each one of the parameters used in the analysis: the exon and intron mean p-distances, the proportion of non-synonymous and synonymous substitutions and the putative PBR variation. The p-values for the Watterson F and Slatkin's exact tests for deviation from neutrality are also shown for each population.

There were 21 amino acid positions with high entropy (≥0.5) in the α 1 domain, which when considered together had an average nucleotide p-distance of 0.24. 11 of the 21 residues corresponded to variable residue positions found in HLA-DRA (out of a total of 20 HLA-DRA polymorphic sites). All 11 occurred within the residues that correspond to the known human Peptide Binding Region (Brown *et al*. 1993), although an exact correlation is difficult to establish since the sequence similarity between the human and Arctic charr peptide sequences is less than 25%.

Within the Arctic charr included in this study, each population had a set of alleles which differed in the number of highly variable amino acids from a minimum of 13 residues to a maximum of 21, each with variability in position along the peptide chain (Table 4-3). These variable residues were mostly concentrated between the 36th and the 63rd position of the α 1 domain, with 8 positions at which all populations showed some degree of entropy. Position 9 was the only one outside this area that showed the same pattern. The other variable positions were more population-specific, and tended to group geographically, usually with patterns shared by Canadian/Swedish populations. Extreme cases were the Lake Aigneau population that possessed several unique variable residues, and the de la Trinité River and Lake Kiryalta populations that lacked variability at many of the common variable sites.

Only 13 different intron sequences were found in this study, so different exons shared a common adjacent intron. The nucleotide p-distance between introns was also lower than in the exons, with an average of 0.037, with 0.029 of intrapopulation diversity.

	Arctic	Atlantic	Siberian	Acadian
Arctic		0104, 1203, 1201, 0801, 0304	0701, 0404	0304, 0401
Atlantic	5		1202, 0103, 0107	0304
Siberian	2	3		-
Acadian	2	1	0	

Table 4-2. MH Class II α second exon alleles shared among lineages.

Number of shared exon allele sequences among geographic lineages. Bellow diagonal: Number of Saal-DAA Alleles shared. Above diagonal: identification numbers of the shared alleles.

Table 4-3. MH Class II α, α 1 domain amino acid entropy.

The table shows the entropy values for each amino acid positions, for all the populations. Shaded are the values with entropy ≥0.5, which is equivalent to approximately 35% of the maximum observed entropy. These sites potentially correspond to peptide binding residues of the MH Class II α molecule. In the first column the amino acid position is shown, and in the next the populations from de la Trinité River (Tri), Lake Aigneau (Aig), Resolute Lake (Res), Gander Lake (Gan), Lake Kiryalta-4 (Kirj4), Lake Kiryalta-3 (Kirj3) and Lake Kamkanda (Kam).

Aa	Tri	Aig	Res	Gan	Sit	Kirj4	Kirj3	Kam
1	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0.52	0	0	0	0	0	0.45
8	0	0	0	0	0	0	0	0
9	0.69	0.60	1.02	0.49	1	0.69	0.60	0.45
10	0	0	0	0	0	0	0	0
11	0	0.41	0.55	0	0.56	0	0	0
12	0	0	0	0	0	0	0	0
13	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0
18	0.69	0.68	0.96	0.78	0.90	0.64	0.56	0
19	0	0.41	0.36	0.49	0	0	0.17	0
20	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0
24	0	0.65	0.47	0.68	0.50	0.35	0.29	0
25	0	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	0
27	0	0	0	0	0	0	0	0
28	0	0.41	0.36	0.55	0	0	0.17	0
29	0	0	0	0	0	0	0	0
30	0	0.41	0.47	0.55	0	0	0.17	0
31	0.69	0.41	0.80	0.84	0.61	0	0.17	0
32	0.69	0	0.36	0.31	0.61	0	0	0
33	0	0	0	0	0	0	0	0
34	0	0.52	0	0	0	0	0	0
35	0	0	0	0	0	0	0	0
36	0.69	0.68	0.69	0.69	0.65	0.69	0.56	0.45
37	0	0	0	0	0	0	0	0

Aa	Tri	Aig	Res	Gan	Sit	Kirj4	Kirj3	Kam
38	0	0.60	0.55	0.41	0.20	0.69	0.66	0.64
39	0.69	1.28	0.96	0.69	1.14	0.69	0.98	1.01
40	0.69	1	1.02	0.69	0.79	0.69	0.72	0.45
41	0	0.41	0.61	0.64	0.61	0	0.17	0
42	0.69	0	0.36	0.64	0.61	0	0.17	0
43	0	0.60	0.69	0.84	0.50	0.64	0.67	0.64
44	0.69	1.49	1.23	1.27	0.97	0.35	0.34	0
45	0	0	0	0	0	0	0	0
46	0	0.65	0.61	0.31	0.50	0.64	0.56	0.64
47	0	0.69	0.61	0.31	0.42	0.35	0.17	0
48	0	0	0	0	0	0	0	0
49	0	0	0	0	0	0	0	0
50	0.69	0.41	0.36	0.55	0.20	0.64	0.60	0.69
51	0	0.41	0.36	1.37	0.61	1.06	1.09	0.69
52	0	0	0	0	0	0	0	0
53	0.69	0.41	0.55	0.49	0.65	0	0	0.64
54	0.69	0.99	1.07	0.86	0.97	0.35	0.17	0.64
55	0.69	0.41	0.79	0.49	0.65	0	0	0.87
56	0	0	0	0	0	0	0	0
57	0.69	0.52	0.47	0.55	0.50	0.69	0.69	0.69
58	0	0	0	0	0	0	0	0
59	0	0	0	0	0	0	0	0
60	0.69	0.41	0.61	0.64	0.67	0.69	0.64	0
61	0.69	0.68	0.89	0.55	0.20	0.69	0.68	1.01
62	0.69	1.06	1.51	1	0.82	0.96	1.08	1.24
63	0	0	0	0	0	0	0	0
64	0	0.52	0.22	0	0	0.35	0.64	0.45
65	0	0	0	0	0	0	0	0
66	0	0	0	0	0	0	0	0
67	0	0	0	0	0.42	0.35	0	0
68	0	0	0	0	0.20	0	0	0
69	0	0	0	0	0	0	0	0
70	0	0	0	0	0	0	0	0
71	0	0	0	0	0	0	0	0
72	0	0.41	0.36	0.49	0.42	0	0.29	0.69
73	0	0	0	0	0	0	0	0

Balancing Selection

For exon 2, the ratio of non-synonymous to synonymous substitutions was 2.7, which is in the expected range for alleles under balancing selection (Satta 1997). This was the case for all the populations studied, with a maximum of 3.1 for Gander Lake and a minimum of 1.8 for Lake Kamkanda (Table 4-1). The putative PBR coding nucleotides had a rate of 3.4 when examined separately, with a maximum in Lake Kiryalta-4 of 8.4 and a minimum of 2 in de la Trinité River. The mean p-distance at synonymous sites was 0.036, a very similar value to the total variation in the introns, suggesting that synonymous variation is due to background mutations. Watterson (Watterson 1978) and Slatkin (Slatkin and Excoffier 1996) tests, which indicate deviations from neutrality, were significant only for Gander Lake population ($p<0.05$). The rest of the populations produced values close to neutrality, excepting Lake Kamkanda, which showed values indicative of purifying selection, although this could be an artifact of the low sample size.

Population differentiation

The pairwise F_{ST} differentiation values among populations according to the second exons is shown in Table 4-4A. The p-values for the FST and Exact test of population differentiation are shown in Table 4-4B. At a significance level of $p<0.05$, all phylogeographic lineages of Arctic charr, as sampled in the current study, are isolated from each other. No populations are isolated within the groups for the F-statistics, however, the exact text of population differentiation, based on haplotype frequencies, report isolation in most populations within the geographic groups.

Intron F_{ST} on the other hand were unable to demonstrate population separations to the same degree the exon data does, but established a division between Arctic-Atlantic, Siberian and

Acadian populations (data not shown). This difference in the exon and intron pairwise F_{ST} p-values indicates that the separation among northern Canadian groups is exon-driven, suggesting that selection pressure, more than genetic drift, affects their allele distribution.

Phylogenetic inference

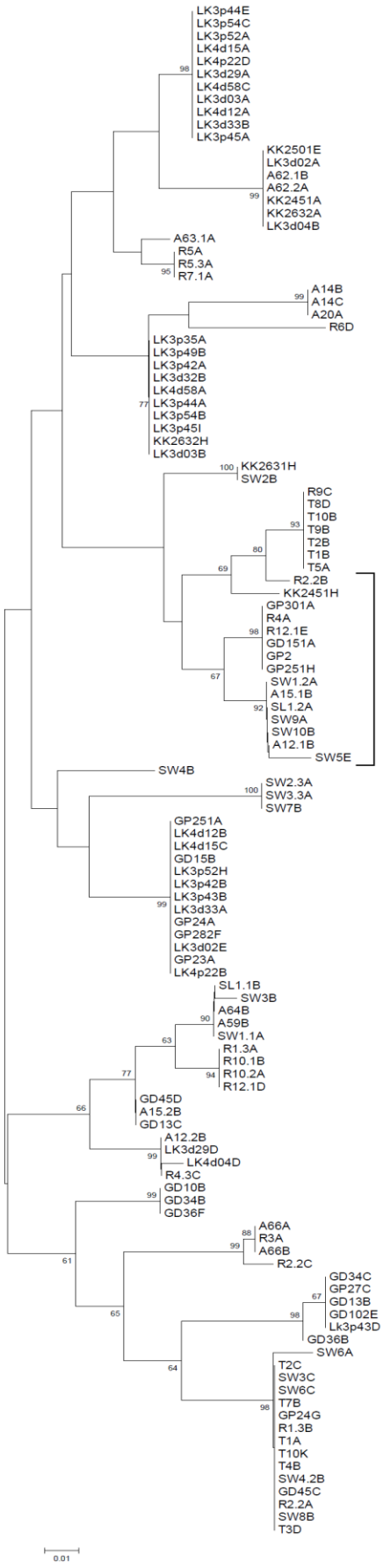
The phylogenetic tree constructed for the exons did not show major monophyletic groups (Figure 4-3, left). Instead, it showed a homogeneous distribution for all the alleles, which may indicates alleles belonging to a single locus. Nevertheless, the exons clustered relative to their geographic distribution, but usually these did not belong to any higher phylogenetic clade. Therefore, none of the highly bootstrapped clusters defined any distinct geographic lineage, although many of them clearly defined specific populations.

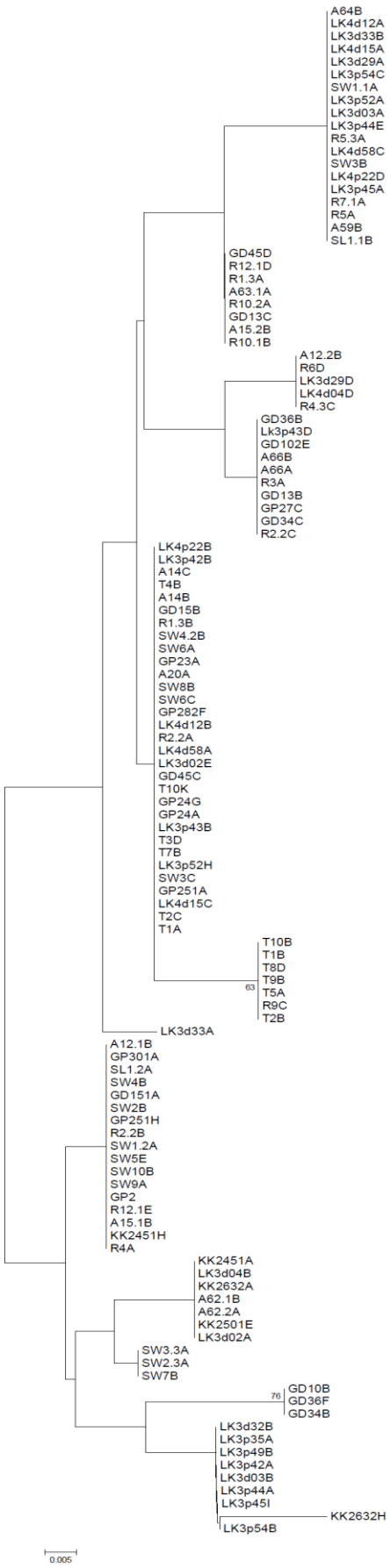
All but 7 alleles were grouped in high-bootstrapped clusters. Eight of those clusters were geographically related. Two groups contained only Canadian alleles. Another 2 clusters consisted exclusively of Lake Kiryalta-3 and -4 samples, containing 21 of the 33 alleles isolated from those populations. Two other groups contained 7 of the 17 alleles from Resolute Lake. There was a single homogeneous group for Lake Aigneau alleles, one for Lake Sitasjaure and one for Gander Lake. Finally there was a cluster of almost exclusively de la Trinité River alleles, containing 6 of the 13 alleles from that population. All other groups included alleles coming from a number of populations and were in general not associated with geographic distribution.

The exons of the alleles that contain the Hpa transposon in the intron 2, also clustered closely, most of them creating two related groups. One group contained Resolute Lake and Gander Lake alleles and another contained Lake Sitasjaure and Lake Aigneau alleles. Another 3 Hpa

containing sequences were dispersed in other clusters, but still within close levels of divergence from the major group mentioned above.

Figure 4-3. MH Class II α phylogenetic reconstruction. Left: The phylogenetic reconstruction using the sequenced second exons with the Nei-Gojobori model, which considers only non-synonymous mutations with a Jukes-Cantor correction. The bracket indicates the alleles containing the Hpa retrotransposon in their introns. Right: The phylogenetic reconstruction using the sequenced second introns with the Kimura-2 model. The trees were bootstrapped 1000 times and only bootstraps over 60% are shown. The names represent series of each population starting with an A for Lake Aigneau, G for Gander Lake, R for Resolute Lake, SW for Lake Sitasjaure, T for de la Trinité River, LK3 for Lake Kiryalta-3, LK4 for Lake Kiryalta-4 and KK for Lake Kamkanda.


LK3p44E
LK3p54C
LK3p52A
LK4d15A
LK4p22D
LK3d29A
LK4d58C
LK3d03A
LK4d12A
LK3d33B
LK3p45A
KK2501E
LK3d02A
A62.1B
A62.2A
KK2451A
KK2632A
LK3d04B
A63.1A
R5A
R5.3A
R7.1A
A14B
A14C
A20A
R6D
LK3p35A
LK3p49B
LK3p42A
LK3d32B
LK4d58A
LK3p44A
LK3p54B
LK3p45I
KK2632H
LK3d03B
KK2631H
SW2B
R9C
T8D
T10B
T9B
T2B
T1B
T5A
R2.2B
KK2451H
GP301A
R4A
R12.1E
GD151A
GP2
GP251H
SW1.2A
A15.1B
SL1.2A
SW9A
SW10B
A12.1B
SW5E
SW4B
SW2.3A
SW3.3A
SW7B
GP251A
LK4d12B
LK4d15C
GD15B
LK3p52H
LK3p42B
LK3p43B
LK3d33A
GP24A
GP282F
LK3d02E
GP23A
LK4p22B
SL1.1B
SW3B
A64B
A59B
SW1.1A
R1.3A
R10.1B
R10.2A
R12.1D
GD45D
A15.2B
GD13C
A12.2B
LK3d29D
LK4d04D
R4.3C
GD10B
GD34B
GD36F
A66A
R3A
A66B
R2.2C
GD34C
GP27C
GD13B
GD102E
Lk3p43D
GD36B
SW6A
T2C
SW3C
SW6C
T7B
GP24G
R1.3B
T1A
T10K
T4B
SW4.2B
GD45C
R2.2A
SW8B
T3D
0.01


A64B
LK4d12A
LK3d33B
LK4d15A
LK3d29A
LK3p54C
SW1.1A
LK3p52A
LK3d03A
LK3p44E
R5.3A
LK4d58C
SW3B
LK4p22D
LK3p45A
R7.1A
R5A
A59B
SL1.1B
GD45D
R12.1D
R1.3A
A63.1A
R10.2A
GD13C
A15.2B
R10.1B
A12.2B
R6D
LK3d29D
LK4d04D
R4.3C
GD36B
Lk3p43D
GD102E
A66B
A66A
R3A
GD13B
GP27C
GD34C
R2.2C
LK4p22B
LK3p42B
A14C
T4B
A14B
GD15B
R1.3B
SW4.2B
SW6A
GP23A
A20A
SW8B
SW6C
GP282F
LK4d12B
R2.2A
LK4d58A
LK3d02E
GD45C
T10K
GP24G
GP24A
LK3p43B
T3D
T7B
LK3p52H
SW3C
GP251A
LK4d15C
T2C
T1A
T10B
T1B
T8D
T9B
T5A
R9C
T2B
LK3d33A
A12.1B
GP301A
SL1.2A
SW4B
GD151A
SW2B
GP251H
R2.2B
SW1.2A
SW5E
SW10B
SW9A
GP2
R12.1E
A15.1B
KK2451H
R4A
KK2451A
LK3d04B
KK2632A
A62.1B
A62.2A
KK2501E
LK3d02A
SW3.3A
SW2.3A
SW7B
GD10B
GD36F
GD34B
LK3d32B
LK3p35A
LK3p49B
LK3p42A
LK3d03B
LK3p44A
LK3p45I
KK2632H
LK3p54B
0.005

A	Aigneau	Gander	Kamkanda	Resolute	Sitasjaure	Trinité	Kirjalta3	Kirjalta4
Aigneau	*							
Gander	0.092	*						
Kamkanda	0.099	0.178	*					
Resolute	-0.008	0.037	0.129	*				
Sitasjaure	0.120	0.043	0.216	0.034	*			
Trinité	0.265	0.147	0.345	0.144	0.092	*		
Kirjalta3	0.150	0.241	0.097	0.208	0.310	0.430	*	
Kirjalta4	0.103	0.172	0.100	0.145	0.241	0.391	-0.038	*

B	Aigneau	Gander	Kamkanda	Resolute	Sitasjaure	Trinité	Kiryalta3	Kiryalta4
Aigneau	*	<0.005	0.19	<0.005	0.01	<0.005	<0.005	<0.005
Gander	0.02	*	<0.005	<0.005	<0.005	<0.005	<0.005	0.01
Kamkanda	0.07	0.02	*	0.01	<0.005	<0.005	0.03	0.01
Resolute	0.51	0.11	0.02	*	<0.005	<0.005	<0.005	<0.005
Sitasjaure	0.01	0.09	<0.005	0.12	*	<0.005	<0.005	<0.005
Trinité	<0.005	0.01	<0.005	0.01	0.05	*	<0.005	<0.005
Kiryalta-3	<0.005	<0.005	0.08	<0.005	<0.005	<0.005	*	0.47
Kiryalta-4	0.03	0.01	0.12	<0.005	<0.005	<0.005	0.74	*

Table 4-4. Population differentiation according to the second exon of the MH Class II α gene.

A: The population pairwise F_{ST} distance values. B. Bellow the diagonal: The p-values for the pairwise F_{ST} analysis (5000 permutations). Above the diagonal: The p-values of the exact test of differentiation using haplotype frequencies (6000 Markov steps). Values ≤0.05 were considered significant for this study and are indicative of the isolation of the populations from each other (shaded).

Discussion

MH allele variation in Arctic charr populations

Arctic charr have the broadest geographical distribution of all salmonids, and in addition have colonized the greatest variety of ecological niches, most of which are nutrient poor (Alekseyev *et al*. 2002). The Arctic charr demonstrates an exceptional capacity for morphological plasticity (Johnson 1980) that allows fast adaptation. The Arctic charr adapts by dividing the original population into habitat and feeding specialized units that are able to exploit habitat-specific resources. With habitat and feeding specialization, however, individuals are likely to be exposed to a variety of new pathogens specific to the environment being exploited (Knudsen 1995). Individuals should therefore be selected for characters that make them more resistant to new pathogens, with the consequence that MH Class II α alleles that are more efficient at presenting peptides from those pathogens and initiating immune responses will become more prevalent in succeeding generations.

Our data suggest that MH class II α allelic divergence and adaptation happened primarily during or subsequent to the speciation of the Arctic charr from other salmonids, and that while later adaptation has occurred, it has been modest. This is reflected by the non-existence of major phylogenetic clusters representing global lineages and also by the intra-population diversity accounting for most of the Arctic charr MH Class II α variation. Although ancestral alleles are maintained in most populations, there still exist differential proportions of those alleles in each population. In addition, several site-specific alleles support the significant isolation of the

lineages as previously identified by mDNA sequencing (Brunner *et al*. 2001) and the isolation of populations within lineage groups.

Because MH Class II α genes are strongly affected by natural selection exerted by the local pathogenic environment of each site, important differences are observed with respect to the actual phylogeny and genotypic distances of the studied populations. For example, the Atlantic Arctic charr group, found to have the lowest genetic divergence with neutral markers (Brunner *et al*. 2001), has above-average MH Class II α divergence which is probably related to the wide geographic extent of this group. Still, F-statistics do not indicate isolation of the populations on each side of the Atlantic, suggesting a global distribution of most alleles. As another example, Atlantic and Arctic MH Class II α alleles form isolated groups and the Atlantic Gander Lake population is not related to the Arctic group as a whole, but Gander Lake Arctic charr does not appear to be separated from Resolute Lake Arctic charr when using F_{ST} p-values.

In a last example, the Acadian population is very diverse when examined using a neutral marker and was found to also have reduced divergence in MH haplotypes. The lower divergence might represent strong selection for specific alleles that confront particular pathogens from the local area.

The Arctic group is genetically the most diverse according to mDNA, likely as a consequence of multiple postglacial invasions sites (Maiers *et al*. 1995). The same result is also obtained for the MH Class II α genes using F-statistics, and just as with the neutral marker, the populations sampled were not isolated from each other.

Although the pairwise F_{ST} comparison did not identify separated populations within previously proposed geographic lineages, there are indications of ongoing adaptation processes as evidenced

by differential haplotype frequencies. The exact test of population differentiation (Table 4-4B) indicates isolation of all populations from each other with the exception of Lakes Kiryalta-3 and -4, and Lake Kamkanda and Lake Aigneau. The first case is an expected result, since both Lakes Kiryalta are connected periodically, and separated by less than 500 meters. The second result may have arisen via the evolutionary convergence of the Lake Kamkanda and Lake Aigneau populations, where the alleles have been maintained in similar proportions. The similarity reflects more traditional lineage groupings that normally classified populations from Northern Canada and Siberia into a single group (Behnke 1984).

It is interesting to note that the exact test is capable of differentiating the populations from Lake Kiryalta and Lake Kamkanda, that are separated by only 20 km and form part of the same drainage basin. The outstanding sensitivity of the method confirms the fact that MH allele distributions can vary significantly even between closely related groups, suggesting that both populations have been exposed to different pathogenic environments and have been rapidly selected for divergent MH Class II α alleles.

Intron evolution

The intron divergence was much lower than observed for the exons and very similar to the variation found for the synonymous substitutions. Given the low variety of intron haplotypes and the fact that some contain different exons, there might not be strong evolutionary linkage among them (Maynard and Haigh 1974). The result suggests background mutations have been the main force affecting their evolution. Nevertheless, the Hpa repeat insertion has had a strong effect on the evolution of the alleles that contain it, causing them to cluster together. The divergence

between the two types of introns, with or without the insertion, is probably large enough to prohibit crossing over, thereby creating a novel Class 2 α lineage. We also found that the variation in the Hpa repeat was much lower than that in the rest of the intron, which may have contributed to the clustering observed in the exon tree, as the repeat may serve as a conserved hot spot for recombination among the alleles that possess it. The existence of recombination hotspots in introns adjacent to the variable exons on the MH genes has been discussed in other species (Ono *et al*. 1993; Schaschl *et al*. 2005), and the possibility that retroelements might be related to the variability of MH genes has been postulated already (Dawkins *et al*. 1999). However, here we present for the first time a putative retrotransposon that has directly influenced the divergence on the adjacent exon. The result reinforces the importance of these elements in the evolution of the MH genes. The Hpa insertion and the consequent creation of this allelic lineage must have occurred fairly recently, since its presence defines a unique intron haplotype, that is associated with exons with reduced divergence compared to the overall population variation (Figure 4-3, left). The fact that the Hpa insert is present in both Atlantic and Arctic samples leads us to believe that it was introduced before the global dispersal of Arctic charr. The Hpa repeat insertion is present in alleles from all populations with the exception of Lake Kiryalta -3 and -4 and de la Trinité River. Considering that neutral genetic data suggests later emergence of both populations with respect to the Arctic lineage (Brunner *et al*. 2001), then the alleles containing the Hpa repeat may have been negatively selected by the environment at these two locations. Coincidently, these two populations are also the least diverse, suggesting that the presence of the Hpa transposon might be positively correlated with the variability of the MH molecules in the other populations, or alternatively, that the lower divergence found in Lake Kirjalta -3 an -4 and de la Trinite river is consequence of out selecting the Hpa containing haplotypes.

Amino acid variation in the α 1 domain

An interesting feature in Arctic charr MH class II α is that its high entropy positions are not the same in all fish, and that each Arctic charr population has a slightly different pattern of variable amino acid positions (Table 4-3). For example, in the de la Trinité population MH Class II α genes, there are conserved residues in positions that are variable elsewhere. This is also the case in the Siberian populations, but particularly for Lake Kiryalta -3 and -4 alleles. On the other hand, Lake Sitasjaure and Gander Lake have many variable positions that are much more conserved in other populations, and Lake Aigneau has some variable positions that are unique.

This kind of divergence may be linked to the adaptation to a new environment and the different pathogens encountered. A subset of pathogens may force the conservation of some residues, as others will be kept polymorphic, as a reservoir of multiple binding sites for eventual new emerging disease vectors.

Alleles

Similar to what it was found for mDNA haplotypes (Brunner *et al*. 2001), some MH Class II α alleles are maintained by distant populations within lineages. However, several MH Class II α alleles are also shared between the geographic lineages, which denote conservation of alleles necessary to confront common features of different pathogens. Most of the shared alleles occured between Atlantic and Arctic, and between Atlantic and Siberian lineages, a clustering tendency

that differs from that reported for mDNA (Brunner *et al*. 2001). In the latter, closer relations were obtained between the Acadian-Atlantic-Siberian groups, setting the Arctic group apart.

Supergroups

The emergence of molecular phylogenetic methods forces the separation of the Arctic-Siberian lineage classic classification into two distinct lineages (Brunner *et al*. 2001). Although MH class II α pairwise F_{ST} supports the division of all the geographic lineages as defined by existing mDNA studies, it postulates a closer linkage between the Arctic-Siberian supergroups, than the Atlantic-Acadian groupings. This difference puts the MH class II α data nearer to the traditional grouping of Arctic charr lineages, largely defined using meristic characters (Behnke 1984; Kornfield and Kircheis 1994), and it is also coincident with the allele distribution seen in Lakes Kamkanda and Aigneau, at both sides of the Arctic. As MH Class II α divergence is dependent on pathogen selection, it has been shaped in part by related environmental forces that have shaped the evolution of meristic characters. Nevertheless, MH divergence also depends on genetic inheritance and is capable of differentiating the Arctic and Siberian populations, just as neutral markers do. Therefore, MH class II α differentiation results lie between those yielded by neutral genetic divergence and phenotypic adaptation studies.

Overall, we concluded that the α subunit of the MH Class II has been strongly subjected to natural selection in the Arctic charr and shows a greater degree of polymorphism in MH class II α than most other species. Although most of the selection seen in this study is ancestral, there are

clear signs of local pressure that have individualized population structures. Therefore MH genes are an ideal population marker to assess polymorphisms and specializations, because they indicate the extent of change in ways that represent both divergence by genetic drift and by natural adaptation.

Acknowledgments

The authors would like to thank G. Furey, M. Shears and M. Bloom for assistance with field collection of samples and J. Witt for assistance with the population analysis. This research was made possible by Discovery grants from the Natural Sciences and Engineering Research Council of Canada to B.D. (217529-99) and M.P.(155928-92), as well a grant to M.P. from the Canadian International Polar Year Programme.

References

Alekseyev, S. S., Samusenok, V. P., Matveev, A. N., and Pichugin, M. Y.: Diversification, Sympatric Speciation, and Trophic Polymorphism of Arctic Charr, *Salvelinus alpinus* Complex, in Transbaikalia. Environmental Biology of Fishes 64: 97-114, 2002

Allendorf, F. and Thorgaard, G.: Tetraploidyand. the evolution of salmonid fishes. Turner BJ, ed, New York, 1984

Behnke, R. J.: Organizing the diversity of the Arctic charr complex. Biology of the Arctic Charr: Proceedings of the International Symposium on Arctic Charr, pp. 3-21, University of Manitoba Press, 1984

Bernatchez, L. and Landry, C.: MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? Journal of Evolutionary Biology 16: 363-77, 2003

Bernatchez, L., Rhydderch J, and F, Kircheis: Microsatellite gene diversity analysis in landlocked Arctic Charr from Maine. Transactions of the American Fisheries Society 131: 1106-1118, 2002

Bonneaud, C., Chastel, O., Federici, P., Westerdahl, H., and Sorci, G.: Complex Mhc-based mate choice in a wild passerine. Proceedings of the Royal Society of London. Series B. 273: 1111-6, 2006

Brown, J. H., Jardetzky, T. S., Gorga, J. C., Stern, L. J., Urban, R. G., Strominger, J. L., and Wiley, D. C.: Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1. Nature 364: 33-9, 1993

Brunner, P. C., Douglas, M. R., Osinov, A., Wilson, C. C., and Bernatchez, L.: Holarctic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. Evolution 55: 573-86, 2001

Bryja, J., Galan, M., Charbonnel, N., and Cosson, J. F.: Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (Arvicolinae). Immunogenetics 58: 191-202, 2006

Consuegra, S., Megens, H. J., Leon, K., Stet, R. J., and Jordan, W. C.: Patterns of variability at the major histocompatibility class II alpha locus in Atlantic salmon contrast with those at the class I locus. Immunogenetics 57: 16-24, 2005a

Consuegra, S., Megens, H. J., Schaschl, H., Leon, K., Stet, R. J., and Jordan, W. C.: Rapid evolution of the MH class I locus results in different allelic compositions in recently diverged populations of Atlantic salmon. Molecular Biology and Evolution 22: 1095-106, 2005b

Dawkins, R., Leelayuwat, C., Gaudieri, S., Tay, G., Hui, J., Cattley, S., Martinez, P., and Kulski, J.: Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. Immunological Reviews 167: 275-304, 1999

de Eyto, E., McGinnity, P., Consuegra, S., Coughlan, J., Tufto, J., Farrell, K., Megens, H. J., Jordan, W., Cross, T., and Stet, R. J.: Natural selection acts on Atlantic salmon major histocompatibility (MH) variability in the wild. Proceedings of the Royal Society of London. Series B. 274: 861-9, 2007

Edgar, R. C.: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-7, 2004

Elliott, J. and Baroudy, E.: The ecology of Arctic charr, *Salvelinus alpinus*, and brown trout, Salmo trutta, in Windermere (northwest England). Nordic Journal of Freshwater Research 71: 33-48, 1995

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

Frost, W.: Breeding Habits of Windermere Charr, *Salvelinus Willughbii* (Gunther), and their Bearing on Speciation of these Fish. Proceedings of the Royal Society of London. Series B, Biological Sciences 163: 232-284, 1965

Gladden, J., Maiers, B., Carmichael, T., and Reist, J.: Mitochondrial DNA Control Region Sequence Variation in Arctic Char (*Salvelinus alpinus* (L.)). Canadian Data Report of Fisheries and Aquatic Sciences 968, Central and Arctic Region Department of Fisheries and Oceans, 1995

Goldstein, D. B. and Pollock, D. D.: Launching microsatellites: a review of mutation processes and methods of phylogenetic interference. Journal of Heredity 88: 335-42, 1997

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebo, S., and Stet, R. J.: MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. Immunogenetics 55: 210-9, 2003

Hall, T. A.: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. 41: 95-98, 1999

Hardee, J. J., Godwin, U., Benedetto, R., and McConnell, T. J.: Major histocompatibility complex class II A gene polymorphism in the striped bass. Immunogenetics 41: 229-38, 1995

Hesthagen, T., Hindar, K., and Jonsson, B.: Effects of acidification on normal and dwarf Arctic charr *Salvelinus alpinus* (L.) in a Norwegian lake. Biological Conservation 74: 115-123, 1995

Johnson, L.: The Arctic charr, *Salvelinus alpinus*. In e. E. K. Balon (ed.): Charrs: salmonid fishes of the genus *Salvelinus*, pp. 19-98, Dr. W. Junk, The Hague, 1980

Jukes, T. and Cantor, C.: Evolution of protein molecules. In e. Munro HN (ed.): Mammalian Protein Metabolism, pp. 21-132, Academic Press, New York, 1969

Kido, Y., Saitoh, M., Murata, S., and Okada, N.: Evolution of the active sequences of the HpaI short interspersed elements. Journal of Molecular Evolution 41: 986-95, 1995

Kimura, M.: A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111-20, 1980

Klemetsen, A., Amundsen, P., Dempson, B., Jonsson, B., Jonsson, N., O'Connell, F., and Mortensen, E.: Atlantic salmon *Salmo salar* L., brown trout *Salmo trutta* L. and Arctic charr *Salvelinus alpinus* (L.): a review of aspects of their life histories. Ecology of Freshwater Fish 12: 1-59, 2003

Knudsen, R.: Relationships between Habitat, Prey Selection and Parasite Infection in Arctic Charr (*Salvelinus alpinus*). Nordic Journal of Freshwater Research 71: 333-344, 1995

Kornfield, I. and Kircheis, F.: Mitochondrial DNA and conservaton of an aboriginal arctic char (*Salvelinus alpinus aquassa*) from Flood Pond, Maine. Canadian Journal of Fisheries and Aquatic Science 51: 62-67, 1994

Kumar, S., Tamura, K., and Nei, M.: MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5: 150-163, 2004

Lamont, S. J.: Impact of genetics on disease resistance. Poultry Science 77: 1111-8, 1998

Lohm, J., Grahn, M., Langefors, A., Andersen, O., Storset, A., and von Schantz, T.: Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. Proceedings of the Royal Society of London. Series B. 269: 2029-33, 2002

Maiers, P., Gladden, B., Carmichael, T., and Reist, J.: Survey of restriction fragment lenght polymorphism variation in Arctic char *Salvelinus alpinus* (L) mitochondrial DNA. Data Report from the Central and Arctic Region, Winnipeg, Department of Fisheries and Oceans, 1995

Maynard, S. and Haigh, J.: The hitchhiking effect of a favorable gene. Genet. Res. 23: 23-35, 1974

Messaoudi, I., Guevara Patino, J. A., Dyall, R., LeMaoult, J., and Nikolich-Zugich, J.: Direct link between mhc polymorphism, T cell avidity, and diversity in immune defense. Science 298: 1797-800, 2002

Nei, M. and Gojobori, T.: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 3: 418-26, 1986

Nordeng, H.: Solution to the "charr problem" based on Arctic Charr (*Salvelinus alpinus*) in Norway. Canadian Journal of Fishery and Aquatic Science 40: 1372-1387, 1993

Ono, H., O'Huigin, C., Tichy, H., and Klein, J.: Major-histocompatibility-complex variation in two species of cichlid fishes from Lake Malawi. Molecular Biology and Evolution 10: 1060-72, 1993

Piertney, S. B. and Oliver, M. K.: The evolutionary ecology of the major histocompatibility complex. Heredity 96: 7-21, 2006

Reche, P. A. and Reinherz, E. L.: Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. Journal of Molecular Biology 331: 623-41, 2003

Sambrook, J., D, R., and J, S.: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, 2001

Satta, Y.: Effects of intra-locus recombination of HLA polymorphism. Hereditas 127: 105-12, 1997

Savvaitova, K.: Patterns of diversity and processes of speciation in Arctic Charr. Nordic Journal of Freshwater Research 71: 81-91, 1995

Schaschl, H., Suchentrunk, F., Hammer, S., and Goodman, S. J.: Recombination and the origin of sequence diversity in the DRB MHC class II locus in chamois (*Rupicapra spp.*). Immunogenetics 57: 108-15, 2005

Skulason, S., Snorrason, S., Noakes, D., Ferguson, M., and Malmquist, H.: Segregation in spawning and early life history among polymorphic Arctic charr, *Salvelinus alpinus*, in Thingvallavatn, Iceland. Journal of Fish Biology 37 (Suppl. A): 225-232, 1989

Slatkin, M. and Excoffier, L.: Testing for linkage disequilibrium in genotypic data using the Expectation-Maximization algorithm. Heredity 76: 377-83, 1996

Steinmetz, M. and Hood, L.: Genes of the major histocompatibility complex in mouse and man. Science 222: 727-33, 1983

Stet, R. J., de Vries, B., Mudde, K., Hermsen, T., van Heerwaarden, J., Shum, B. P., and Grimholt, U.: Unique haplotypes of co-segregating major histocompatibility class II A and class II B alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class II genotypes. Immunogenetics 54: 320-31, 2002

Takasaki, N., Murata, S., Saitoh, M., Kobayashi, T., Park, L., and Okada, N.: Species-specific amplification of tRNA-derived short interspersed repetitive elements (SINEs) by retroposition: a process of parasitization of entire genomes during the evolution of salmonids. Proceedings of the National Academy of Science 91: 10153-7, 1994

Volpe, J. P. and Ferguson, M. M.: Molecular genetic examination of the polymorphic Arctic charr *Salvelinus alpinus* of Thingvallavatn, Iceland. Molecular Ecology 5: 763-72, 1996

Watterson, G. A.: The Homozygosity Test of Neutrality. Genetics 88: 405-417, 1978

Wilson, A. J., Gislason, D., Skulason, S., Snorrason, S. S., Adams, C. E., Alexander, G., Danzmann, R. G., and Ferguson, M. M.: Population genetic structure of Arctic charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales. Molecular Ecology 13: 1129-42, 2004

Chapter 5. Global MH Class II β polymorphism in Arctic charr (*Salvelinus alpinus* L.) and adaptation to local environments

Nucleotide sequence data reported are available in the GenBank databases under the accession numbers: EU159587-EU159662

Introduction

The current territory occupied by Arctic charr (*Salvelinus alpinus* L.) was colonized only after the last glaciation which occurred 10000-15000 years ago (Bernatchez *et al*. 2002);(Johnson 1980). In the colonization process, several different morphotypes have been generated that can confound population characterization based on meristic markers (Alekseyev *et al*. 2002; Allendorf and Thorgaard 1984). Sometimes the different morphotypes are reproductively isolated and show genetic variation (Elliott and Baroudy 1995; Frost 1965; Hesthagen *et al*. 1995; Nordeng 1993; Skulason *et al*. 1989), but most of the time differentiation is difficult to observe due to the short reproductive isolation time of the populations.

Nevertheless, thorough analysis of meristic characters has facilitated the grouping of the many different forms into 4 lineages (Behnke 1984) including: a group covering European and Greenland Arctic charr, a group for northern Canada and Siberia, a group for Alaska and western Russia, and a group for eastern North America. Subsequent genetic characterization using mitochondrial DNA (mDNA) divided the Arctic charr into five relatively homogeneous geographic lineages: an Atlantic lineage including southeastern Canadian and European Arctic charr; an Arctic lineage for northern North America including western Greenland; a Beringian lineage for north Pacific Dolly Varden (*Salvelinus malma*); a Siberian lineage for northern Russia, and an Acadian group for landlocked populations found largely in southern Québec and Maine (Brunner *et al*. 2001) (Figure 5-1).

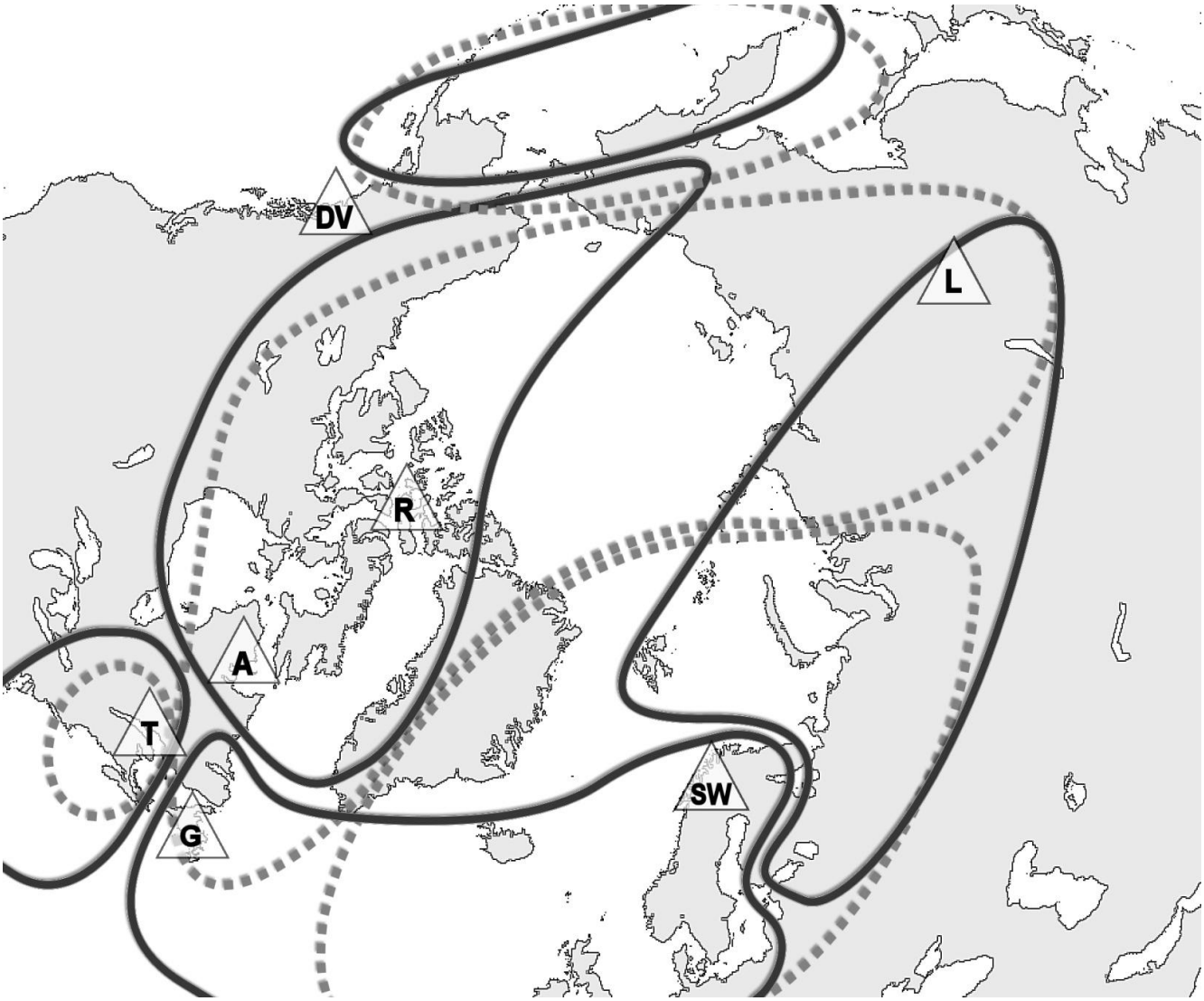


Figure 5-1. Arctic charr phylogeographic lineages. The map shows in dashed lines the limits of the geographic lineages classified using meristic characters (Behnke 1984) and in solid lines the geographic limits of the lineages according to mDNA sequencing (Brunner *et al*. 2001). The populations sampled for this study are indicated in triangles. The names represent location of each population: A for Lake Aigneau, G for Gander Lake, R for Resolute Lake, SW for Lake Sitasjaure, T for de la Trinité River, L for Lake Kiryalta and DV for Dolly Varden samples from Juneau.

Attempts to establish genetic relationships between groups of Arctic charr, however, continue to show a bewildering variety of results according to method, sample size and sample selection (Goldstein and Pollock 1997; Thorpe 1995). To help resolve some of the differences in lineal classifications, we characterized the Arctic charr complex by observing the polymorphism of the β subunit of the Major Histocompatibility Class II receptor. The molecule exposes pathogen derived peptides to T-lymphocytes, which then determine whether those particles are self or non-self and initiate an immune response if required (Steinmetz and Hood 1983). Because of the enormous diversity of pathogens, MH receptors are very polymorphic at the population level, so as to increase the possibility that at least some members will survive disease outbreaks (Grimholt *et al*. 2003; Reche and Reinherz 2003). In that sense, individuals that confront similar pathogens tend to carry common patterns in their MH molecules (Bonneaud *et al*. 2006; Lamont 1998). However, because of the evolutionary processes that affect the MH polymorphism (Bernatchez and Landry 2003), the pathogen-specific patterns are loose and a wide divergence in allele structure is kept within the population (Dawkins *et al*. 1999; Messaoudi *et al*. 2002). Therefore, the MH alleles found in distinct populations are defined by the genetic line of the individuals as well as by the selective pressure of the pathogens present in the environments they inhabit (Bryja *et al*. 2006; de Eyto *et al*. 2007).

There are two types of MH receptors present in most jawed vertebrates, MH Class I and MH Class II. The latter are composed of two subunits: Class II α and Class II β. Each one has 5 domains, normally encoded by 5 separated exons. The second exon encodes the α 1 and β 1 domains, respectively, that bind the antigen to be presented to T-lymphocytes (Peptide Binding Region) and are the highly variable area of the molecule (Steinmetz and Hood 1983).

The β subunit of the Class II MH molecule in humans contains most of the dimer's polymorphism; its variable residues are usually part of the peptide binding region (PBR) (Reche and Reinherz 2003) and it has been extensively used to characterize wild populations (Dorschner *et al*. 2000; Landry and Bernatchez 2001; Miller *et al*. 2001). This study examined whether or not the β subunit was capable of distinguishing among recently isolated Arctic charr populations sampled from across the natural distributional range of the species. The working hypothesis of this study was that the high variation in the β subunit of the Class II MH molecule would be complemented by the influence of differential pathogen selection in the different locations, thereby facilitating distinguishing among studied populations.

Materials and Methods

Fish samples

Muscle samples of fresh caught Arctic charr were immersed in 70% ethanol or RNA preserving solution (25 mM $C_6H_5NaO_7$, 10 mM EDTA and 4 M $((NH_4)_2SO_4)$ and shipped to our laboratory from the sampling areas: Lake Kiryalta (57°08'N; 119°27'E) in the Transbaikal region of eastern Russia, part of the Siberian lineage described by Brunner (Brunner *et al*. 2001). Lake Kiryalta is geographically divided into two water bodies (Lake Kiryalta -3 and -4), but connected by a 50 m canal, so considered as only one population for this study; Lake Sitasjaure (68°00'N; 17°25'E) in Sweden and Gander Lake, Newfoundland (48°55'N; 54°35'W) both part of the Atlantic lineage; Lake Aigneau (57°14'N; 70°07'W) in Québec and Resolute Lake (74°41'N; 94°57'W) in Nunavut, Canada, part of the Arctic lineage; the de la Trinité River (49°25'N; 67°18'W) in Québec, part of the Acadian lineage; and Dolly Varden (*Salvelinus malma*) samples from Juneau (58°22'N 134°35'W), part of the Beringian lineage (Figure 5-1).

DNA extraction

Samples in ethanol and RNA preserving solution were placed in TE buffer (100 mM Tris, 1mM EDTA) for 2 hrs, and then incubated for 3 hrs in 100 mM Tris, 10 mM EDTA, 240 mM NaCl, 1% SDS and 300 ug/ml Proteinase K at 55 ºC. This was followed by two extractions, each with one volume of phenol and one extraction with one volume of chloroform. The DNA was precipitated with 1/10 volumes of 3 M CH2COONa pH 5.2 and 2 volumes of ethanol, and then

washed with 70% ethanol. The final pellet was then resuspended in water and stored at 4 ºC until used for PCR.

PCR and cloning

PCR forward (SAALDABF: 5' – GAT ACT CCT CAA AGG ACC TG– 3') and PCR reverse (SAALDABR: 5'- CTC AGC CAG GTC ACT CTG – 3') primers were designed to match the beginning of the second exon, and the beginning of the third exon, respectively. The primers were based on the full length sequences of Arctic charr (Chapter 2).

The conditions for the PCR were at 95 ºC of initial denaturation for 2 min, plus 35 cycles of 95 ºC of denaturation for 30 sec, 57 ºC of primer annealing for 40 sec and 72 ºC of extension for 1.5 min. A final extension at 72 ºC for 8 min was done to ensure the addition of an A overhang at the end of the complementary strand.

The PCR amplicons were separated by agarose electrophoresis in a 1% gel to verify size and then extracted directly from the gel using GenElute Agarose Spin Columns (Sigma-Aldrich, St Louis, MO). The fragments were inserted in pGEM-T vectors as per manufacturer instructions (Promega Corporation, Madison, WI). XL-blue strains of *E. coli* bacteria were made competent by washing them twice and then re-suspending them with Thawing buffer (10mM PIPES, 15mM CaCl-2H_2O, 250 mM KCl, 55mM $MnCl_2$-4H_2O, pH 6.7). The bacteria were transformed by the Inoue procedure (Sambrook *et al.* 2001) and grown in LB plates with ampicillin (100 ug/ml) plus 1 uM IPTG and 100 ug/ml X-gal. Eight to twelve white colonies were selected and grown in LB media. The bacteria were used for PCR with SAALDABF primer and a Reverse primer for Denaturant Gradient Gel Electrophoresis (DGGE), designed to match the end of the second exon

(SAALDABDGGER: CGG GCG GGG GCG GCG GGC CGG GCG CGG GGC GCG GCG GGC GCC AGT ACG GCG CTG TAG). The amplicons were separated in a denaturing gradient of 20-45% for 15.5 hrs at 60 °C (Bio-Rad laboratories, Richmond, CA). Only the alleles that differed between samples were selected for DNA sequencing in order to avoid redundancies. The plasmids selected for sequencing were extracted from the grown bacteria with GenElute Plasmid Miniprep Kit (Sigma-Aldrich, St Louis, MA) according to manufacturer instructions, and sequenced using the T7 and Sp6 primers with a big dye terminator v3.1 (Applied Biosystems, Foster City, CA) in a 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA).

Since the PCR was yielding fragments over 1200 bp, which presented some difficulties for amplification and cloning, a degenerate reverse primer was designed in the middle of the second intron (SAALDABR2: TAG TCA TAC TGT AAT GKC ACT AC). SAALDABF1 and SAALDABR2 were used under the same experimental conditions mentioned above to obtain all the sequences used on this study.

Sequencing and data analysis

The sequences obtained were aligned using the software Muscle (Edgar 2004). An alignment of the previously sequenced full length Class II β cDNA sequence (Chapter 2), was used as a guide for the definition of the end of the second exon and the beginning of the intron. The intron and exon sequences were analyzed separately. The part of the intron used for the phylogenetic analysis was the 297-312 bp region from the beginning of the intron, to 10 bp upstream of the position of the primer SAALDABR2. BioEdit (Hall 1999) was used to define the open reading frames and to translate them into protein sequences. This software was also used to obtain

entropy values (H) that measure the variability of each amino acid residue position (Shannon, 1948 *fide in* Reche and Reinherz 2003).

Following similar criteria used by other authors (Consuegra *et al*. 2005a; Consuegra *et al*. 2005b; Reche and Reinherz 2003), we considered highly entropic sites, having values ≥0.375, to be approximately 25% of the maximum observed entropy in the molecule. MEGA 3 (Kumar *et al*. 2004) was used to find and correct sequences containing singleton errors and to calculate the diversity of the sequences among and within populations, the rates of non-synonymous/synonymous substitutions, and to draw the phylogenetic trees using the Kimura-2 parameter model (Kimura 1980) for the intron sequences and the Nei-Gojobori model (Nei and Gojobori 1986) for non-synonymous substitutions in the exon sequences. Arlequin software 3.01 (Excoffier *et al*. 2005) was used to calculate the allelic frequencies, pairwise F-statistics (F_{ST}) and to perform the exact test for population differentiation.

Results

Allelic sequences

The mean p-distance for the exons of the 76 sequences obtained out of 65 fish samples was 0.07, with 0.053 of intrapopulation diversity. 41 different exon alleles were found of 243, 246 or 249 bp, out of 65 fish samples. Five of the identified alleles were present in frequencies over 0.05, leaving 21 unique alleles.

The mean p-distances in all populations were similar to the total, with the exception of Gander Lake (0.04) and de la Trinité River (0.02) populations. The de la Trinité River population also had only 3 different alleles, half of the average per population. Details of the diversity of each population are shown in Table 5-1 and the translated allele sequences for the β 1 domain are shown in Figure 5-2.

From the 80 amino acid positions in the β 1 domain, 27 presented high entropy (H>0.375). The highly entropic sites had a mean p-distance of 0.47, 14 of which matched with comparable sites in one of the human HLA Class II β genes, and 8 of which corresponded with peptide binding residues on the human Class II β molecule. However, the Arctic charr and the human Class II β proteins were under 35% similar, making it difficult to make exact structural matches.

Most populations had a similar number and position of highly entropic sites. However, Gander Lake and Lake Sitasjaure populations had less variable sites than the average. An extreme case was the de la Trinité River population that had only one highly entropic site. The de la Trinité

River population had other low variable sites, in a number and position that corresponded to the highly variable sites of the other populations. Details on the variation of the highly entropic sites of each population are shown in Table 5-2.

The introns presented significantly lower variation, diverging in average a 0.038 with 0.03 of intrapopulation diversity. Introns presented 29 haplotypes, only 3 of which had frequencies over 0.05, with a single haplotype holding 0.17. Overall, 13 unique alleles were found. The second intron of the Class II β gene was approximately 780 bp long, 320-352 bp of which were filled with a 32 bp conserved fragment repeated 10-11 times (ACA GTA TGA CTA GTC AGC TAT GTA GTG ACA TT), ending 176 bp downstream from the beginning of exon 3. The stretch is also present in the same intron, in the Class II β of other salmonids and conserved nearly a 100% in most cases, typically only as a single copy.

Balancing selection

A relatively high ratio of non-synonymous/synonymous substitutions of 3.3, indicated balancing selection acting on the polymorphic residues. The high ratio was present in all populations studied and maximized in the Lake Sitasjaure population at 4.9. In Lake Kiryalta the rate of non-synonymous/synonymous substitutions was at its lowest (Table 5-1), where as in Lake Aigneau the ratio approximated the values for other southern Canadian populations (Gander Lake and de la Trinité River).

Synonymous substitutions averaged 0.026 in Arctic charr, a slightly lower value than for intron variation, which may indicate some intron hitchhiking (Maynard and Haigh 1974) in which

linkage with the exons has partially submitted them to the evolutionary conditions of the coding region and preserved greater variability than would occur under strict neutrality.

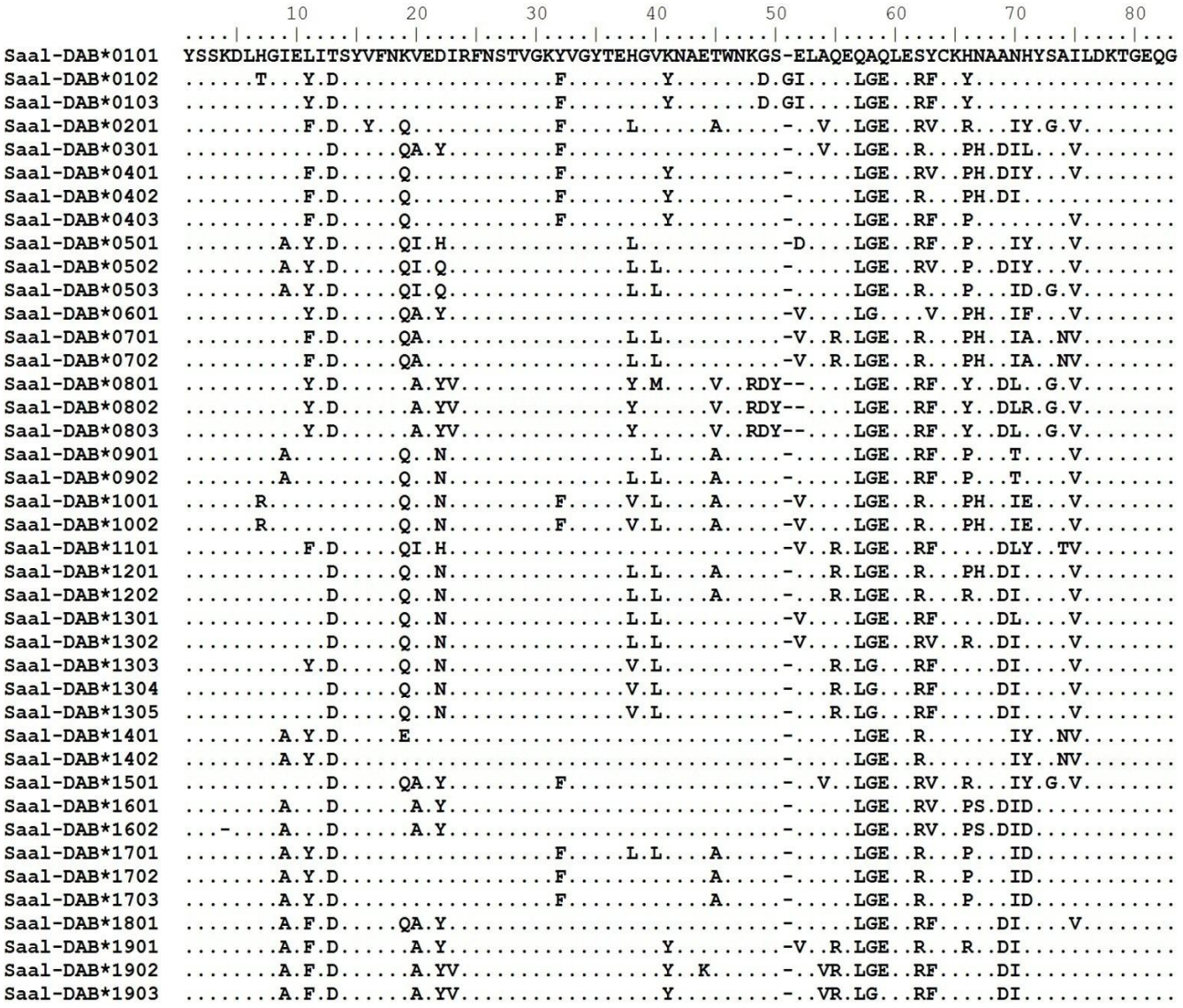


Figure 5-2. MH Class II β alleles.

The translated sequences of the alleles found in this study for the second exon of the MH Class II β gene in Arctic charr.

Populations	Fish Samples	Exon Divergence	Non-synonymous Subst	Synonymous Subst	dn/ds Ratio
Aigneau (Arc)	11	0.06	0.07	0.02	3.27
Resolute (Arc)	8	0.07	0.09	0.02	4.69
Dolly Varden (Ber)	5	0.06	0.08	0.02	4.38
Sitasjaure (Atl)	9	0.05	0.06	0.01	4.94
Gander (Atl)	13	0.04	0.05	0.02	2.72
Kiryalta (Sib)	11	0.06	0.07	0.03	2.22
De la Trinité (Aca)	8	0.02	0.03	0.01	3.21
Lineages					
Arctic	19	0.07	0.08	0.02	4.10
Beringia	5	0.06	0.08	0.02	4.39
Atlantic	22	0.04	0.05	0.02	2.74
Siberia	11	0.06	0.07	0.03	2.22
Acadian	8	0.02	0.03	0.01	3.13

Table 5-1. MH Class II β second exon divergence in the populations.

The table shows the sample number for each population studied and the lineages that they are part of (as defined by Brunner *et al.*, 2001). The distance values for the parameters used in the analysis are displayed. The lineage to which each population belongs to appears in parentheses: Arc: Arctic, Ber: Beringia, Atl: Atlantic, Sib: Siberia, Aca: Acadia

Pos	Aig	Res	Gan	Sit	Kir	Trin	D.Var	Pos	Aig	Res	Gan	Sit	Kir	Trin	D.Var
1	0	0	0	0	0	0	0	41	0.68	0.45	0	0.66	0	0	0.64
2	0	0	0	0	0	0	0	42	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	43	0	0	0	0	0	0	0
4	0	0	0	0	0	0.35	0	44	0.56	0.29	0	0	0	0	0
5	0	0	0	0	0	0	0	45	0	0.64	0.52	0.66	1.04	0.35	0.45
6	0	0	0	0	0	0	0	46	0	0	0	0	0	0	0
7	0	0.89	0	0	0	0	0	47	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	48	0	0	0	0	0.56	0.35	0
9	0.64	0.68	0.41	0.66	0	0.35	0.69	49	0	0.29	0	0.30	0.56	0.35	0
10	0	0	0	0	0	0	0	50	0	0	0	0	0.56	0.35	0
11	0.96	1.08	0.89	1.09	0.82	0.35	1.10	51	0	0.29	0	0.30	0	0	0
12	0	0	0	0	0	0	0	52	0.45	1.13	0.26	0.30	0.56	0.35	0.64
13	0	0.64	0.26	0.66	0	0	0	53	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	54	0.56	0.29	0	0	0.68	0	0.45
15	0	0	0	0	0	0	0	55	0.56	0.56	0	0	0.45	0	0.64
16	0	0	0	0	0.29	0	0	56	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	57	0	0	0.26	0	0	0	0
18	0	0	0	0	0	0	0	58	0	0	0.26	0	0	0	0
19	0.56	0.45	0.76	0.66	0.56	0	0.69	59	0.64	0	0.26	0	0	0	0.64
20	0.92	0.92	0.52	0	0.56	0	0.69	60	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0	61	0	0	0	0	0	0	0
22	1.29	1.47	0.76	0.66	0.72	0	1.01	62	0	0	0.26	0	0	0	0.45
23	0.56	0.29	0	0	0.56	0.35	0.45	63	0.82	0.68	0.80	1.07	0.96	0.35	1.01
24	0	0	0	0	0	0	0	64	0	0	0	0	0	0	0
25	0	0	0	0	0	0	0	65	0	0	0	0	0	0	0
26	0	0	0	0	0	0	0	66	0.68	0.82	0.83	0.30	1.04	0.35	1.01
27	0	0	0	0	0	0	0	67	0.56	0.64	0.26	0.59	0.69	0.35	0.45
28	0	0	0	0	0	0	0	68	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	69	0.56	0.56	0.60	0.59	0.45	0	0.64
30	0	0	0	0	0	0	0	70	0.29	0.72	0.26	0.92	0.56	0.35	0.64
31	0	0	0	0	0	0	0	71	0.72	1.36	1.08	1.07	1.03	0.68	0.87
32	0.45	0.68	0.52	0.66	0.68	0	0.45	72	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	73	0	0.56	0	0	0.68	0.35	0
34	0	0	0	0	0	0	0	74	0.45	0.45	0.68	0	0	0	0
35	0	0	0	0	0	0	0	75	0.64	0.45	0.65	0.66	0	0.35	0.69
36	0	0	0	0	0	0	0	76	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	77	0	0	0	0	0	0	0
38	1.08	1.10	0.41	0.59	1.04	0.35	0.64	78	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	79	0	0	0	0	0	0	0
40	0.68	0.68	0.41	0.66	0.72	0	0.64	80	0	0	0	0	0	0	0

Table 5-2. MH Class II β, β 1 domain amino acid entropy.

The entropy values for each amino acid positions for all the populations. The values with entropy ≥0.375 are shaded, which is equivalent to approximately 25% of the maximum observed entropy. In the first column the amino acid position is shown, and in the next the populations from Lake Aigneau (Aig), Resolute Lake (Res), Gander Lake (Gan), Lake Sitasjaure (Sit), Lake Kiryalta (Kir), de la Trinité River (Tri) and Juneau Dolly Varden (D Var)

Population differentiation

The pairwise F_{ST} of the MH Class II β second exon are shown in Table 5-3. At a significance level of $p<0.05$, all the phylogeographic lineages appeared isolated, with the exception of the Beringian lineage, represented by *Salvelinus malma* samples of Dolly Varden fish from the Juneau location. Most Dolly Varden sequences do not cluster together in the phylogenetic tree and are spread over the whole Arctic charr MH Class II β phylogeny. However, Dolly Varden alleles remained fairly unique and the Juneau population appeared isolated when considering the results of the exact test for population differentiation. Within the lineages, the populations studied appeared isolated from each other when considering the p-values of the pairwise F_{ST} and exact tests. Aside from the Juneau Dolly Varden samples, only the populations from Lake Sitasjaure and Resolute Lake did not separate from each other with the F_{ST} test. Nevertheless, the pairwise F_{ST} p-value for the Sitasjaure-Resolute comparison fell within the error margin area for population differentiation (Table 5-3a.1).

The intron pairwise FSTs did not demonstrate isolation between the Beringian and the Atlantic and Arctic lineages, or between the Arctic and Siberian lineages. However, the introns exact test results indicated isolation of all lineages (Table 5-3b.2).

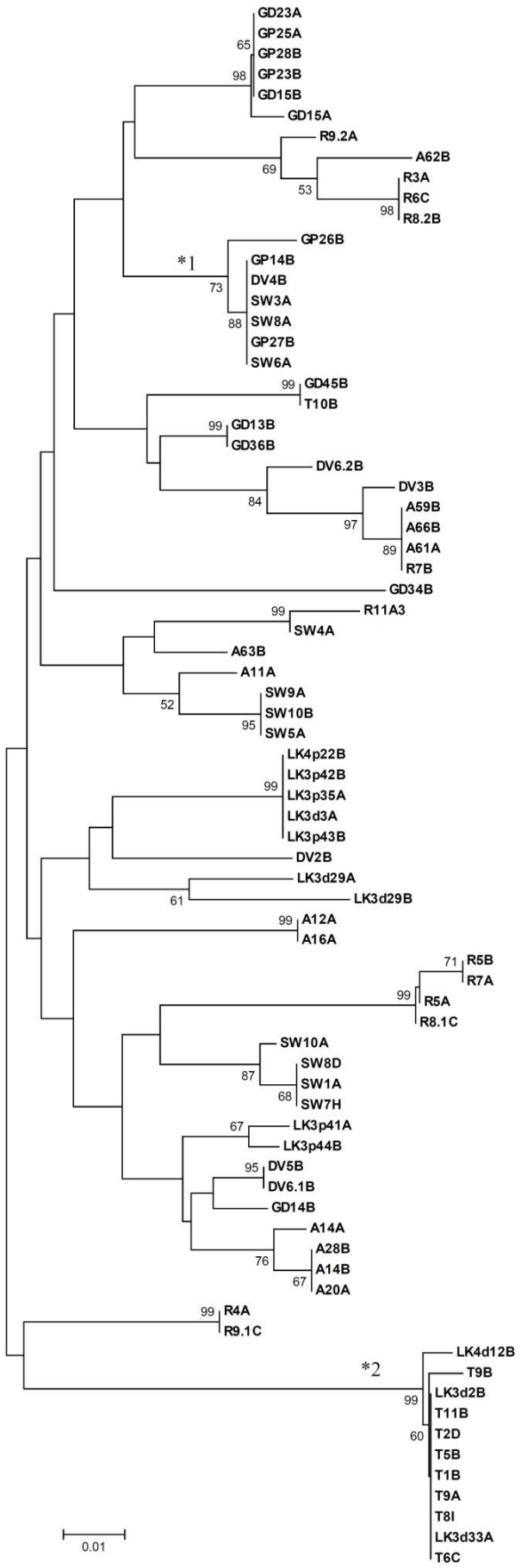
Phylogenetic inference

In the exon tree (Figure 5-3, left), only 11 samples were not grouped with a confident bootstrap (>75%). Most of the samples were part of consistent clades, although there was not a major subdivision in the phylogenetic tree, which indicated exons belonging to the same loci.

Eleven clusters grouped individuals that belonged to the same population and may correspond to specific alleles for survival in specific habitats. Another 5 subclusters grouped individuals from different populations, some of them separated by considerable geographic distances, ie. Figure 5-3 (*2): the grouping of Lake Kiryalta and de la Trinité River. The Lake Kiryalta-de la Trinité River grouping possibly represents an allele family relevant for survival in both areas and contains all except one of de la Trinité River alleles. The de la Trinité River population was the only Arctic charr population that did not have unique characteristic alleles.

In the intron phylogenetic tree (Figure 5-3, right) on the other hand, 20 samples did not group in confidently bootstrapped branches. The remainder of the samples formed 13 groups, of which 10 were related geographically and 3 were not. The Atlantic lineage that in the exons forms a major cluster containing Lake Sitasjaure and Gander Lake individuals (*1), was also present in the intron tree. An unexpected clustering between Arctic and Siberian alleles (*3) was observed only in the intron phylogenetic tree.

Figure 5-3. MH Class II β phylogenetic reconstruction. Left: The phylogenetic reconstruction using the sequenced second exons with the Nei-Gojobori model, which considers only non-synonymous mutations. Right: The phylogenetic reconstruction using the sequenced second introns with the Kimura-2 parameter model. The trees were bootstrapped 1000 times and only bootstraps over 50% are shown. The names represent series of each population starting with an A for Lake Aigneau, G for Gander Lake, R for Resolute Lake, SW for Lake Sitasjaure, T for de la Trinité River, LK for Lake Kiryalta and DV for Dolly Varden samples from Juneau.


GD23A
GP25A
GP28B
GP23B
GD15B
GD15A
R9.2A
A62B
R3A
R6C
R8.2B
GP26B
*1
GP14B
DV4B
SW3A
SW8A
GP27B
SW6A
GD45B
T10B
GD13B
GD36B
DV6.2B
DV3B
A59B
A66B
A61A
R7B
GD34B
R11A3
SW4A
A63B
A11A
SW9A
SW10B
SW5A
LK4p22B
LK3p42B
LK3p35A
LK3d3A
LK3p43B
DV2B
LK3d29A
LK3d29B
A12A
A16A
R5B
R7A
R5A
R8.1C
SW10A
SW8D
SW1A
SW7H
LK3p41A
LK3p44B
DV5B
DV6.1B
GD14B
A14A
A28B
A14B
A20A
R4A
R9.1C
*2
LK4d12B
T9B
LK3d2B
T11B
T2D
T5B
T1B
T9A
T8I
LK3d33A
T6C
0.01

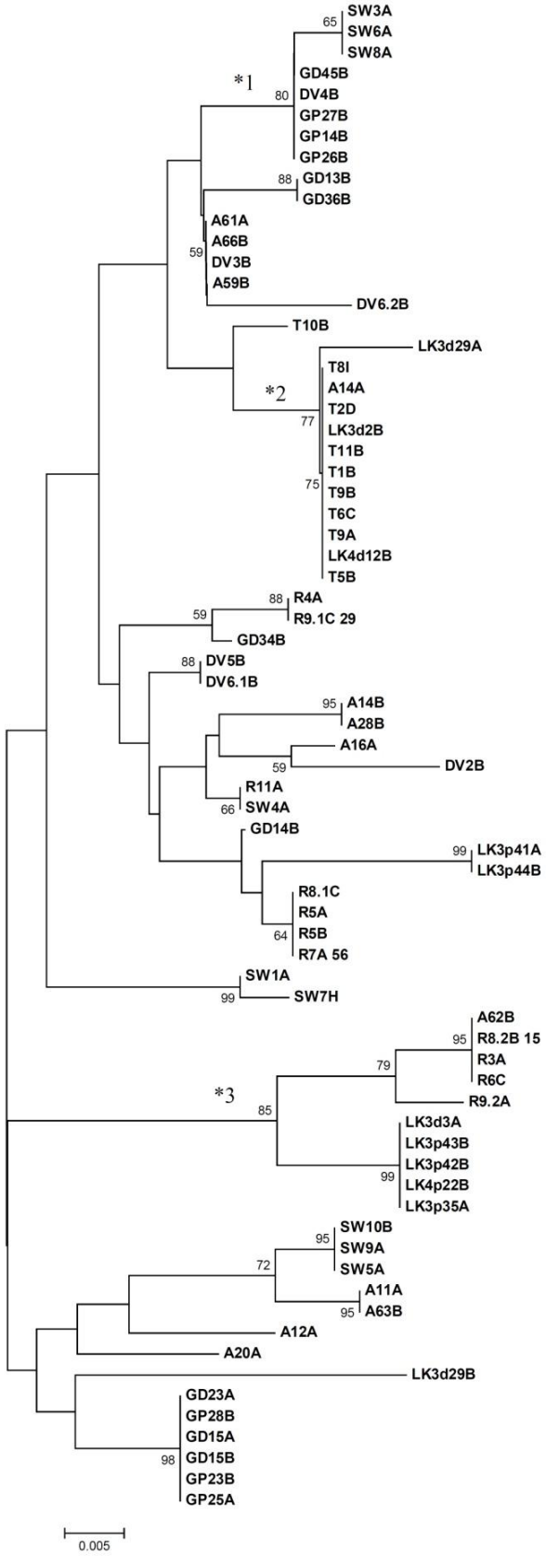

SW3A
SW6A
SW8A
*1
GD45B
DV4B
GP27B
GP14B
GP26B
GD13B
GD36B
A61A
A66B
DV3B
A59B
DV6.2B
T10B
LK3d29A
T8I
A14A
*2
T2D
LK3d2B
T11B
T1B
T9B
T6C
T9A
LK4d12B
T5B
R4A
R9.1C 29
GD34B
DV5B
DV6.1B
A14B
A28B
A16A
DV2B
R11A
SW4A
GD14B
LK3p41A
LK3p44B
R8.1C
R5A
R5B
R7A 56
SW1A
SW7H
A62B
R8.2B 15
R3A
R6C
R9.2A
*3
LK3d3A
LK3p43B
LK3p42B
LK4p22B
LK3p35A
SW10B
SW9A
SW5A
A11A
A63B
A12A
A20A
LK3d29B
GD23A
GP28B
GD15A
GD15B
GP23B
GP25A
0.005

Table 5-3. Population differentiation according to the second exon of the MH Class II β gene.

Below the diagonal are the differentiation values for the pairwise F_{ST} analysis (5000 permutations). Above the diagonal are the p-values of the exact test of differentiation using haplotype frequencies (6000 Markov steps). For the F_{ST} and exact test, p-values <0.05 were considered significant for this study, and are indicative of the isolation of the populations from each other. The relations where differentiation was found are not-shaded. Table 5-3a shows the F_{ST} values and exact test p-values for the exons, Table 5-3b shows both tests for the introns. Tables numbers ending in .1 shows the values for each population. Table ending in .2 shows the differentiation of the phylogeographic lineages (Brunner *et al*. 2001). The average error was 0.002 for both tests in exons and introns.

5-3 a.1

FST\Exact	Aigneau	Dolly Varden	Gander	Kiryalta	Resolute	Sitasjaure	De la Trinité
Aigneau		0.025	<0.001	<0.001	0.005	<0.001	<0.001
Dolly Varden	0.004		0.047	0.012	0.020	0.045	0.005
Gander	0.215	0.113		<0.001	<0.001	0.007	0.001
Kiryalta	0.173	0.095	0.284		<0.001	0.001	0.012
Resolute	0.102	0.047	0.186	0.182		<0.001	0.001
Sitasjaure	0.159	0.078	0.135	0.207	0.084		0.001
De la Trinité	0.574	0.550	0.638	0.414	0.537	0.607	

5-3 a.2

	Arctic	Bering	Atlantic	Siberia	Acadian
Arctic		0.002	<0.001	<0.001	<0.001
Bering	-0.005		0.038	0.027	0.004
Atlantic	0.107	0.067		<0.001	<0.001
Siberia	0.149	0.095	0.232		0.01
Acadian	0.494	0.550	0.573	0.414	

5-3 b.1

FST\Exact	Aigneau	Dolly Varden	Gander	Kiryalta	Resolute	Sitasjaure	De La Trinité
Aigneau		0.198	<0.001	0.001	0.007	0.004	<0.001
Dolly Varden	0.044		0.007	0.005	0.010	0.037	<0.001
Gander	0.102	0.092		<0.001	<0.001	<0.001	<0.001
Kiryalta	0.100	0.181	0.232		<0.001	0.001	0.003
Resolute	0.166	0.236	0.217	0.049		0.001	<0.001
Sitasj	0.033	0.071	0.133	0.159	0.224		<0.001
De la Trinité	0.312	0.382	0.364	0.321	0.481	0.363	

5-3 b.2

	Arctic	Bering	Atlantic	Siberia	Acadia
Arctic		0.026	<0.001	<0.001	0.002
Bering	0.086		0.008	0.008	0.001
Atlantic	0.072	0.035		<0.001	<0.001
Siberia	0.034	0.181	0.168		0.002
Acadia	0.303	0.382	0.276	0.321	

	Beringia	Atlantic	Siberia	Acadian
Arctic	0.065	0.064	0.076	0.097
Beringia		0.057	0.068	0.083
Atlantic			0.074	0.087
Siberia				0.071

Table 5-4. MH Class II β second exon sequence p-distances between the lineages previously identified by Brunner *et al*. (2001).

Discussion

The peptide binding region of the β subunit of the Class II MH molecule is the most polymorphic coding region in vertebrates (Grimholt *et al*. 2003) and most of its polymorphic sites encode for residues that bind the antigen to be presented to T-lymphocytes (Reche and Reinherz 2003). Reports on populations containing different Class II β alleles living in close geographic areas and having a common high rate of non-synonymous/synonymous substitutions (Landry and Bernatchez 2001; Miller *et al*. 2001) suggest that the polymorphism is tightly associated with the survival of populations in specific habitats.

Origins of the polymorphism of the β 1 domain

The molecular origins of the high polymorphism of MH genes remain a topic of debate. Some theories suggest that conserved repeated elements may work as hotspots for recombination (Schaschl *et al*. 2005), facilitating the patchwork structure that is seen in many Class II β gene second exons (Dawkins *et al*. 1999). Depending on the nature of the repeats, repetition may promote polymerase slippage, increasing the probabilities of error during the replication of the DNA strand (Weitzmann *et al*. 1997). Almost every second intron in all MH Class II β genes studied to date have simple GT or GA repeated sequences preserved through speciation (Schwaiger *et al*. 1994). Only in cichlids has a more complex repeated motif of 12 bp in the MH class II β been reported (Ono *et al*. 1993). In the Arctic charr MH Class II β gene reported here there was a contiguous repetition of a 32 bp motif in the second intron. A similar motif is present in a single copy within the corresponding intron of several Oncorhynchus species MH Class II β

genes, but only in Arctic charr and in two Atlantic salmon (*Salmo salar*) samples reported so far (AJ439067; AJ39069) has the sequence been replicated. The phenomenon may represent the initiation in the salmonidae family of a process that might lead to more radical allelic changes within the MH class II β genes. The presence of the repeat at the end of the intron has not yet had a clear effect on the Arctic charr β 1 domains, although the β 1 domains are on average slightly more diverse in Arctic charr than in other salmonids. The replication also indicates a closer Class II β MH relation between the genera *Salvelinus* and *Salmo* than either has with the genus *Oncorhynchus*.

MH Class II β allele variation in Arctic charr populations

The relatively high interpopulation diversity seen in the data presented here was probably acquired recently, during the Arctic charr geographic range expansion. The high interpopulation diversity allowed differentiation between all studied populations, with the exception of Dolly Varden for which the pairwise F_{ST} and the exact tests of differentiation indicated similarities to other Arctic charr populations. Although population differentiation refers to the isolation of the lineages, the phylogenetic tree was mostly shaped by allelic differentiation among populations and did not group individuals according to their lineage. The only clear exceptions were 6 samples from populations on both sides of the Atlantic lineage range that had related exon alleles (*1).

The isolation of the studied populations within the suggested Arctic charr lineages contrasts with the results obtained using mDNA where the lineages identified did not contain subdivisions (Brunner *et al*. 2001). Local selective pressure on specific MH Class II β alleles has probably

promoted recent adaptive differentiation that has been too rapid to affect the evolution of the neutral marker. Nevertheless, studies on microsatellites within restricted areas have indicated the presence of isolated populations of Arctic charr (Bernatchez *et al*. 1998; Wilson *et al*. 2004), so subtle neutral changes may also have occurred in the recent post-glacial colonization period.

The intron phylogenetic tree contains a cluster similar to one in the exon tree that groups Gander Lake and Lake Sitasjaure individuals into the Atlantic lineage (*1), a relationship that was also supported by mDNA-RFLP (Wilson *et al*. 1996), and that argues for the common origin of these populations. However, unexpected clusters situate fish from Lake Kiryalta and de la Trinité River on the same branch (*2), and individuals from Lake Kiryalta together with Arctic lineage individuals (*3). The branch that includes Lake Kiryalta and de la Trinité River is also similar to a cluster in the exon phylogenetic tree (*2), and probably relates to fitting alleles that were kept on those populations, but eliminated from the Atlantic lineage. Interestingly the alleles on this branch were among the few alleles maintained in the de la Trinité River population and may be specific to the environmental conditions of local habitat. The relationship between the introns from Lake Kiryalta and the Arctic lineage (*3) provides an example of the evolutionary independence and low level of linkage that the second exon of MH Class II β has with the rest of the gene. Although the exons are completely different, the intron has been conserved in both populations which indicate common origin for the allele families SAAL-DAB*05 and SAAL-DAB*03. Nevertheless, since the studies on mDNA (Brunner *et al*. 2001) proposed an early departure of the Arctic lineage from the rest of the Arctic charr complex, it has to be assumed that the alleles were selected out of the other lineages, and that similar environments and exposure to related pathogens might have force retention of the alleles only in Arctic and Siberian populations.

Supergroups

Since the MH class II β polymorphism is strongly affected by the environment, some discordance is observed between the MH Class II β marker and the mDNA data (Brunner *et al*. 2001), particularly with respect to the relationship among the 5 geographic lineages that Brunner *et al*. defined (Table 5-4). For example, the Siberian and Atlantic lineages were found to be very close to each other using mDNA, but for MH class II β the Atlantic lineage was closer to the Beringian lineage, suggesting preservation of similar alleles in both groups. mDNA suggested that the Siberian lineage was evolutionarily situated between the Atlantic and the Beringian lineages, so the alleles that join both lineages must have been depleted from the intermediate Siberian lineage. In another case, according to the neutral mDNA marker (Brunner *et al*. 2001) the Arctic lineage was the first to differentiate from the rest of the Arctic charr complex, and the Acadian group would have separated subsequently from Europe and Asia. However, the most divergent MH Class II β group was the Acadian lineage, which was most differentiated with respect to the Arctic lineage. The separation might indicate differentiation caused by the very different latitudes that both populations inhabit. Geographic areas with different temperatures and water salinities offer a different pathogenic environment (Harvell *et al*. 1999) that would select for different MH alleles (Bryja *et al*. 2006). In terms of diversity, the Arctic lineage was the most diverse whether mDNA or MH Class II β allele composition was used as the basis for analysis. mDNA data, however, could not identify subdivisions within the Arctic lineage, whereas the populations from Lake Aigneau and Resolute Lake were clearly separated using the MH Class II β allele composition.

Amino acid entropy on the β 1 domain

The PBR polymorphism of the MH molecule within a population is intended to increase the possibilities of binding to peptides from diverse array of pathogens. Particular combinations of amino acids in the PBR are better than others at binding to certain pathogenic peptides (Lohm *et al*. 2002), so it is expected that the residues that contribute to the binding task will be fixed within a defined population. On the other hand, because of the evolutionary nature of the MH genes (Bernatchez and Landry 2003), the residues that are not essential for binding current pathogens are kept polymorphic and can vary among different individuals in a population. For any two populations confronted by different pathogens, individuals within any one population will posses polymorphic residues derived from their genetic lineage and different fixed amino acid positions depending on the pathogens encountered within their occupied habitats. The different populations of Arctic charr studied here presented variation in which specific amino acid positions were kept polymorphic, even within populations from the same geographic lineage. This characteristic is in part responsible for their discrimination by the pairwise F_{ST} and exact tests. For example, within the Arctic lineage, the populations studied differed in 8 positions (Table 5-2), 4 that are only variable in the Resolute Lake population, and other 4 that are only variable in the Lake Aigneau population. Within the Atlantic lineage 7 positions were unique, 3 only variable in the Gander Lake population and 4 only in the Lake Sitasjaure population. The observation provides direct evidence of the influence of natural selection on the divergence of the MH Class II β molecule between closely related geographic groups and confirms the value of the MH Class II β population marker for separating populations according to genetic line and adaptation to a specific habitat niche.

Although most of the variation found in the Arctic charr MH class II β gene arose early in the evolutionary history of the species, important changes have taken place during the most recent interglacial colonization period. The changes that have occurred have been influenced by both genetic drift and natural selection. MH Class II β thus allows differentiating between closely related populations that have not been identified as unique by neutral genetic markers. Although such populations may have a common recent origin, they have nonetheless adapted to their niches and its associated pathogenic assemblage. Clearly, adaptation and directive selection can be witnessed by observing the changes in the MH class II β allele sequences and frequencies of studied populations.

Acknowledgements

The authors would like to thank M. O'Connell and B. Dempson for assistance with field collection of samples and J. Witt for assistance with the population analysis. They would also like to thank René J.M. Stet for critical reading of this manuscript (and much more). This research was made possible by Discovery grants from the Natural Sciences and Engineering Research Council of Canada to B.D. (217529-99) and M.P. (155928-92), as well a grant to M.P. from the Canadian International Polar Year Programme.

References

Alekseyev, S. S., Samusenok, V. P., Matveev, A. N., and Pichugin, M. Y.: Diversification, Sympatric Speciation, and Trophic Polymorphism of Arctic Charr, *Salvelinus alpinus* Complex, in Transbaikalia. Environmental Biology of Fishes 64: 97-114, 2002

Allendorf, F. and Thorgaard, G.: Tetraploidyand. the evolution of salmonid fishes. Turner BJ, ed, New York, 1984

Behnke, R. J.: Organizing the diversity of the Arctic charr complex. Biology of the Arctic Charr: Proceedings of the International Symposium on Arctic Charr, pp. 3-21, University of Manitoba Press, 1984

Bernatchez, L., Dempson, B., and Martin, S.: Microsatellite gene diversity analysis in anadromous arctic char, *Salvelinus alpinus*, from Labrador, Canada. Canadian Journal of Fisheries and Aquatic Sciences 55: 1264-1272, 1998

Bernatchez, L. and Landry, C.: MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? Journal of Evolutionary Biology 16: 363-77, 2003

Bernatchez, L., Rhydderch J, and F, Kircheis: Microsatellite gene diversity analysis in landlocked Arctic Charr from Maine. Transactions of the American Fisheries Society 131: 1106-1118, 2002

Bonneaud, C., Chastel, O., Federici, P., Westerdahl, H., and Sorci, G.: Complex Mhc-based mate choice in a wild passerine. Proceedings of the Royal Society of London. Series B. 273: 1111-6, 2006

Brunner, P. C., Douglas, M. R., Osinov, A., Wilson, C. C., and Bernatchez, L.: Holarctic phylogeography of Arctic charr (*Salvelinus alpinus* L.) inferred from mitochondrial DNA sequences. Evolution 55: 573-86, 2001

Bryja, J., Galan, M., Charbonnel, N., and Cosson, J. F.: Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (Arvicolinae). Immunogenetics 58: 191-202, 2006

Consuegra, S., Megens, H. J., Leon, K., Stet, R. J., and Jordan, W. C.: Patterns of variability at the major histocompatibility class II alpha locus in Atlantic salmon contrast with those at the class I locus. Immunogenetics 57: 16-24, 2005a

Consuegra, S., Megens, H. J., Schaschl, H., Leon, K., Stet, R. J., and Jordan, W. C.: Rapid evolution of the MH class I locus results in different allelic compositions in recently diverged populations of Atlantic salmon. Molecular Biology and Evolution 22: 1095-106, 2005b

Dawkins, R., Leelayuwat, C., Gaudieri, S., Tay, G., Hui, J., Cattley, S., Martinez, P., and Kulski, J.: Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. Immunological Reviews 167: 275-304, 1999

de Eyto, E., McGinnity, P., Consuegra, S., Coughlan, J., Tufto, J., Farrell, K., Megens, H. J., Jordan, W., Cross, T., and Stet, R. J.: Natural selection acts on Atlantic salmon major histocompatibility (MH) variability in the wild. Proceedings of the Royal Society of London. Series B. 274: 861-9, 2007

Dorschner, M., Duris, T., Bronte, C., Burnham, Curtis, M., and Phillips, R.: High levels of MHC class II allelic diversity in lake trout from Lake Superior. Journal of Heredity 91: 359-63, 2000

Edgar, R. C.: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-7, 2004

Elliott, J. and Baroudy, E.: The ecology of Arctic charr, *Salvelinus alpinus*, and brown trout, Salmo trutta, in Windermere (northwest England). Nordic Journal of Freshwater Research 71: 33-48, 1995

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

Frost, W.: Breeding Habits of Windermere Charr, *Salvelinus Willughbii* (Gunther), and their Bearing on Speciation of these Fish. Proceedings of the Royal Society of London. Series B, Biological Sciences 163: 232-284, 1965

Goldstein, D. B. and Pollock, D. D.: Launching microsatellites: a review of mutation processes and methods of phylogenetic interference. Journal of Heredity 88: 335-42, 1997

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebo, S., and Stet, R. J.: MHC polymorphism and disease resistance in Atlantic salmon (*Salmo salar*); facing pathogens with single expressed major histocompatibility class I and class II loci. Immunogenetics 55: 210-9, 2003

Hall, T. A.: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. 41: 95-98, 1999

Harvell, C. D., Kim, K., Burkholder, J. M., Colwell, R. R., Epstein, P. R., Grimes, D. J., Hofmann, E. E., Lipp, E. K., Osterhaus, A. D., Overstreet, R. M., Porter, J. W., Smith, G. W., and Vasta, G. R.: Emerging marine diseases--climate links and anthropogenic factors. Science 285: 1505-10, 1999

Hesthagen, T., Hindar, K., and Jonsson, B.: Effects of acidification on normal and dwarf Arctic charr *Salvelinus alpinus* (L.) in a Norwegian lake. Biological Conservation 74: 115-123, 1995

Johnson, L.: The Arctic charr, *Salvelinus alpinus*. In e. E. K. Balon (ed.): Charrs: salmonid fishes of the genus Salvelinus, pp. 19-98, Dr. W. Junk, The Hague, 1980

Kimura, M.: A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16: 111-20, 1980

Kumar, S., Tamura, K., and Nei, M.: MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. Briefings in Bioinformatics 5: 150-163, 2004

Lamont, S. J.: Impact of genetics on disease resistance. Poultry Science 77: 1111-8, 1998

Landry, C. and Bernatchez, L.: Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*). Molecular Ecology 10: 2525-39, 2001

Lohm, J., Grahn, M., Langefors, A., Andersen, O., Storset, A., and von Schantz, T.: Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. Proceedings of the Royal Society of London. Series B. 269: 2029-33, 2002

Maynard, S. and Haigh, J.: The hitchhiking effect of a favorable gene. Genet. Res. 23: 23-35, 1974

Messaoudi, I., Guevara Patino, J. A., Dyall, R., LeMaoult, J., and Nikolich-Zugich, J.: Direct link between mhc polymorphism, T cell avidity, and diversity in immune defense. Science 298: 1797-800, 2002

Miller, K. M., Kaukinen, K. H., Beacham, T. D., and Withler, R. E.: Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. Genetica 111: 237-57, 2001

Nei, M. and Gojobori, T.: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 3: 418-26, 1986

Nordeng, H.: Solution to the "charr problem" based on Arctic Charr (*Salvelinus alpinus*) in Norway. Canadian Journal of Fishery and Aquatic Science 40: 1372-1387, 1993

Ono, H., O'HUigin, C., Tichy, H., and Klein, J.: Major-histocompatibility-complex variation in two species of cichlid fishes from Lake Malawi. Molecular Biology and Evolution 10: 1060-72, 1993

Reche, P. A. and Reinherz, E. L.: Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. Journal of Molecular Biology 331: 623-41, 2003

Sambrook, J., D, R., and J, S.: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, 2001

Schaschl, H., Suchentrunk, F., Hammer, S., and Goodman, S. J.: Recombination and the origin of sequence diversity in the DRB MHC class II locus in chamois (*Rupicapra spp*.). Immunogenetics 57: 108-15, 2005

Schwaiger, F. W., Weyers, E., Buitkamp, J., Ede, A. J., Crawford, A., and Epplen, J. T.: Interdependent MHC-DRB exon-plus-intron evolution in artiodactyls. Molecular Biology and Evolution 11: 239-49, 1994

Skulason, S., Snorrason, S., Noakes, D., Ferguson, M., and Malmquist, H.: Segregation in spawning and early life history among polymorphic Arctic charr, *Salvelinus alpinus*, in Thingvallavatn, Iceland. Journal of Fish Biology 37 (Suppl. A): 225-232, 1989

Steinmetz, M. and Hood, L.: Genes of the major histocompatibility complex in mouse and man. Science 222: 727-33, 1983

Thorpe, J. E.: Present and future charr research: reactions of a salmon biologists. Nordic Journal of Freshwater Research 71: 438-444, 1995

Weitzmann, M. N., Woodford, K. J., and Usdin, K.: DNA secondary structures and the evolution of hypervariable tandem arrays. Journal of Biological Chemestry 272: 9517-23, 1997

Wilson, A. J., Gislason, D., Skulason, S., Snorrason, S. S., Adams, C. E., Alexander, G., Danzmann, R. G., and Ferguson, M. M.: Population genetic structure of Arctic charr, *Salvelinus alpinus* from northwest Europe on large and small spatial scales. Molecular Ecology 13: 1129-42, 2004

Wilson, C., Hebert, P., Reist, J. and Dempson, B.: Phylogeography and postglacial dispersal of Arctic charr *Salvelinus alpinus* in North America. Molecular Ecology 5, 1996

Chapter 6. Differentiation of sympatric Arctic charr morphotypes using Major Histocompatibility Class II genes

Introduction

Arctic charr (*Salvelinus alpinus* L.) are the most northerly distributed and dominant salmonid species in Arctic post-glacial lakes (Johnson 1980). Inhabited lakes were colonized only 10000-15000 years ago (Bernatchez *et al*. 2002) and generally contain few other fish species (Knudsen *et al*. 2006). Most are oligotrophic ecosystems, with many having dissimilar unoccupied feeding niches dominated by a regular epipelagic-litoral area with low benthic diversity and a low productivity profoundal (Knudsen *et al*. 2006). Facilitated by their extraordinary phenotypic plasticity, Arctic charr colonized the recently formed lakes by exploiting the different resource areas through subpopulation adaptation (Alekseyev *et al*. 2002; Allendorf and Thorgaard 1984; Klemetsen *et al*. 2003). As a result, in many post-glacial lakes the Arctic charr presents two distinctive morphotypes that differ in feeding ecology, growth rate, size, age of maturity and morphology (Parker *et al*. 2001). The morphotypes are also distinguishable from one another because they present differential depth distributions, time and/or place of spawning, color and/or other meristic characters. The most common sympatric Arctic charr morphotypes are site related, with the "Normal" morphotype being larger, inhabiting the littoral and/or pelagic areas and after resorting to piscivory (cannibalism) as they grow. By contrast, the smaller or "dwarf" morphotype tends to inhabit profoundal habitats and matures at a younger age and smaller size than the normal morphotype. Dwarf Arctic charr are often pale in colour, have an upper jaw that is larger than the lower jaw; eyes that are larger with respect to the head and longer pectoral fins than the "Normal" form (Jonsson and Jonsson 2001). The dichotomy in mouth position and pectoral fin size often corresponds with trophic adaptation in fish (Lavin and McPhail 1986) and

Arctic charr trophic adaptations has been related in a number of populations (Guiguer *et al*. 2002; Knudsen *et al*. 2006).

The characters that define normal or dwarf Arctic charr morphotypes appear to be inherited, with experiments having demonstrated that progeny of normal Arctic charr have higher growth rate and later sexual maturity than dwarf Arctic charr progeny (Skualson *et al*. 1996; Svedang 1990). Speciation however has not yet occurred, so normal-dwarf hybrid progeny can be obtained and possess intermediate size, growth and age of maturity characteristics (Svedang, 1990).

Lack of complete speciation and the presumably short interval of time since the initiation of differentiation has not allowed sufficient accumulation of genetic drift to facilitate the easy differentiation of the forms with many standard genetic methods (Danzmann *et al*. 1991; Hindar 1986; Magnusson and Ferguson 1987; Volpe and Ferguson 1996). Nevertheless, recent studies using microsatellites on Arctic charr morphotypes with known differences in spawning seasons have shown differentiation of stocks (Westgaard *et al*. 2004), confirming reproductive isolation and indicating that the phenotypic distinction has a genetic component in some instances.

As an alternative to detect genetic differences between Arctic charr morphotypes, we studied the polymorphism of the Major Histocompatibility Class II receptor. The molecule exposes pathogen derived peptides to T-lymphocytes, which then determine whether those particles are self or non-self and initiates an immune response if it is required (Steinmetz and Hood 1983). MH receptors are very polymorphic at the population level, thereby increasing the possibilities of initiating an immune response to an ever growing diversity of pathogens (Dawkins *et al*. 1999; Lohm *et al*. 2002; Messaoudi *et al*. 2002). However, MH alleles that confront local pathogens are rapidly selected either positively or negatively, such that stable populations tend to have common

patterns in their MH polymorphism as determined by the environment that they inhabit (Bernatchez and Landry 2003; Bonneaud *et al*. 2006; Lamont 1998). There are two types of MH receptors present in most jawed vertebrates, MH Class I and MH Class II. While Class I is generally associated with ancestral polymorphism in salmonids (Miller *et al*. 2006; Miller and Withler 1997), Class II has been shown to differentiate recently separated populations (Hardee *et al*. 1995; Landry and Bernatchez 2001; Miller *et al*. 2001). MH Class II receptors are composed of two subunits: Class II α and Class II β. Each one has 5 domains, normally encoded by 5 separated exons. The second exon encodes the α 1 and β 1 domains, respectively, that bind the antigen to be presented to T-lymphocytes (Peptide Binding Region) and, thus are the highly variable area of the molecule (Steinmetz and Hood 1983).

As sympatric Arctic charr morphotypes tend to inhabit different areas of the same lake and have different ecological behaviours, each one has been shown to contain different parasitic loads. For example, normal Arctic charr in Norwegian populations tend to be infected by parasites transmitted by limnetic copepods, whereas dwarf Arctic charr are infected with parasites related to benthic feeding (Knudsen *et al*. 1997). Normal Arctic charr also tend to be infected with a larger variety of pathogens than dwarf Arctic charr, likely due to broader habitat utilization (Knudsen *et al*. 1997).

In this study we expected that the different species of pathogens found in sympatric morphotypes would select for different alleles of MH Class II. Although morphotypes might not have been separated from each other for long enough to have produced detectable genetic drift, the differential selection of MH alleles essential for niche-specific survival would allow genetic isolation of studied morphotypes.

Materials and Methods

Fish samples

Samples from lakes with known sympatric populations at Arctic charr were obtained from Lake Kiryalta (57°08'N; 119°27'E) in the Transbaikal region of eastern Russia, and Gander Lake, Newfoundland (48°55'N; 54°35'W) using multimesh experimental gillnets and classified into morphotype grouping based on colour and size. Lake Kiryalta is geographically divided into two water bodies (Lake Kiryalta -3 and -4) but connected by a 50 m canal, so it was considered as only one geographic area for this study. Classification of Lake Kiryalta samples was based on size and colour descriptions given in Alekseyev *et al*. (2002), whereas classification of Gander Lake samples was based on color and size descriptions given in O'Connell and Dempson (2002). All classifications were performed in collaboration with the authors of the cited studies.

Dorsal muscle samples were cut from each specimen and immersed in 70% ethanol or RNA preserving solution (25 mM $C_6H_5NaO_7$, 10 mM EDTA and 4 M $((NH_4)_2SO_4)$ and shipped to our laboratory.

DNA extraction

Samples in ethanol and RNA preserving solution were placed in TE buffer (100 mM Tris, 1mM EDTA) for 2 hrs, and then incubated for 3 hrs in 100 mM Tris, 10 mM EDTA, 240 mM NaCl, 1% SDS and 300 ug/ml Proteinase K at 55 ºC. Incubation was followed by two extractions, each with one volume of phenol and one extraction with one volume of chloroform. The DNA was

precipitated with 1/10 volumes of 3 M CH2COONa pH 5.2 and 2 volumes of ethanol, and then washed with 70% ethanol. The final pellet was then resuspended in water and stored at 4 ^{o}C until used for PCR.

PCR and cloning

Primers were designed to amplify the second exon and the second intron in MH Class II α and β respectively. The primers were based on the full length sequences of Arctic charr (Chapter 2). The Class II β PCR forward (SAALDABF: 5' – GAT ACT CCT CAA AGG ACC TG– 3') and PCR reverse (SAALDABR2: TAG TCA TAC TGT AAT GKC ACT AC) and Class II α PCR forward (SAALDAAF: 5' - CTG GAT GCA GTG ATT CAG ATG – 3') and PCR reverse (SAALDAAR: 5'- GAC GTG GCA GAT GAG AGT G – 3') primers were used respectively under the same conditions for the PCR reaction: 95 °C of initial denaturation for 2 min, plus 35 cycles of 95 °C of denaturation for 30 sec, 57 °C of primer annealing for 40 sec and 72 °C of extension for 1.5 min. A final extension at 72 °C for 8 min was done to ensure the addition of an A overhang at the end of the complementary strand.

The PCR amplicons were separated by agarose electrophoresis in a 1% gel to verify size and then extracted directly from the gel using GenElute Agarose Spin Columns (Sigma-Aldrich, St Louis, MO). The fragments were inserted in pGEM-T vectors as per manufacturer instructions (Promega Corporation, Madison, WI). XL-blue strains of *E. coli* bacteria were made competent by washing them twice and then re-suspending them with Thawing buffer (10mM PIPES, 15mM CaCl-2H2O, 250 mM KCl, 55mM MnCl2-4H2O, pH 6.7). The bacteria were transformed by the Inoue procedure (Sambrook *et al.* 2001) and grown in LB plates with ampicillin (100 ug/ml) plus

1 uM IPTG and 100 ug/ml X-gal. Eight to twelve white colonies were selected and grown in LB media. The bacteria were used for PCR using the regular forward primer and a reverse primer for Denaturant Gradient Gel Electrophoresis (DGGE), designed to match the end of the second exon (SAALDAADGGER: CGG GCG GGG GCG GCG GGC CGG GCG CGG GGC GCG GCG GGC GCT ATT GTC TCT GGT GGG TYC for Class II α and SAALDABDGGER: CGG GCG GGG GCG GCG GGC CGG GCG CGG GGC GCG GCG GGC GCC AGT ACG GCG CTG TAG for Class II β). The amplicons were separated in a denaturing gradient of 20-45% and 15-50% for Class II α and β, respectively, then run for 15.5 hrs at 60 ℃ (Bio-Rad laboratories, Richmond, CA). To avoid redundancies, only the alleles that differed between samples were selected for DNA sequencing. The plasmids selected for sequencing were extracted from bacterial cultures with the GenElute Plasmid Miniprep Kit (Sigma-Aldrich, St Louis, MA) according to manufacturer instructions, and sequenced using the T7 and Sp6 primers with a big dye terminator v3.1 (Applied Biosystems, Foster City, CA) in a 3130xl Genetic Analyzer (Applied Biosystems, Foster City, CA).

Sequencing and data analysis

The sequences obtained were aligned using the software Muscle (Edgar 2004). An alignment of the previously sequenced full length Class II α and β cDNA sequences (Chapter 2), were used as a guide for the definition of the end of the second exon and the beginning of the intron. The intron and exon sequences were analyzed separately. BioEdit (Hall 1999) was used to define the open reading frames and to translate them into protein sequences. MEGA 4 (Tamura *et al*. 2007) was used to find and correct sequences containing singleton errors, to calculate the diversity of

the sequences among and within populations, the rates of non-synonymous/synonymous substitutions, and to draw the phylogenetic trees using the Nei-Gojobori model (Nei and Gojobori 1986) for non-synonymous substitutions in the exon sequences. Arlequin 3.01 software (Excoffier *et al*. 2005) was used to calculate the allelic frequencies and pairwise F-statistics (F_{ST}).

Results

Gander Lake

α 1 and β 1 domains MH Class II alleles

Twelve α1 and eight β1exon sequences were obtained from normal Arctic charr samples. Nine α1 and 6 β1exon sequences were obtained for dwarf Arctic charr.

Seven alleles of Class II α and 8 of Class II β were found in Gander Lake. Gander Lake dwarf Arctic charr samples had 4 Class II α alleles and Gander Lake Normal had 7, thus three unique Class II α alleles were found in normal Arctic charr, whereas dwarf Arctic char did not have exclusive Class II α alleles. Normal Arctic charr presented 6 Class II β alleles, 5 of them unique for the morphotype. Dwarf Arctic charr had three Class II β alleles, two of them unique. A single Class II β allele was shared among both populations: Saal-DAB*0201, which had a frequency in Gander Lake samples of 0.36.

The mean exon p-distance of Gander Lake samples was 0.078 for Class II α and 0.043 for Class II β. From the total site variation, 0.007 for MH Class II α and 0.006 for Class II β corresponded to diversity between normal and dwarf morphotypes (interpopulation diversity). A mean p-distance of 0.071 for Class II alpha was found in both Gander Lake normal and dwarf Arctic charr. For Class II β the mean p-distance of the normal morphotype was 0.05 and for the dwarf morphotype p-distance was 0.02.

The second exon of MH Class II α in Gander Lake presented a ratio of 3.1 between non-synonymous and synonymous substitutions. The interpopulation non-synonymous substitution

diversity in Gander Lake was 0.083 with extremely low interpopulation synonymous substitutions (0.0015). The rate among both parameters for the diversity between morphotypes had a value of 59.1. The non-synonymous substitutions within morphotype were similar for normal and dwarf Arctic charr, but the synonymous substitutions differed, giving a rate of 4.3 to the Normal Arctic charr and only 2.1 to the dwarf Arctic charr.

The second exon of MH Class II β presented a non-synonymous/synonymous substitution rate of 2.7. Both populations differed in their substitutions; In the normal morphotype non-synonymous and synonymous substitutions measured, respectively, 0.06 and 0.029. In the dwarf morphotype the p-distances were 0.028 and 0.003 respectively, which generated a ratio of 9.4. The average sequence p-distance between Normal and dwarf morphotype in Gander Lake was 0.084 for the second exon of MH Class II α and 0.045 for Class II β.

Intron variation

The mean p-distance in the MH Class II α second intron was 0.043 in Gander Lake, with only 0.006 of diversity between the morphotypes. The normal morphotype diverged 0.052 and the dwarf morphotype only diverged 0.022. The average p-distance between the morphotypes for Class II α second introns was 0.045. For Class II β the average p-distance in the second intron of the gene was 0.024, all of which corresponded to intrapopulation diversity. Both the Normal and dwarf morphotypes had virtually equal intron internal variation, and the p-distance between them was 0.023. The divergence values for the exons and introns, for Class II α and β are given in Table 6-1. The haplotype table is shown in Figure 6-1.

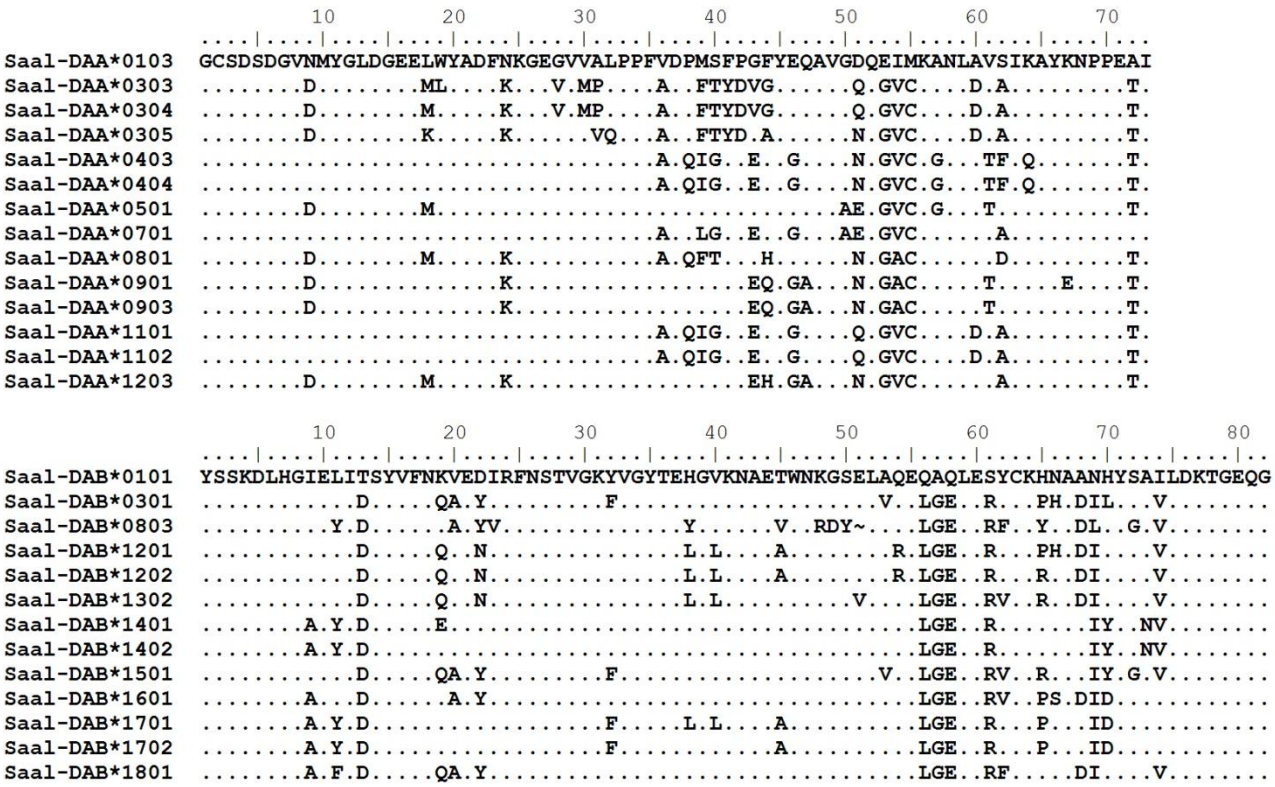


Figure 6-1. Normal and dwarf Arctic charr alleles.

The Figure shows the translated sequence of the second exons found for Gander Lake and Lake Kiryalta for Class II α (alleles DAA) and for Class II β (alleles DAB)

a

		Exon distan	Interpop. divers	Rate dn/ds	Interpop. dn/ds Rate	Intron distan
Class II α	Gander	0.078	0.007	3.067	59.133	0.043
	Kiryalta	0.050	<0.001	3.846	<0.001	0.060
Class II β	Gander	0.043	0.006	2.715	2.442	0.024
	Kiryalta	0.060	0.011	2.201	2.852	0.043

b

	Class II α				Class II β			
	Samp numb	Exon distan	Rate dn/ds	Intron distan	Samp numb	Exon distan	Rate dn/ds	Intron distan
Gander								
Normal	12	0.071	4.300	0.051	8	0.052	2.086	0.025
Dwarf	9	0.071	2.077	0.022	6	0.022	9.360	0.025
Kiryalta								
Normal	17	0.052	3.875	0.069	6	0.069	2.012	0.043
Dwarf	15	0.050	4.067	0.027	6	0.028	2.269	0.040

Table 6-1. Class II α and Class II β divergence in Arctic charr morphotypes.

Table 6-1a shows the divergence values for both genes for Gander Lake and Lake Kiryalta, including the exons mean p-distance, interpopulation diversity, non-synonymous/synonymous (dn/ds) mutation rate, the interpopulation ratio of non-synonymous/synonymous substitutions and the intron mean p-distance. Table 1b shows the sample number and the divergence values for each morphotype in both lakes: Kir (Kiryalta), Gan (Gander).

Population differentiation

Pairwise F-statistics (F_{ST}) between Gander Lake morphotypes Class II α second exon sequences indicated separation of the morphotypes (Table 6-2a). When considering the second introns of the Class II α gene, no isolation of the populations was found (data not shown).

F_{ST} applied to the second exon of the Class II β gene indicated morphotype isolation of both morphotypes (Table 6-2b). The same result was obtained from an analysis of the introns, with a positive FST isolation (data not shown).

Phylogenetic reconstruction

The Nei-Gojorobi phylogenetic tree constructed with the non-synonymous substitutions in the Class II α second exons shows clustering among the dwarf and the normal morphotypes. The tree has three primary well bootstrapped branches, the first clusters two Normal samples, the second includes 80% of the dwarf Arctic charr samples and the third includes 80% of the normal Arctic charr samples. (Figure 6-2a).

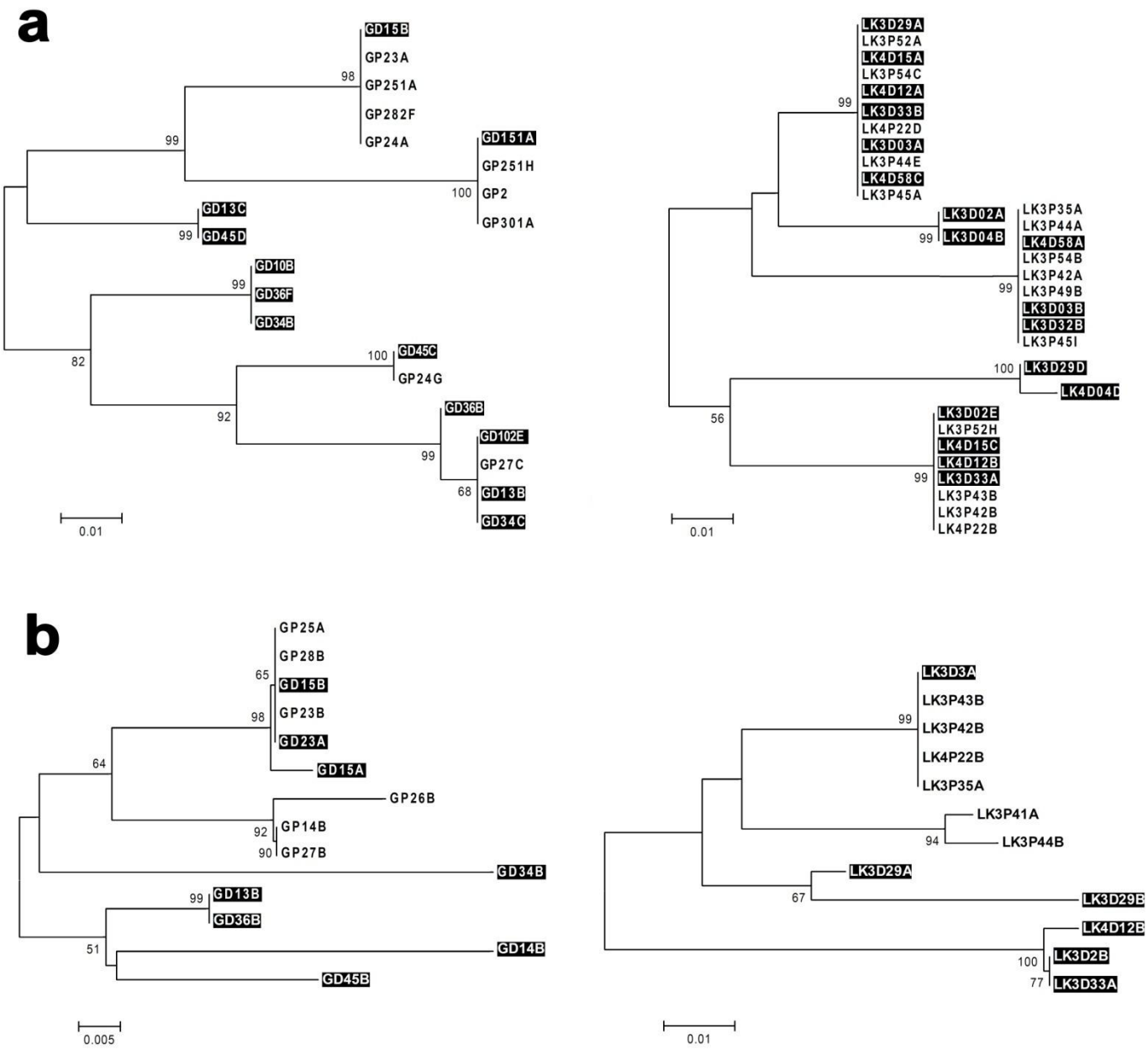
For MH Class II β second exons, the Nei-Gojobori non-synonymous phylogenetic tree also has three well bootstrapped branches, one with a mix of Normal and dwarf Arctic charr samples and the other two branches including exclusively either Normal or dwarf morphotypes samples (Figure 6-2b).

Figure 6-2. Normal and dwarf Arctic charr phylogenetic reconstruction.

The phylogenetic trees constructed from the exon sequences with the Nei-Gojobori model, considering only non-synonymous mutations. Figure 6-2a shows the phylogenetic trees based on Class II α and Figure 6-2b the phylogenetic trees based on Class II β. The trees on the left are for Gander Lake and the trees in the right are for Lake Kiryalta. Normal Arctic charr morphotype samples are shaded, dwarf morphotype samples are not shaded.

Gander Lake

Lake Kiryalta



Lake Kiryalta

α 1 and β 1 domain alleles

Seventeen α1 and six β1exon sequences were obtained from normal Arctic charr samples. Fifteen α1 and six β1exon sequences were obtained for dwarf Arctic charr.

Nine MH Class II α alleles were found in Lake Kiryalta, three of them shared among the two morphotypes. Five alleles were unique to the normal Arctic charr morphotype and only one was unique for the dwarf Arctic charr morphotype. Seven MH Class II β alleles were found with only one (Saal-DAB*1402) shared among morphotypes that had a total frequency of 0.42. Four Class II β alleles were found only in Normal Arctic charr, while other two were present only in dwarf Arctic charr.

In Lake Kiryalta the mean MH Class II p-distance was 0.05 for the α second exon and 0.06 for the β. There was no interpopulation diversity between the two morphotypes in the second exon of the Class II α gene, and 0.01 of interpopulation diversity in the Class II β gene. Both morphotypes had similar internal diversity for Class II α of 0.05. For Class II β the normal morphotype had a mean p-distance of 0.07, while the dwarf morphotype had 0.028. The divergence between morphotypes was 0.05 for Class II α and 0.07 for Class II β.

The rate non-synonymous/synonymous substitutions of the second exon of the MH Class II α gene was 3.9 in Lake Kiryalta, with very close rates for both Normal and dwarf morphotypes, of 3.9 and 4.1 respectively. The total non-synonymous/synonymous rate for the second exon in MH

Class II β was 2.2, with also a very similar rate of 2 and 2.3 for Normal and dwarf Arctic charr respectively.

Intron variation

MH Class II α second intron presented a mean p-distance of 0.06 with 0.012 of inter-morphotype diversity. The divergence was higher in the normal morphotype (0.07) versus the dwarf morphotype (0.028). MH Class II α second intron sequences for the morphotypes diverged by 0.07. The mean p-distance of MH Class II β second introns was 0.043, with 0.044 coming from the normal morphotype and 0.041 from the dwarf morphotype. The average sequence variation between morphotypes was 0.044. The divergence values for Class II α and β, in the second exons and introns are shown in Table 6-1. The haplotype table for Class II α and β is shown in Figure 6-1.

Population differentiation

Pairwise F_{ST} of the second exon sequences of MH Class II α did not suggest the isolation of the morphotypes in Lake Kiryalta. The same result was obtained when comparing the MH Class II α second introns (data no shown). However, F_{ST} estimates did suggest the isolation of the morphotypes when applied over the MH Class II β second exon sequences (Table 6-2). The isolation of Normal and dwarf Arctic charr was also supported by the F_{ST} analysis of the second introns of the MH Class II β gene (data not shown).

Phylogenetic reconstruction

For the second exon MH Class II α tree, there was no clear isolation of the morphotypes. The Class II β second exon on the other hand showed a pattern consistent with the isolation of the morphotypes in Lake Kiryalta, with one of the well bootstrapped phylogenetic branches, including only normal Arctic charr samples and the other two almost exclusively dwarf Arctic charr samples (Figure 6-2b).

a	Gan.Normal	Gan.dwarf.	Kir.Normal	Kir.dwarf
Gan.Normal		0.04285	<0.0001	0.00020
Gan.dwarf	0.15932		0.00179	0.00159
Kir.Normal	0.31199	0.19811		0.55902
Kir.dwarf	0.33186	0.23217	-0.01789	

b	Gan.Normal	Gan.dwarf	Kir.Normal	Kir.dwarf
Gan.Normal		0.04245	0.00060	0.00139
Gan.dwarf	0.15028		0.00218	0.00337
Kir.Normal	0.26268	0.43915		0.02123
Kir.dwarf	0.3392	0.62751	0.31822	

Table 6-2. Population differentiation among Arctic charr morphotypes.

The Tables show the results of the pairwise F_{ST} analysis (5000 permutations) using the exons of the MH Class II α (Table 6-2a) and MH Class II β (Table 6-2b). Below the diagonal are shown the F_{ST} distance values, above the diagonal are shown the p-values for the F_{ST} analysis. P-values ≤ 0.05 were considered significant for this study and are indicative of the isolation of the populations/morphotypes from each other (not shaded).

Discussion

MH Class II α and MH Class II β polymorphism

There was a high divergence of MH Class II α and MH Class II β genes in Gander Lake and Lake Kiryalta Arctic charr expected for populations recently separated, most of the diversity within each lake was intrapopulation. Thus only a small percentage of the observed variation has occurred separately within each morphotype stock. Nevertheless, an extraordinary rate of non-synonymous/synonymous substitutions in the Class II α interpopulation diversity indicated that most of the changes that had occurred since the separation of both morphotypes was related to adaptation to pathogens within the separate ecological niches.

In both, Lake Kiryalta and Gander Lake, for MH Class II α and β genes there has been a loss of total variability in the dwarf morphotype in comparison to normal morphotype. Similar lower variability was also observed in the intron polymorphism (data not shown) and in the number of alleles that each morphotype contained. Thus it is likely that several alleles were lost during the separation of the morphotypes. The loss of MH variability is unlikely to be related to a species bottleneck, since the dwarf morphotype has been found in other studies to be as diverse as the normal morphotype with respect to neutral molecular markers (Volpe and Ferguson 1996; Westgaard *et al*. 2004). On the other hand it has been reported in Norwegian lakes that the normal morphotype is infected by significantly greater higher number of parasite species than the dwarf morphotype (Knudsen *et al*. 1997). Differential parasite loading is likely in Gander Lake, where the normal morphotype has been shown to have more diverse trophic behavior than the dwarf form (Power *et al*. 2005). A diverse trophic behavior is usually related to a higher number

of parasite infections (Knudsen 1995; Lafferty *et al*. 2006; Seilacher *et al*. 2007). Consequently, the lost of MH diversity in the dwarf morphotype may be related to the need for a lower number of MH alleles to confront a lower variety of pathogens. In Gander Lake, the lack of diversity in MH Class II α conveyed to a lower adaptation capacity in the dwarf morphotype, given by the non-synonymous/synonymous substitution ratio, but that appeared to be compensated with a much higher ratio in the Class II β gene. In Lake Kiryalta on the other hand, the lower allelic diversity of dwarf Arctic charr was counterweighted by a higher relative number of non-synonymous substitutions. Despite the lower diversity of dwarf Arctic charr MH genes, their adaptive changes were relatively equivalent to the normal morphotype, which might indicate specific adaptation to confront fewer parasite species, but still very divergent.

MH Class II α and MH Class II β in the adaptation of Arctic charr

In mammals, MH Class II α genes have relatively low polymorphism, and most of the α subunit polymorphic sites are not residues that bind the peptide to be presented to T-lymphocytes (Reche and Reinherz 2003). Mammalian Class II β on the other hand is highly polymorphic and in contrast with what occurs in Class II α, most of the highly variable residues correspond to amino acids that contact the peptide (Reche and Reinherz 2003). In mammals it is inferred that most of the variability in the Peptide Binding Region (PBR) of the MHC Class II dimer is given by the divergence of the β subunit. Because of these reports, almost every fish population study using MH molecules has been done with the β subunit as a marker (Dorschner *et al*. 2000; Landry and Bernatchez 2001; Miller *et al*. 2001). This tendency is also seen in salmonid research, although it

has been proved that teleost Class II α as well as Class II β are both subjected to strong natural selection (Bryja *et al*. 2006; Hardee *et al*. 1995; Piertney and Oliver 2006; Stet *et al*. 2002).

It is not surprising, then, to verify that both α and β subunits in Arctic charr can be equivalently useful in the differentiation of the populations. More interestingly, we observed that for the different populations studied, the relative value of α and β subunits as indicators of differentiation and adaptation could vary, and that some populations are better differentiated by the α subunit while others are better differentiated by the β subunit. For example, while Class II α is more variable than Class II β in the Gander Lake population, in Lake Kiryalta the situation is the opposite. Also, in Gander Lake both morphotypes are differentiated by an extraordinary level of adaptation in Class II α given by an unprecedented inter-morphotype rate of non-synonymous/synonymous substitutions. On the other hand in Lake Kiryalta the inter-morphotype diversity of the alpha gene is nonexistent indicating few recent changes. Even within lakes, in the Gander Arctic charr the α subunit presents more adaptive changes in the normal morphotype while the β subunit presents remarkably higher adaptation levels in the dwarf morphotype. These differences were also observed in the levels of isolation of the stocks within each lake: Gander Lake morphotypes were isolated with similar pairwise differences for Class II α and β, while Lake Kiryalta morphotypes were isolated only when considering Class II β, which pairwise differences doubled the ones shown between Gander lake morphotypes.

Arctic charr Morphotypes

The differentiation of the morphotypes in Lake Kiryalta and Gander Lake is given by the different alleles present in each population as well as allele frequencies that Normal and dwarf

Arctic charr hold. The morphotype-specific alleles are probably necessary to confront specific pathogens that Normal and dwarf Arctic charr find in their niches. In both lakes, the morphotypes have different life histories, inhabit different environments and use different prey recourses: In Lake Kiryalta, the dwarf morphotype consumes mainly zooplankton while the normal morphotype is piscivorous (Alekseyev *et al*. 2002). In Gander Lake, Normal and dwarf charr live at significantly different depths, and their diets and trophic levels have been shown to differ by stomach content and by stable isotope analysis (Power *et al*. 2005). Such habitat and feeding differences should be reflected by the pathogens that Normal and dwarf charr encounter, as has been reported in other lakes where morphotypes exist (Knudsen *et al*. 1997).

Morphotype differentiation

To date, Arctic charr morphotypes living in sympatry have only been differentiated using genetic markers in cases where the reproductive isolation given by differential spawning times has been evident (Westgaard *et al*. 2004). For most of the studies the separation has been obscure (Danzmann *et al*. 1991; Hindar 1986; Magnusson and Ferguson 1987; Volpe and Ferguson 1996) and even in the cases where the division occurred the differentiation values were relatively low (Westgaard *et al*. 2004). Although some reports have shown the heritability of the characters specific to the dwarf or normal morphotype (Svedang 1990), the genetic results to date have enforced the argument that the segregation of morphotypes has its origin in phenotypic plasticity acted on by the varying environment, influences found in the niches of postglacial lakes (Jonsson and Jonsson 2001; Skualson *et al*. 1996). Over time the lakes have acquired environmental stability and thus only recently have facilitated the effective separation of the morphotypes

through genetic drift. The stabilization has been very recent though, and the resulting genetic drift is too small to be detected by most neutral molecular markers. Although differentiation between the morphotype can be established with MH approaches, the low levels of differentiation leave open the question of whether Normal and dwarf morphotype are functionally independent stocks.

The MH molecule variation results in part from genetic drift, but a strong component of the polymorphism in a given population is established by the natural selection of the alleles (Bryja *et al*. 2006; de Eyto *et al*. 2007). MH alleles can be selected out of, or fixed in, a population in an extremely short period of time since the natural selection exerted by pathogens rapidly eliminates the individuals that are not fit for the ecological niche.

The intermediate character of the MH molecules as a population marker, influenced as they were by drift and natural selection, allowed the differentiation among Normal and dwarf morphotypes in both Gander Lake and Lake Kiryalta. Although the separation of the morphotypes occurred relatively recently in evolutionary terms, the fish have adapted rapidly acquiring population specific Class II α and Class II β alleles, which establish them as separate populations with heritable genetic characters allowing them to survive in their specific environments.

The fact that each morphotype possess exclusive alleles implies that movement from one stock (habitat) to another happens rarely, because most fish do not have the necessary alleles for survival in the other niche. Nevertheless, balancing selection in MH genes forces the maintenance of ancestral alleles in the population (Klein 1987), as observed by the occurrence of a few alleles shared among morphotypes. In Gander Lake both morphotypes share one Class II β allele and 4 Class II α alleles, while in Lake Kiryalta the morphotypes share one Class II β

allele and 3 Class II α alleles. The sharing of alleles implies that a few fish from each stock maintain the ability to move into the other niche, a phenomenon that would contribute to keep the state of semi-speciation witnessed in Arctic charr populations where the different morphotypes rarely get to be fully separated (Klemetsen *et al*. 2003). Since the MH Class II dimmers shared by morphotypes are present in only few individuals, for fishery management purposes the morphotypes stocks should still be considered as separate species as has been suggested elsewhere (i.e. Jonsson and Jonsson 2001). Given the low number of MH alleles shared by sympatric populations, the removal of one morphotype by anthropogenic stresses (e.g. fishing, pollution) will not be compensated by the adaptation of the remaining morphotype to the vacant niche except over evolutionary time scales. Therefore, high phenotypic plasticity alone, will not expedite differentiation and new niche occupancy without co-occurring evolutionary change.

References

Alekseyev, S. S., Samusenok, V. P., Matveev, A. N., and Pichugin, M. Y.: Diversification, Sympatric Speciation, and Trophic Polymorphism of Arctic Charr, *Salvelinus alpinus* Complex, in Transbaikalia. Environmental Biology of Fishes 64: 97-114, 2002

Allendorf, F. and Thorgaard, G.: Tetraploidyand. the evolution of salmonid fishes. Turner BJ, ed, New York, 1984

Bernatchez, L. and Landry, C.: MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? Journal of Evolutionary Biology 16: 363-77, 2003

Bernatchez, L., Rhydderch J, and F, Kircheis: Microsatellite gene diversity analysis in landlocked Arctic Charr from Maine. Transactions of the American Fisheries Society 131: 1106-1118, 2002

Bonneaud, C., Chastel, O., Federici, P., Westerdahl, H., and Sorci, G.: Complex Mhc-based mate choice in a wild passerine. Proceedings of the Royal Society of London. Series B. 273: 1111-6, 2006

Bryja, J., Galan, M., Charbonnel, N., and Cosson, J. F.: Duplication, balancing selection and trans-species evolution explain the high levels of polymorphism of the DQA MHC class II gene in voles (Arvicolinae). Immunogenetics 58: 191-202, 2006

Danzmann, R., Ferguson, M., Skualson, S., Snorranson, S., and Noakes, D.: Mitochondrial DNA diversity among four sympatric morphotypes of Arctic charr, *Salvelinus alpinus* L, from Thingvallavatn, Iceland. Journal of Fish Biology 39: 649-659, 1991

Dawkins, R., Leelayuwat, C., Gaudieri, S., Tay, G., Hui, J., Cattley, S., Martinez, P., and Kulski, J.: Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. Immunological Reviews 167: 275-304, 1999

de Eyto, E., McGinnity, P., Consuegra, S., Coughlan, J., Tufto, J., Farrell, K., Megens, H. J., Jordan, W., Cross, T., and Stet, R. J.: Natural selection acts on Atlantic salmon major histocompatibility (MH) variability in the wild. Proceedings of the Royal Society of London. Series B. 274: 861-9, 2007

Dorschner, M. O. D., T. Bronte, C. R. Burnham Curtis, M. K. Phillips, R. B.: High levels of MHC class II allelic diversity in lake trout from Lake Superior. Journal of Heredity 91: 359-63, 2000

Edgar, R. C.: MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-7, 2004

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

Goldsby, R. A., Kindt, T. J., Osborne, B. A., and Kuby, J.: Kuby immunology. W.H. Freeman, New York, 2000

Guiguer, K., Reist, J., Power, M., and Babaluk, J.: Using stable isotopes to confirm the trophic ecology of Arctic charr morphotypes from Lake Hazen, Nunavut, Canada. Journal of Fish Biology 60: 348-362, 2002

Hall, T. A.: BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. Nucleic Acids Symposium Series. 41: 95-98, 1999

Hardee, J. J., Godwin, U., Benedetto, R., and McConnell, T. J.: Major histocompatibility complex class II A gene polymorphism in the striped bass. Immunogenetics 41: 229-38, 1995

Hindar, K.: Genetic differenciation among local populations and morphotypes of Arctic charr, *Salvelinus alpinus*. Biologicl journal of the Linnean society 27: 269-285, 1986

Johnson, L.: The Arctic charr, *Salvelinus alpinus*. In e. E. K. Balon (ed.): Charrs: salmonid fishes of the genus Salvelinus, pp. 19-98, Dr. W. Junk, The Hague, 1980

Jonsson, B. and Jonsson, N.: Polymorphism and speciation in Arctic charr. Journal of Fish Biology 58: 605-638, 2001

Klein, J.: Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Human Immunology 19: 155-62, 1987

Klemetsen, A., Amundsen, P., Dempson, B., Jonsson, B., Jonsson, N., O'Connell, F., and Mortensen, E.: Atlantic salmon Salmo salar L., brown trout Salmo trutta L. and Arctic charr *Salvelinus alpinus* (L.): a review of aspects of their life histories. Ecology of Freshwater Fish 12: 1-59, 2003

Knudsen, R.: Relationships between Habitat, Prey Selection and Parasite Infection in Arctic Charr (*Salvelinus alpinus*). Nordic Journal of Freshwater Research 71: 333-344, 1995

Knudsen, R., Klemetsen, A., Amundsen, P. A., and Hermansen, B.: Incipient speciation through niche expansion: an example from the Arctic charr in a subarctic lake. Proceedings of the Royal Society of London. Series B. 273: 2291-8, 2006

Knudsen, R., Kristoffersen, R., and Amundsen, P. A.: Parasite communities in two sympatric morphotypes of Arctic charr, *Salvelinus alpinus* (L.), in northern Norway. Canadian Journal of Zoology 75: 2003-2009, 1997

Lafferty, K., Dobson, A., and Kuris, A.: Parasites dominate food web links. Proceedings of the National Academy of Science 103: 11211-11216, 2006

Lamont, S. J.: Impact of genetics on disease resistance. Poultry Science 77: 1111-8, 1998

Landry, C. and Bernatchez, L.: Comparative analysis of population structure across environments and geographical scales at major histocompatibility complex and microsatellite loci in Atlantic salmon (*Salmo salar*). Molecular Ecology 10: 2525-39, 2001

Lavin, P. and McPhail, J.: Adaptive divergence of trophic phenotype among freshwater populations of the threespine stickleback (G*asterosteus aculeatus*). Canadian Journal of Fisheries and Aquatic Sciences 43: 2455-2463, 1986

Lohm, J., Grahn, M., Langefors, A., Andersen, O., Storset, A., and von Schantz, T.: Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. Proceedings of the Royal Society of London. Series B. 269: 2029-33, 2002

Magnusson, K. and Ferguson, M.: Genetic analysis of four syrnpatric rnorphs of Arctic charr, *Salvelinus alpinus*, from Thingvallavatn, Iceland. Environmental Biology of Fishes 20: 67-73, 1987

Messaoudi, I., Guevara Patino, J. A., Dyall, R., LeMaoult, J., and Nikolich-Zugich, J.: Direct link between mhc polymorphism, T cell avidity, and diversity in immune defense. Science 298: 1797-800, 2002

Miller, K. M., Kaukinen, K. H., Beacham, T. D., and Withler, R. E.: Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. Genetica 111: 237-57, 2001

Miller, K. M., Li, S., Ming, T. J., Kaukinen, K. H., and Schulze, A. D.: The salmonid MHC class I: more ancient loci uncovered. Immunogenetics 58: 571-89, 2006

Miller, K. M. and Withler, R. E.: Mhc diversity in Pacific salmon: population structure and trans-species allelism. Hereditas 127: 83-95, 1997

Nei, M. and Gojobori, T.: Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Molecular Biology and Evolution 3: 418-26, 1986

O'Connell, F. and Dempson, B.: The biology of Arctic charr, *Salvelinus alpinus*, of Gander Lake, a large, deep, oligotrophic lake in Newfoundland, Canada. Environmental Biology of Fishes 64: 115-126, 2002

Parker, H., Noonburg, E., and Nisbet, R.: Models of Alternative Life-History Strategies, Population Structure and Potential Speciation in Salmonid Fish Stocks. The Journal of Animal Ecology 70: 260-272, 2001

Piertney, S. B. and Oliver, M. K.: The evolutionary ecology of the major histocompatibility complex. Heredity 96: 7-21, 2006

Power, M., O'Connell, F., and Dempson, B.: Ecological segregation within and among Arctic char morphotypes in Gander Lake, Newfoundland. Environmental Biology of Fishes 73: 263-274, 2005

Reche, P. A. and Reinherz, E. L.: Sequence variability analysis of human class I and class II MHC molecules: functional and structural correlates of amino acid polymorphisms. Journal of Molecular Biology 331: 623-41, 2003

Sambrook, J., Russell D, Sambrook J: Molecular Cloning: A Laboratory Manual. Cold Spring Harbor, 2001

Seilacher, A., Reif, W. E., and Wenk, P.: The parasite connection in ecosystems and macroevolution. Naturwissenschaften 94: 155-69, 2007

Skualson, S., Snorrason, S., Noakes, D., and Ferguson, M.: Genetic basis of life history variations among sympatric morphotypes of Arctic char, *Salvelinus alpinus*. Canadian Journal of Fisheries and Aquatic Sciences 53: 1807-1813, 1996

Stet, R. J., de Vries, B., Mudde, K., Hermsen, T., van Heerwaarden, J., Shum, B. P., and Grimholt, U.: Unique haplotypes of co-segregating major histocompatibility class II A and class II B alleles in Atlantic salmon (*Salmo salar*) give rise to diverse class II genotypes. Immunogenetics 54: 320-31, 2002

Svedang, H.: Genetic basis of life-history variation of dwarf and normal Arctic charr, *Salvelinus alpinus* (L.), in Stora Rosjon, central Sweden. Journal of Fish Biology 36: 917-932, 1990

Tamura, K., Dudley, J., Nei, M., and Kumar, S.: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596-9, 2007

Volpe, J. P. and Ferguson, M. M.: Molecular genetic examination of the polymorphic Arctic charr *Salvelinus alpinus* of Thingvallavatn, Iceland. Molecular Ecology 5: 763-72, 1996

Westgaard, J., Klemetsen, A., and Knudsen, R.: Genetic differences between two sympatric morphotypes of Arctic charr confirmed by microsatellite DNA. Journal of Fish Biology 65: 1185-1191, 2004

Chapter 7. General Discussion

General Discussion

MHC molecules arose very early in vertebrate evolution, as they are present in all but jawless vertebrates, and are an integral part of a novel survival strategy against pathogens: the adaptive immune system. Most of the adaptive capacity of the immune system relies on the existence of a wide variety of molecules that are capable of protecting against a wide variety of pathogens (Dawkins *et al*. 1999; Lohm *et al*. 2002). MHC molecules are the main source of the variation at the evolutionary scale, and their polymorphism is dependent both on genetic drift and on the natural selection of the most adequate alleles for a particular niche (Bernatchez and Landry 2003; Bonneaud *et al*. 2006; Lamont 1998).

The use of MH gene polymorphism here allowed the differentiation of Arctic charr populations and sympatric morphotypes. MH Class II, and particularly the β subunit, appeared to be the best population marker. Most Arctic charr Class I α gene polymorphisms were ancestral, as has been seen in other salmonids (Shum *et al*. 2001). They were not derived during recent colonization events and thus they were not useful in differentiating the populations. However, UGA, a non classical gene in other species, actually showed recent polymorphism in Arctic charr and was capable of differentiating among phylogeographic lineages.

MH Class II linkage

When comparing the alleles obtained for all the MH Class II genes, it is apparent that they are not distributed randomly amongst individuals. Although substantial variation was observed, in

many cases alleles were segregated in pairs, and certain Class II α allele families were often found with specific Class II β allele families in the same individual. Class II α and Class II β genes are located in the same chromosome in fish (Bingulac-Popovic *et al*. 1997) and, therefore, may be evolutionarily linked and tend to recombine and segregate together. On another hand, allele combinations were generally conserved within populations, but not between them. Although some alleles were present in more than one population, the combinations tended to be unique for each stock. It has been reported in salmonids that there is a synergic effect of certain allele combinations towards pathogen resistance (Kjoglum *et al*. 2006). Thus it is possible that in Arctic charr, specific combinations of alleles are selected by the pathogens of each ecosystem, in a higher hierarchy of specificity for MH polymorphism.

MH genes in Arctic charr populations

Each MH gene in the genome has been selected under particular evolutionary forces and since some genes are older than others (Sato *et al*. 2001), each Arctic charr MH gene has acquired distinct variability.

The highest MH divergence found in this study was for Class I α genes UBA and UCA, and the lowest was for Class I α genes UGA and UEA. Class II genes had an intermediate total variation, but the highest ratios of non-synonymous/synonymous substitutions. The ratios of the Class I genes were all similar, except the UEA genes whose ratios indicated strong positive selection. The UGA Class I gene compensated for low total divergence with an outstanding inter-lineage variation that suggested independent recent changes for each population. Considering the positive non-synonymous/synonymous substitution rate, the variations are probably adaptations

to the new environment. Class II β genes also held high interpopulation variation, and also the highest non-synonymous/synonymous substitution ratios. The strong selection seen on the Class II β gene is reflected by the fact that it was the only marker capable of differentiating all the populations studied, and consequently the phylogeographic lineages previously identified using mDNA (Brunner *et al*. 2001).

Interestingly, none of the MH genes clearly separated the populations of Lake Sitasjaure and Resolute Lake. Even for Class II β the p-value fell within the error margin of the method. Taxonomy studies using meristic characters and studies using non-selected genetic markers have indicated these two populations should belong to separate lineages (Behnke 1984; Brunner *et al*. 2001; Savvaitova 1995), so the similarity between the two seen here cannot have a phylogenetic connotation. Interestingly, Lake Sitasjaure and Resolute Lake are located at relatively similar latitudes, and are the northernmost samples that we could obtain. Both lakes Sitasjaure and Resolute are very oligotrophic and the biota they can sustain is very limited (Markager and Vincent 2001; Sahlberg and Rahm 2005). The environmental limitations also apply to microorganisms (Kling *et al*. 1992), and only specialized pathogens might survive in such waters (Mock and Thomas 2005; Staley and Gosink 1999). Thus it is plausible that pathogens inhabiting both Lake Sitasjaure and Resolute Lake have common features that select for similar MH alleles in the resident Arctic charr populations. The Atlantic and Arctic phylogeographic lineages defined by Brunner are farther genetically from each other, than each is to the Siberian lineage according to mDNA (Brunner *et al*. 2001), so the specifically selected alleles would have been maintained from the common ancestor of both populations, but deleted from the Siberian Arctic charr lineage. On the other hand the similarity between Lake Sitasjaure and Resolute Lake recalls the traditional division of the Arctic charr complex based on meristic characters, where

the Siberian and the Arctic lineages are within one group adjacent to the Atlantic lineage (Behnke 1984). The similarity between the data obtained by traditional taxonomy and our studies is not surprising, since meristic characters and MH genes are both subject to strong environmental selection that is specific for an ecological niche.

On the other hand, the information compiled using MH gene polymorphism data gives us more concrete information on the specialization of each stock. This is because the alleles of MH genes are discrete units, unlike meristic characters which, are continuous variables. Since each individual carries relatively few MH alleles (Grimholt *et al*. 2003), most alleles are directly related to the resistance to local pathogens. Holding those alleles allows survival in one ecological niche or the other, but intermediate conditions should be rare. Thus MH gene polymorphism indicates a very clear separation between stocks of Arctic charr, because it depends on discrete characters that are essential for survival within particular ecosystems.

Arctic charr sympatric stocks

It has been suggested that the maintenance over time of more than one Arctic charr morphotype in the same water body is due to the fact that intermediate forms are less efficient in food consumption (Jonsson and Jonsson 2001). Complementary to that, this study can now add that the separation might also be maintained by a differential adaptation to pathogens as seen by MH allele content. A stock that is settled and adapted to a particular ecological niche possesses specific MH alleles. Migrating to a different niche requires dealing with the novel pathogens in that niche and for most individuals this is not possible. Nevertheless, because of the evolutionary forces that guide MH evolution, some very old alleles are always kept in the population (Klein

1987). Therefore, few individuals might still hold some alleles that permit their migration to a different stock, a phenomenon that would contribute to keeping the state of semi-speciation witnessed in Arctic charr populations, where the different morphotypes rarely become fully separated (Klemetsen *et al*. 2003). However, hybrids (intermediate forms), would by definition only have 50% of the specialized immune response capability that native forms from the ecological niche possess. This suggests intermediate forms would be less competitive and would remain as only a small fraction of the population. The presence of wide polymorphism and the permanence of ancestral alleles also suggests that one of the morphotypes could repopulate the other's niche upon selective depletion of the other. The potential for cross-recovery should be directly related with the number of individuals holding the MH alleles needed to survive in the other morphotype's niche. The number of individuals examined in the present work was too few to estimate these values, but the fact that very few alleles were shared between the sympatric stocks, indicates that the chances of cross-recovery, although present, are likely to be low. Furthermore, the recovery would occur only over evolutionary time, not rapidly enough to avoid a population collapsing. As has been previously suggested (Jonsson and Jonsson 2001), sympatric stocks should be considered different species for management purposes. Naturally, and given the variability of the Arctic charr complex, there might be cases where sympatric stocks are less specialized in their immune system and other characters, and then the possibilities for cross-repopulation might increase.

MH genes and conservation genetics

Since MH polymorphism gives information on the relatedness of the individuals and their interaction with the environment, it can be used to define Evolutionary Significant Units (ESU), populations with unique characteristics that make them an indispensable piece of certain ecosystem (Miller *et al*. 2001). Defining ESU's allows implementation of separate management strategies for different units of the same species. Although the separation of populations using MH genes polymorphism may obscure conclusions about their exact phylogenetic relationship, analytical results indicate how adapted the populations are to their own environments. From the conservation and management point of view such information is very useful because a population that is highly adapted to its ecosystem may be a strong interactor whose remove could have significant impact on the community as a whole (Palmer *et al*. 1997; Vila 1998).

References

Behnke, R. J.: Organizing the diversity of the Arctic charr complex. Biology oftheSlArctic Charr: Proceedings ofthe International Symposium on Arctic Charr, pp. 3-21, University of Manitoba Press, 1984

Bernatchez, L. and Landry, C.: MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? Journal of Evolutionary Biology 16: 363-77, 2003

Bingulac-Popovic, J., Figueroa, F., Sato, A., Talbot, W. S., Johnson, S. L., Gates, M., Postlethwait, J. H., and Klein, J.: Mapping of mhc class I and class II regions to different linkage groups in the zebrafish, Danio rerio. Immunogenetics 46: 129-34, 1997

Bonneaud, C., Chastel, O., Federici, P., Westerdahl, H., and Sorci, G.: Complex Mhc-based mate choice in a wild passerine. Proceedings of the Royal Society of London. Series B, Biological Sciences 273: 1111-6, 2006

Brunner, P. C., Douglas, M. R., Osinov, A., Wilson, C. C., and Bernatchez, L.: Holarctic phylogeography of Arctic charr (Salvelinus alpinus L.) inferred from mitochondrial DNA sequences. Evolution 55: 573-86, 2001

Dawkins, R., Leelayuwat, C., Gaudieri, S., Tay, G., Hui, J., Cattley, S., Martinez, P., and Kulski, J.: Genomics of the major histocompatibility complex: haplotypes, duplication, retroviruses and disease. Immunological Reviews 167: 275-304, 1999

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebo, S., and Stet, R. J.: MHC polymorphism and disease resistance in Atlantic salmon (Salmo salar); facing pathogens with single expressed major histocompatibility class I and class II loci. Immunogenetics 55: 210-9, 2003

Jonsson, B. and Jonsson, N.: Polymorphism and speciation in Arctic charr. Journal of Fish Biology 58: 605-638, 2001

Kjoglum, S., Larsen, S., Bakke, H. G., and Grimholt, U.: How specific MHC class I and class II combinations affect disease resistance against infectious salmon anaemia in Atlantic salmon (Salmo salar). Fish and Shellfish Immunology 21: 431-41, 2006

Klein, J.: Origin of major histocompatibility complex polymorphism: the trans-species hypothesis. Human Immunology 19: 155-62, 1987

Klemetsen, A., Amundsen, P., Dempson, B., Jonsson, B., Jonsson, N., O'Connell, F., and Mortensen, E.: Atlantic salmon Salmo salar L., brown trout Salmo trutta L. and Arctic charr Salvelinus alpinus (L.): a review of aspects of their life histories. Ecology of Freshwater Fish 12: 1-59, 2003

Kling, G., Fry, B., and O'Brien, W.: Stable Isotopes and Planktonic Trophic Structure in Arctic Lakes. Ecology 73: 561-566, 1992

Lamont, S. J.: Impact of genetics on disease resistance. Poultry Science 77: 1111-8, 1998

Lohm, J., Grahn, M., Langefors, A., Andersen, O., Storset, A., and von Schantz, T.: Experimental evidence for major histocompatibility complex-allele-specific resistance to a bacterial infection. Proceedings of the Royal Society of London. Series B, Biological Sciences 269: 2029-33, 2002

Markager, S. and Vincent, W.: Light absorption by phytoplankton: development of a matching parameter for algal photosynthesis under different spectral regimes. Journal of Plankton Research 23: 1373-1384, 2001

Miller, K. M., Kaukinen, K. H., Beacham, T. D., and Withler, R. E.: Geographic heterogeneity in natural selection on an MHC locus in sockeye salmon. Genetica 111: 237-57, 2001

Mock, T. and Thomas, D.: Recent advances in sea-ice microbiology. Environmental Microbiology 7: 605-619, 2005

Palmer, M., Ambrose, R., and Poff, L.: Ecological theory and community restoration ecology. Restoration Ecology 5: 291-300, 1997

Sahlberg, J. and Rahm, L.: Light limitation of primary production in high latitude reservoirs. Hydrology and Earth System Sciences 9: 707-720, 2005

Sato, A., Mayer, W. E., Tichy, H., Grant, P. R., Grant, B. R., and Klein, J.: Evolution of Mhc class II B genes in Darwin's finches and their closest relatives: birth of a new gene. Immunogenetics 53: 792-801, 2001

Savvaitova, K.: Patterns of diversity and processes of speciation in Arctic Charr. Nordic Journal of Freshwater Research 71: 81-91, 1995

Shum, B. P., Guethlein, L., Flodin, L. R., Adkison, M. A., Hedrick, R. P., Nehring, R. B., Stet, R. J., Secombes, C., and Parham, P.: Modes of salmonid MHC class I and II evolution differ from the primate paradigm. Journal of Immunology 166: 3297-308, 2001

Staley, J. and Gosink, J.: Poles apart: Biodiversity and Biogeography of sea ice bacteria. Annual reviews in Microbiology 53: 189-215, 1999

Vila, M.: Efectos de la diversidad de especies en el funcionamiento de los ecosistemas. Orsis 13: 105-117, 1998