TURUN YLIOPISTON JULKAISUJA
ANNALES UNIVERSITATIS TURKUENSIS

SARJA - SER. AII OSA - TOM. 220

BIOLOGICA - GEOGRAPHICA - GEOLOGICA

Phylogeography and population genetics of north European Atlantic salmon (*Salmo salar* L.)

Anni Tonteri

TURUN YLIOPISTO
Turku 2008

From the Laboratory of Genetics, Department of Biology
University of Turku
Turku, Finland

Supervised by:
Professor Craig R. Primmer
University of Turku
Finland

Professor Jaakko Lumme
University of Oulu
Finland

Reviewed by:
Senior Research Associate Jukka Vuorinen
University of Joensuu
Finland

Senior Scientist Einar Eg Nielsen
Danish Institute for Fisheries Research
Denmark

Examined by:
Research Geneticist Timothy L. King
U.S. Geological Survey
USA

Cover image: Atlantic salmon spawning in a redd.
Jodie O. Pennanen, © Atlantic Salmon Federation

ISBN 978-951-29-3495-9 (PRINT)
ISBN 978-951-29-3496-6 (PDF)
ISSN 0082-6979
Painosalama Oy - Turku, Finland 2008

To my father, the best dad in the world

Phylogeography and population genetics of north European Atlantic salmon (*Salmo salar* L.)

Anni Tonteri

This thesis is based on the following articles, which are referred to in the text by their Roman numerals:

I Tonteri A, Titov S, Veselov A, Zubchenko A, Koskinen MT, Lesbarrères D, Kaluzchin S, Bakhmet I, Lumme J & Primmer CR (2005). Phylogeography of anadromous and non-anadromous Atlantic salmon (*Salmo salar*) from northern Europe. *Annales Zoologici Fennici* 42: 1-22.

II Tonteri A, Veselov AJe, Zubchenko A, Lumme J & Primmer CR (2008). Microsatellites reveal clear genetic boundaries among Atlantic salmon (*Salmo salar* L.) populations from Barents and White Seas. *Manuscript*.

III Tonteri A, Veselov AJe, Titov S, Lumme J & Primmer CR (2007). The effect of migratory behaviour on genetic diversity and population divergence: A comparison of anadromous and freshwater Atlantic salmon *Salmo salar*. *Journal of Fish Biology* 70 (Supplement C): 381–398.

IV Tonteri A, Vasemägi A, Lumme J & Primmer CR (2008). Evolutionary immunogenetics: Do immune relevant loci exhibit different selection pressures compared to random loci in Atlantic salmon (*Salmo salar* L.)? *Manuscript*.

ABSTRACT

Although abundant in the number of individuals, the Atlantic salmon may be considered as a threatened species in many areas of its native distribution range. Human activities such as building of power plant dams, offshore overfishing, pollution, clearing of riverbeds for timber floating and badly designed stocking regimes have diminished the distribution of Atlantic salmon. As a result of this, many of the historical populations both in Europe and northern America have gone extinct or are severely depressed. In fact, only 1% of Atlantic salmon existing today are of natural origin, the rest being farmed salmon. All of this has lead to a vast amount of research and many restoration programmes aiming to bring Atlantic salmon back to rivers from where it has vanished. However, many of the restoration programmes conducted thus far have been unsuccessful due to inadequate scientific research or lack of its implementation, highlighting the fact that more research is needed to fully understand the biology of this complex species.

The White and Barents Seas in northwest Russia are among the last regions in Europe where Atlantic salmon populations are still stable, thus forming an important source of biodiversity for the entire European region. Salmon stocks from this area are also of immense economic and social importance for the local people in the form of fishing tourism. The main aim of this thesis was to elucidate the post-glacial history and population genetic structure of north European and particularly northwest Russian Atlantic salmon, both of which are aspects of great importance for the management and conservation of the species. Throughout the whole thesis, these populations were studied by utilizing microsatellites as the main molecular tool.

One of the most important discoveries of the thesis was the division of Atlantic salmon from the White and Barents Seas into four separate clusters, which has not been observed in previous studies employing nuclear markers although is supported by mtDNA studies. Populations from the western Barents Sea clustered together with the northeast Atlantic populations into a clearly distinguishable group while populations from the White Sea and eastern Barents Sea were separated into three additional groups. This has important conservation implications as this thesis clearly indicates that conservation of populations from all of the observed clusters is warranted in order to conserve as much of the genetic diversity as possible in this area.

The thesis also demonstrates how differences in population life histories within a species, migratory behaviour in this case, and in their phylogeographic origin affect the genetic characteristics of populations, namely diversity and divergence levels. The anadromous populations from the Atlantic Ocean, White Sea and Barents Sea possessed higher levels of genetic diversity than the anadromous populations form the Baltic Sea basin. Among the non-anadromous populations the result was the opposite: the Baltic freshwater populations were more variable. This emphasises the importance of taking the life history of a population into consideration when developing conservation strategies: due to the limited possibilities for new genetic diversity to be generated via gene flow, it is expected that freshwater Atlantic salmon populations would be more vulnerable to extinction following a population crash and thus deserve a high conservation status.

In the last chapter of this thesis immune relevant marker loci were developed and screened for signatures of natural selection along with loci linked to genes with other functions or no function at all. Also, a novel landscape genomics method, which combines environmental information with molecular data, was employed to investigate whether immune relevant markers displayed significant correlations to various environmental variables more frequently than other loci. Indications of stronger selection pressure among immune-relevant loci compared to non-immune relevant EST-linked loci was found but further studies are needed to evaluate whether it is a common phenomenon in Atlantic salmon.

CONTENTS

INTRODUCTION

In order to adapt to changes in their environment organisms need to have heritable genetic variation that evolution can act upon. Loss of genetic diversity leads to increased extinction risk as a species' ability to cope with environmental changes reduces. This is one of the issues, among many others, that conservation genetics deals with and aims to minimize by employing population genetic techniques (Frankham *et al.* 2002).

Simultaneously, it is vitally important to understand the effects that the last glaciations have had on the distribution and genetic diversity of various organisms in order to fully comprehend the genetic structure of populations and the implications for conservation. For freshwater fish species the assumption is that during the last ice age they had to escape the advancing ice masses and survive in different refugia as old freshwater habitats were destroyed. Multiple glacial advances and retreats caused further disturbances, while in some regions large lakes were formed at the edge of the ice sheet providing opportunities for dispersal over vast areas. Today, all of this can be seen as lower levels of genetic diversity and lower sister species divergence in freshwater fish species from formerly glaciated areas compared to species from non-glaciated areas (Bernatchez & Wilson 1998).

The Atlantic salmon is one of the species that have been considerably influenced by the last glaciations as vast ice sheets covered large parts of its present day distribution range both in Europe and in North America. The species is highly valued by the fishing industry and recreational fishermen and certainly deserves the names, such as *the king of the underwater world* (Verspoor 2007) and *the icon of the salmonids* (Vähä 2007), that it has also been called. But despite the cultural and economic esteem people give to Atlantic salmon, humans, through their activities, have diminished many natural Atlantic salmon populations or even driven them to extinction (Parrish *et al.* 1998, WWF 2001). In fact, only 1% of Atlantic salmon existing today are of natural origin, the rest being farmed salmon (Verspoor 2007). All of this has lead to a vast amount of research and many restoration programmes aiming to bring Atlantic salmon back to rivers from where it has vanished. However, many of the restoration programmes conducted thus far have been unsuccessful due to inadequate scientific research or lack of its implementation (Verspoor 2007), highlighting the fact that more research is needed to fully understand the biology of this complex species. Especially the genetics of the species needs to be better comprehended as it lies at the heart of the biological character, survival and reproduction of Atlantic salmon (Verspoor 2007). In addition, the glacial and postglacial history of Atlantic salmon needs to be resolved as it is one of the key determinants of the species genetic structure.

Phylogeography

Phylogeography is an integrative field of study that combines information from several disciplines including molecular and population genetics, ethology, demography, phylogenetics, and historical geography to explain the genetic structure of modern populations (Avise 1998). In particular, it focuses on how historical factors, often the Weichselian glaciations ca. 117 000-10 000 years ago (e.g. Svendsen *et al.* 2004), have influenced the geographical distribution of gene lineages. The most common molecular tool used in phylogeography has been mitochondrial DNA (mtDNA), whose relatively high mutation rate is well suited to studies examining events that took place during the last few million years, while for the study of more recent events (the last 10 000 years) markers with higher mutation rate, such as microsatellites, are needed (Hewitt 2004).

Phylogeography is important in the context of conservation and can be used to identify management units and evolutionary significant units (e.g. Avise 2000). A management unit (MU) may be defined as a population that exchanges migrants with other populations but so few that it will remain as a demographically independent unit at the present time (Avise 2000); such populations are typically exemplified by shallow phylogeographic separation. Evolutionarily significant units (ESU's), first defined by Ryder (1986), display deep phylogeographical separation (Avise 2000). Hence, ESU's are populations that have a distinct, long-term evolutionary history and, according to Moritz (1994), can be defined as populations that are monophyletic for mtDNA alleles and show significant divergence in allele frequencies at nuclear loci. Although this definition of ESU's has received some criticism over the years and may not be applicable for all species (e.g. Crandall *et al.* 2000, Fraser & Bernatchez 2001), it is often useful when making conservation plans for individual species (Avise 2000). Furthermore, the concept may be employed to identify geographical regions within which several species display phylogenetically distinct populations, thus making the area a candidate for high conservation priority (Avise 2000).

Advances and retreats of the Weichselian ice sheets and formation of ice-dammed lakes

Three times during the Weichselian period (ca. 117 000-10 000 years ago) the Eurasian Artic was affected by major glaciations (e.g. Svendsen *et al.* 2004). The maximum early Weichselian ice sheet (ca. 90 000-80 000 years ago; Figure 1a) covered the Barents and Kara Seas from Svalbard in the west to the Taimyr Peninsula in the east and extended all the way to mainland Russia (Svendsen *et al.* 1999, 2004). In Scandinavia the ice sheet was much more restricted and covered only mainland Norway, northern and central Sweden and Finnish Lapland (Svendsen 2004). The ice sheet blocked north-flowing rivers, which led to the formation of large ice-dammed lakes at the edge of the ice. One of these paleolakes was Lake Komi, which was situated in the Pechora lowlands to the west of the Ural Mountains and extended as far as the White Sea basin (Mangerud *et al.* 2001, 2004).

The first regrowth of the ice sheet occurred during the Middle Weichselian when the maximum ice sheet (ca. 60 000-50 000 years ago; Figure 1b) extended again from Svalbard to the Russian mainland but did not reach as far to the east as during the Early Weichselian (Svendsen *et al.* 2004). In Scandinavia, however, the glaciation was more extensive as the Baltic Sea basin and all of Norway, Sweden and Finland were covered by ice (Svendsen *et al.* 2004). During this period ice dammed lakes existed most likely in the White Sea basin and in the Pechora Lowlands (Mangerud *et al.* 2004).

The most recent glacial readvance reached its maximum extent about 20 000 years ago during the Late Weichselian when the ice sheet covered large areas of northern Europe including most of the British Isles and Denmark, northern parts of Germany and Poland, the Baltic Countries and all of Norway, Sweden and Finland (e.g. Svendsen *et al.* 1999, 2004; Figure 1c). At this time the ice cover probably did not reach all the way to mainland Russia (Svendsen 1999). The eastern limit of the ice sheet is not fully resolved but most likely its extent was limited allowing rivers to freely discharge into the Arctic Ocean (Mangerud *et al.* 2004). However, many ice-dammed lakes were formed at the edge of the Scandinavian Ice Sheet and later as the glaciers started to recede approximately 14 200 years ago the Baltic Ice Lake was formed (e.g. Björck 1995, Mangerud *et al.* 2004; Figure 1c) followed by the brackish and freshwater Yoldia Sea (ca. 11 300-10 700 years ago), freshwater Ancylus Lake (ca. 10 700-10 100 years ago) and brackish Littorina Sea stages (ca. 10 100-800 years ago; Andrén *et al.* 2000).

Ladoga and Onega, two large lakes belonging to the Baltic Sea basin, were

deglaciated by 12 750 years ago after which Ladoga formed a bay of the Baltic Ice Lake until the end of the ice lake stage (Björck 1995, Saarnisto & Saarinen 2001). Although the Baltic Ice Lake never reached Lake Onega due to its high elevation (Saarnisto *et al.* 1995), a connection between Ladoga and Onega has existed throughout most post-glacial periods (Saarnisto *et al.* 1995). Later, ca. 10 700-10 100 years ago (Andrén *et al.* 2000), Ladoga formed a bay of the Anculys Lake (Björck 1995).

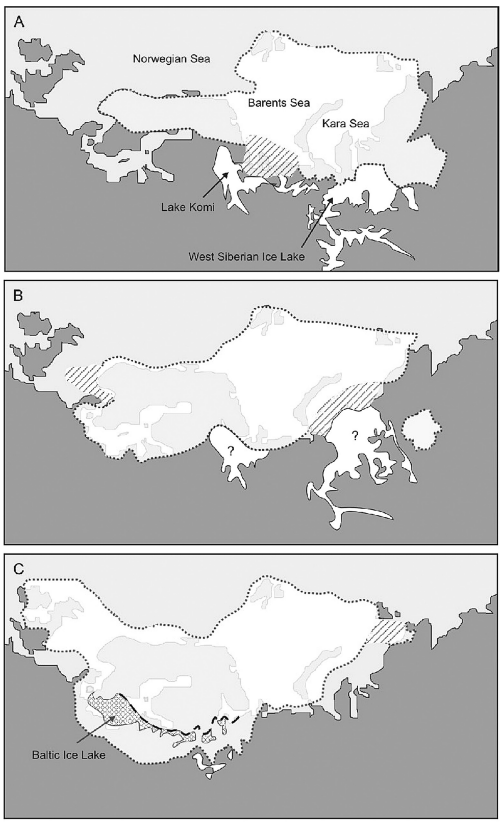

Figure 1. The extent of the Eurasian ice sheet during a) Early (ca. 90 000 years ago), b) Middle (ca. 60 000 years ago), and c) Late Weichselian glacial maxima (ca. 20 000 years ago) and the ice lakes that existed during those times (after Mangerud *et al.* 2004). The hatching denotes areas where the extent of the ice sheet has not been resolved and in c) the extent of the ice sheet ca. 14 000 years ago is represented with a dashed line and the location of the Baltic Ice Lake with cross hatching. Ice lakes in b) are only hypothetical and their existence has not been proven (Mangerud *et al.* 2004).

The Atlantic salmon

The Atlantic salmon (*Salmo salar* L.) belongs to the family Salmonidae and is, together with the brown trout (*Salmo trutta* L.), the only representative of the genus *Salmo* (Webb *et al.* 2007) although the addition of species *Acantholingua ohridana* and *Sternopygus obtusirostis* to the genus has been suggested (Phillips *et al.* 2000, Snoj *et al.* 2002, Crespi & Fulton 2004). Atlantic salmon are native to the northern Atlantic Ocean and the indigenous distribution range extends in northern America from the Hudson river in New York to outer Ungava Bay in Quebec, and in Europe from Iceland southwards to the British Isles and the Douro river in northern Portugal, and eastwards to the Baltic Sea and the Pechora river in northwest Russia (MacCrimmon & Gots 1979).

The Atlantic salmon is a migratory fish and has two different life-history forms depending on whether or not the populations undergo a marine migration phase. Most of the populations are sea-migrating (i.e. anadromous), meaning that they spawn and spend their juvenile years (up to five years) in rivers after which they migrate to the sea to feed and mature (Hutchings & Jones 1998). After spending one to four winters at sea, mature individuals usually return to their natal rivers for spawning (Mills 1989, Hutchings & Jones 1998). In addition to the sea-migrating ones, populations that spend their entire life cycle in freshwater exist as well. These non-anadromous populations were formed following the last ice age as rapid land upheaval isolated the populations from the sea and created lakes large enough for the freshwater-adapted refugial populations to thrive (Berg 1985). In such populations smolts migrate from a river to a lake and back again as their migration to and from the sea is inhibited by geographical barriers. In Europe, freshwater Atlantic salmon can be found in 13 locations in Norway, Sweden, Finland, and northwest Russia (MacCrimmon & Gots 1979, Berg 1985) and in North America in several small lakes in Maine, New Brunswick, Nova Scotia, Quebec,

Labrador, and Newfoundland (MacCrimmon & Gots 1979). The only known exception to this is the non-anadromous Atlantic salmon in river Namsen in Norway that do not migrate to a lake but spend their entire life cycle in the river (Berg 1985).

Due to its strong homing behaviour, the Atlantic salmon is naturally substructured into genetically differentiated and reproductively isolated populations (Ståhl 1987). There is a clear division between the North American and European salmon (Ståhl 1987, Bermingham *et al.* 1991, McConnell *et al.* 1995a, b, Verspoor *et al.* 1999, King *et al.* 2001), and the European salmon can further be divided into two groups: the Eastern Atlantic and the Baltic salmon (Ståhl 1987, Bermingham *et al.* 1991, Bourke *et al.* 1997, Verspoor *et al.* 1999, Nilsson *et al.* 2001, Consuegra *et al.* 2002). Some of the non-anadromous populations group to the European cluster and some to the Baltic cluster (Ståhl 1987).

North European Atlantic salmon colonized their current habitats following the last glaciation but despite abundant research, no consensus has been reached on the origin of the current populations. Recently it has been suggested that the Baltic Sea was colonized exclusively from a single eastern ice lake refugium (Nilsson *et al.* 2001), or from up to three distinct refugia: the Gulf of Bothnia from an Atlantic refugium, the Gulf of Finland from an eastern ice lake refugium, and the southern Main Basin from a refugium that was presumably located in the basin of the rivers Neman, Vistula, Odra, and Elbe (Säisä *et al.* 2005). The freshwater populations from Lakes Ladoga and Onega have been proposed to originate from the eastern ice lake refugium (Nilsson *et al.* 2001, Säisä *et al.* 2005).

For northwest Russian Atlantic salmon colonization from two directions has been suggested. Evidence for colonization of the northern Kola Peninsula from the eastern Atlantic Ocean (i.e. from the Iberian peninsula, the British Isles, and the North Sea) has been presented in several studies (Verspoor *et al.* 1999, Consuegra *et al.* 2002, Asplund *et al.* 2004, Makhrov *et al.* 2005, Säisä *et al.* 2005) as well as immigration from the western Atlantic Ocean (Asplund *et al.* 2004, Makhrov *et al.* 2005). The Atlantic salmon found in the area of the White Sea and the eastern Barents Sea probably originate from a northeastern glacial refugium (Kazakov & Titov 1991, Asplund *et al.* 2004) although some support for western Atlantic immigration has been found on the west coast of the White Sea as well (Makhrov *et al.* 2005). In addition, Makhrov *et al.* (2005) suggested colonization from the Baltic basin into the White Sea.

Nowadays, human activities such as building of power plant dams, offshore overfishing, pollution, clearing of riverbeds for timber floating and badly designed stocking regimes have diminished the distribution of Atlantic salmon and many of the historical populations both in Europe and northern America have gone extinct or are severely depressed (Parrish *et al.* 1998). In fact, the White and Barents Seas in northwest Russia are among last areas in Europe where Atlantic salmon populations are still stable (Parrish *et al.* 1998). Other threats to the Atlantic salmon are fish diseases and parasites, such as *Gyrodactylus salaris*. Atlantic salmon from the Baltic Sea is tolerant to this parasite while in salmon from the Atlantic stock it reproduces unrestrained (e.g. Bakke *et al.* 1990, Peeler 2006) causing secondary infections in pre-smolt juvenile Atlantic salmon and commonly resulting in death. With devastating consequences, the parasite was introduced to Norway from the Baltic in the early 1970's and since then 45 Norwegian rivers have been infected (e.g. Kudersky *et al.* 2003, Peeler 2006). In 1992 the parasite was introduced to the river Keret in the White Sea causing 98% loss of juvenile production (Kudersky *et al.* 2003).

Microsatellite DNA and other molecular tools

Microsatellites, also known as short tandem repeats (Weber & Wong 1989), are stretches of DNA where identical 1-6 base pairs long motifs

are repeated one after the other, and depending on the length of the repeat unit they can also be called mono-, di- or trinucleotide repeats etc. A perfect microsatellite locus consists of an uninterrupted sequence of repeat units, such as $(TC)_{15}$, a compound locus of sequences of different repeat units, such as $(GAGT)_{10}(GT)_5(GAGT)_{14}$, and an interrupted microsatellite locus of a repeated sequence intermittent with non-repetitive sequence, such as $(CA)_8GTCCT(CA)_{16}$. Microsatellite alleles differ from one another by the number of repeat units present and one locus can have several tens of alleles (e.g. up to 48 alleles at a locus in King *et al.* (2005)).

The high polymorphism of microsatellite loci is explained by their mutation rate which has often been reported to range between $\sim 1 \times 10^{-3}$ and $\sim 1 \times 10^{-4}$ mutations per locus per generation (e.g. Weber & Wong 1993, Shimoda *et al.* 1999, Kayser *et al.* 2000, Xu *et al.* 2000). It is commonly believed that the size variation observed at microsatellite loci is generated through polymerase slippage and incorrect realignment during the replication of repetitive DNA (i.e. slipped strand mispairing, Levinson & Gutman 1987) leading to a new allele, the length of which differs by one or a few repeat units from the original allele. Indeed, mutations at microsatellite loci have often been noted to change the allele length by one repeat unit (e.g. Primmer *et al.* 1996, Brinkmann *et al.* 1998), although changes of multiple repeat units have also been reported (e.g. Primmer *et al.* 1996, Brinkmann *et al.* 1998, Fitzsimmons 1998, Primmer *et al.* 1998, Gardner *et al.* 2000). In agreement with these observations and the suggested mutation mechanism are the stepwise mutation model (SMM, Ohta & Kimura 1973), according to which mutations always change the length of a microsatellite in a stepwise fashion by adding or deleting one repeat unit, and the two phase model (TPM, Di Rienzo 1994), which states that most mutations are changes of one repeat unit but that sometimes changes of two or more units happen also.

Despite the complexity of microsatellite evolution and the open questions regarding their mutation model it is nevertheless clear that a considerable amount of evolutionary information is contained in different allele lengths (MacHugh 1996). This and the fact that microsatellites are evenly distributed in the genome (e.g. Litt & Luty 1989, Dietrich *et al.* 1996), codominant and selectively neutral (Litt & Luty 1989, Queller *et al.* 1993), highly polymorphic (Litt & Luty 1989, Tautz 1989, Weber & May 1989), and by using the polymerase chain reaction (PCR) easy to score (Queller *et al.* 1993) make them useful molecular tools for several types of studies. Indeed, they have been used, for example, in parentage testing, forensics, and population genetics (e.g. Queller *et al.* 1993) and, as one of the newest applications, to identify genomic regions affected by natural selection (e.g. Vasemägi & Primmer 2005).

However, there are some serious problems associated with microsatellites which need to be considered. The stepwise mutation model may lead to allele size homoplasy (Estoup *et al.* 1995), which implies that two alleles can be of the same size (identical in state) without being of the same origin (identical by descent) and which can cause overestimation of population relatedness if not taken into consideration. Another problem arises if a PCR primer-binding sequence has been changed by mutations that prevent the primer from annealing properly. As a result, the allele in question will not be PCR amplified (i.e. it is a null allele, Callen *et al.* 1993), which can lead to overestimation of homozygosity.

Prior to DNA based techniques, allozyme (i.e. enzyme isomer) electrophoresis was the dominant method utilized in population genetics studies. Mutations in protein coding DNA may lead to non-synonymous substitutions, which change the amino acid composition of an enzyme and potentially also its net charge and conformation, which in turn enables the identification of different alleles by means of electrophoresis. Technically allozyme electrophoresis is easy to implement and may be applied to any organism allowing screening of a large number of loci (Hansen *et al.* 2007).

The down side of the method is, however, that many loci are monomorphic or have only a low level of variation (Hansen *et al.* 2007).

The classical tool of phylogeographic studies is mitochondrial DNA (mtDNA) (e.g. Avise 1998). Unlike nuclear DNA, mtDNA is haploid, maternally inherited, primarily selectively neutral and mainly non-recombining, and has a relatively high mutation rate. It is the latter two characteristics in particular that make it relatively easy to reconstruct phylogenies of haplotypes and thus make mtDNA such a good tool for phylogeographical studies (e.g. Hansen *et al.* 2007). These special features of mtDNA, however, lead to a few limitations. Mitochondrial studies are commonly based on a small number of genes, and always on just one independently segregating locus, which may lead to erroneous population genetic inference (e.g. Pamilo & Nei 1988). In addition, some empirical studies have shown that selection may complicate mtDNA patterns (Hey 1997), and furthermore, due to the maternal inheritance, analysis of mtDNA may not give a correct picture of the genetic structure of populations if the migration rates of males and females differ.

Research objectives

Although the Atlantic salmon has been the subject of numerous studies, several aspects related to its glacial and post-glacial history have remained unresolved, and furthermore, the species' genetic structure in the easternmost distribution range has not been studied adequately. All of these are issues that should be clarified in order to have a sound genetic basis for the conservation and management of the species.

This thesis addresses several issues related to the population genetics of north European Atlantic salmon and specifically aims to:

1. Elucidate the phylogeography and population genetic structure of Atlantic salmon from northern Europe by combining information obtained with different types of molecular markers and with dense and extensive sampling coverage. Utilization of nuclear markers with differing mutation rates, such as microsatellites and allozymes, may help to provide a clearer picture of population relationships as the different marker types may resolve relationships over different evolutionary time scales. (Chapters **I** and **II**)

2. Assess the contribution of distinct northwest Russian Atlantic salmon populations to the region's overall genetic diversity with the aim of using the information for making management and conservation plans. In addition, to investigate the suitability of the microsatellite data obtained to serve as a baseline to assign individuals caught in offshore fisheries to their population of origin. (Chapter **II**)

3. Establish the importance of anadromous migration, population size, and population glacial history in determining the genetic diversity and divergence of Atlantic salmon populations. As both anadromous and non-anadromous life histories occur in the species, these offer an opportunity to thoroughly investigate the effects of anadromy and glacial history on genetic diversity and to assess their implications for conservation. (Chapter **III**)

4. Develop immune relevant microsatellite markers and, by employing them, investigate whether immune relevant genes exhibit different selection pressures compared to random loci in Atlantic salmon. As pathogen load in general is expected to be higher in the south than in the north due to temperature differences, and the studied populations differ in susceptibility to the deadly parasite *Gyrodactylus salaris*, immune relevant loci are expected to show signs of stronger selection than random loci. (Chapter **IV**)

MATERIAL AND METHODS

Atlantic salmon populations included in the study

Altogether 37 anadromous and 8 non-anadromous Atlantic salmon populations from northern Europe, with a special emphasis on northwest Russian populations, were sampled for the different studies constituting this thesis (Figure 2; for detailed maps, see Chapters **I-IV**). Human impact on northwest Russian Atlantic salmon has been relatively minor mainly due to the remoteness of the area. Hence, the White and Barents Seas are among the last regions in Europe where Atlantic salmon populations are still stable (Parrish *et al.* 1998), thus forming an important source of biodiversity for the entire European region. Salmon stocks from this area are also of immense economic and social importance for the local people in the form of fishing tourism.

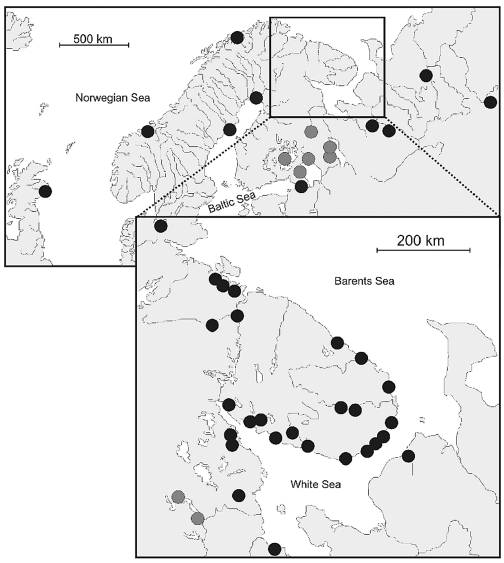

Figure 2. The Atlantic salmon populations sampled for this thesis (see Chapters **I-IV** for detailed maps and names of the populations). Anadromous populations are marked with black circles and non-anadromous with grey circles.

The main study methods

In all of the studies included in this thesis microsatellites have been the main molecular tool utilized, although in Chapter **I** allozymes and mtDNA were also employed. Details of the data gathering processes, such as PCR conditions and fragment analysis, can be found in the original papers (Chapters **I-IV**).

In every study the level of genetic diversity was estimated with the observed number of alleles (A), expected and observed heterozygosities (H_e and H_o, respectively), and allelic richness (A_r) while in the estimation of genetic divergence the θ estimator of Wright's F_{ST} (Weir & Cockerham 1984) and, in Chapter **I**, also its stepwise mutation model based analogue ρ_{ST} (Rousset 1996), were used. Additionally, the data were also checked for discrepancies from Hardy-Weinberg equilibrium (HWE) and genotypic linkage equilibrium (LE).

In Chapters **I** and **II** either phylogenetic trees based on Cavalli-Sforza and Edwards (1967) chord distance (D_{CE}) or principal component analysis (PCA) were employed to study the genetic relationships between north European Atlantic salmon populations. The trees in Chapter **I** were built using allozyme data and microsatellite data alone and by combining the two data sets.

To elucidate if the north European populations analysed in Chapter **I** originate from a single or several refugia, an allele size permutation test was utilized (Hardy *et al.* 2003). The divergence estimate calculated after allele size permutation, pR_{ST}, should equal F_{ST} and thus the method can be interpreted as testing whether $F_{ST} = R_{ST}$, where R_{ST} is an SMM-based analogue of F_{ST}. When the contribution of stepwise-like mutations to genetic differentiation is negligible compared to genetic drift and migration, F_{ST} and R_{ST} should give similar estimates of genetic divergence, while if stepwise-like mutations have

contributed significantly to divergence $R_{ST} > F_{ST}$. Taking into account the generation time of salmon and the approximate microsatellite mutation rate, stepwise-like mutations should not have contributed significantly to the divergence of populations colonized from the same glacial refugium, but should have contributed to the divergence of populations colonized from different refugia (Estoup & Angers 1998). Hence, allele size permutation provides a method for testing whether a particular region was colonized from one or several refugia.

In Chapter **II**, a Mantel test was utilized to examine whether the geographical distance between the populations and their genetic divergence as estimated with $F_{ST}/(1-F_{ST})$ were associated (i.e. if there was any sign of isolation by distance, IBD). To study if the strength of IBD was constant over different geographical ranges, the slope of IBD regression was calculated by including population pairs of increasing geographical distance one by one into the analysis. For example, all the populations separated by 200 km or less were included into the calculation of the 200 km regression slope and by adding the next pair into the analysis a slope at 205 km was obtained.

The suitability of the data obtained in Chapter **II** to serve as a reference for individual assignment in offshore fisheries was examined using a self-assignment method. For likelihood estimation the direct approach using the Bayesian method of Rannala & Mountain (1997) was chosen.

In order to assess the conservation value of the Russian Atlantic salmon populations, the percentage of total genetic diversity (GD%, Crozier *et al.* 1999) retained in various groups comprised of the populations was estimated. In addition, each population's contribution to the total allelic richness among all the Russian populations was estimated (C^r_T) and further partitioned into two components: population's contribution to total allelic richness due to 1) its own diversity (C^r_S) and 2) its divergence from the other populations (C^r_D, Petit *et al.* 1998).

In Chapter **III** levels of genetic diversity and population divergence between anadromous and non-anadromous populations were compared. As loci with high within-population heterozygosity (H_e), such as microsatellites, may give underestimates of population divergence when measured with F-statistics estimates of G'_{ST}, a standardized measure of genetic differentiation independent of the degree of within-population genetic variation, were also calculated (Hedrick 1999, 2005).

To examine the effect of stock size and phylogeographic origin on genetic diversity, the populations were categorised into seven groups according to the estimated number of adults ascending to a particular river each year (for details, see Table 2 in Chapter **III**). A general linear model (GLM) was utilized to study the relationship between A_r, H_e, or H_o and stock size and phylogeographic origin in anadromous and non-anadromous populations. For interpretation of the significant interaction between phylogeographic origin and life history type, least square means (LSMEAN) of A_r, H_e, and H_o were calculated for anadromous and non-anadromous populations from the Baltic Sea basin and the basins of the Atlantic Ocean, White Sea and Barents Sea.

Populations included in the study described in Chapter **IV** were chosen so that they expressed different susceptibility to *Gyrodactylus salaris* and originated from different environments (freshwater Lake Ladoga, brackish Baltic Sea, marine Atlantic Ocean) and different latitudes (from 61°N to 71°N). As the pathogen load in general was expected to be higher in the south than in the north due to temperature differences, and as the studied populations differed in *G. salaris* susceptibility, the immune relevant loci included in the study were expected to show signs of stronger selection than the random loci, especially when comparing the Lake Ladoga and Baltic Sea populations to the Atlantic populations.

For the development of immune relevant microsatellite markers needed in this type of study, the literature was searched for Atlantic

salmon expressed sequence tags (EST's) from genes differentially regulated upon exposure to various pathogens, including the bacterium *Aeromonas salmonicida* (Tsoi *et al.* 2004, Martin *et al.* 2006), the parasite *Gyrodactylus salaris* (Matejusová *et al.* 2006, Collins *et al.* 2007), and saprolegniaceae water moulds (Roberge *et al.* 2007). These EST's were screened for di-, tri-, and tetranucleotide microsatellites, and the microsatellite-containing EST's were further screened for duplicates, which were discarded. Data obtained with these immune relevant microsatellites was complemented with data from presumably neutral microsatellite loci obtained in the study presented in Chapter **I** and with data from gene-linked loci used in Vasemägi *et al.* (2005).

In the identification of loci under selection two methods based on the idea of genetic hitchhiking introduced by Maynard-Smith & Haigh (1974) were utilized. The first method is based on the idea that loci under selection usually show unusually high or low genetic differentiation, which can be quantified using F_{ST}. In the method of Beaumont and Nichols (1996), the F_{ST} of each locus is plotted against its heterozygosity and, in order to detect outliers, coalescent simulations are performed to obtain a distribution of F_{ST} as a function of heterozygosity and to determine quantiles that cover, for example, 95% of the data points. Putatively selected loci thus fall outside the interval determined by the quantiles.

The other method utilized is based on the notion that microsatellite loci that are linked to a selected region of the genome are expected to show reduction in variability (e.g. Slatkin 1995, Schlötterer *et al.* 1997). Simulations have indicated that ln-transformed ratios of expected heterozygosity in two populations (*lnRH*) follow neutral distribution (Schlötterer 2002, Kauer *et al.* 2003) and thus loci that have undergone a selective sweep are located at the tails of the *lnRH* distribution.

In addition to traditional hitchhiking methods a marker group approach was also employed. The markers were grouped into immune relevant and non-immune relevant EST-linked loci and into anonymous microsatellites that are *a priori* assumed to experience different selection pressures. This method differs from the common neutrality tests as it does not concentrate on the identification of single loci that deviate from neutral expectations and is therefore expected to have several advantages over traditional hitchhiking mapping studies, such as lower rates of false positives.

Also, a landscape genomics method, which combines environmental information with molecular data (Joost *et al.* 2007), was employed to investigate whether immune relevant markers displayed significant correlations to biotic and abiotic environmental conditions more frequently than other loci. This spatial analysis method (SAM) tests for association between allele frequencies at marker loci and environmental variables by building logistic regression models for all possible marker-environmental variable pairs and using the likelihood ratio (*G*) and Wald tests to determine the significance of the models. Here, SAM was utilized to test: 1.) if the three locus groups differed in terms of selection pressure they have experienced, 2.) if gene-linked loci (i.e. immune relevant loci plus loci linked to other EST's) have been under different selection pressures compared to anonymous micro-satellites, and 3.) if immune relevant genes have been under different selection pressures compared to random loci (i.e. loci linked to other EST's plus anonymous microsatellite loci).

RESULTS AND DISCUSSION

Phylogeography and population genetic structure of north European Atlantic salmon

Microsatellite data identified three clusters among north European Atlantic salmon populations while phylogeographic resolution with allozyme data alone was relatively limited as only a small number of nodes on the phylogenetic tree were supported by bootstrap values > 50 % (Figures 3a, b, Chapter **I**). The analysis of the combined microsatellite-allozyme data set identified the same three groups as the microsatellite data alone but with higher bootstrap support for some key nodes making the within-cluster population relationships clearer (Figure 3c, Chapter **I**). All populations from the Baltic Sea basin (Lakes Saimaa, Onega and Ladoga and rivers Tornionjoki, Vindelälven and Neva) clustered together with bootstrap support of 63%, while the non-anadromous populations from the White Sea basin (Lakes Luzhma, Pisto and Kamennoe) formed another group supported with a bootstrap value of 73%. The remaining populations included all those from the White and Barents Seas, and the Atlantic Ocean. Support for this cluster as a separate group was not high, and highly supported nodes within the cluster tended to be for populations situated geographically close to each other.

The global multilocus estimates of F_{ST} and R_{ST} between the above groups were 0.103 and 0.118, respectively (Figure 4, Chapter **I**). The observed multilocus R_{ST} lay above the upper limit of the 95% confidence interval of the null distribution of the permuted pR_{ST} and was statistically significant (P = 0.016, Figure 4, Chapter **I**) indicating that stepwise-like mutations have contributed to genetic divergence and therefore, postglacial colonization of northern Europe from more than one glacial refugium is statistically supported. The result remained statistically significant (P = 0.017) when the three non-anadromous populations from the White Sea basin (Luzhma, Pisto, Kamennoe) were excluded from the analysis. However, for other group pairings or at the within group level, there was no indication of stepwise-like mutations having contributed to genetic divergence (Figure 4, Chapter **I**).

Colonization of the Baltic Sea and Lakes Ladoga and Onega from a single glacial refugium, most likely the Baltic Ice Lake, is supported by the grouping of all Baltic Basin populations together with moderately high bootstrap support (Chapter **I**). This post-glacial colonization scenario is in line with that proposed by Nilsson *et al.* (2001), based on mtDNA, but is not concordant with colonization scenarios where a significant contribution from a western refugium (North Sea) has been proposed for part (Koljonen *et al.* 1999, Säisä *et al.* 2005) or all (Verspoor *et al.* 1999) of the Baltic region. Given the highly supported separation of all Baltic Sea basin populations from the Scottish River Dee population (91% bootstrap support, Figure 3c, Chapter **I**), any significant contribution of North Sea stocks to the recolonization of the Baltic Sea seems unlikely.

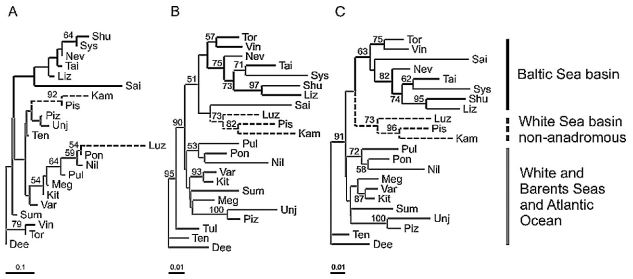

Figure 3. Neighbour-joining phylograms based on a) allozyme data, b) microsatellite data, and c) the combination of both marker types based on D_{CE} distances. The numbers indicate percent bootstrap support for each node over 2 000 replications. Only values over 50% are shown. (Chapter **I**)

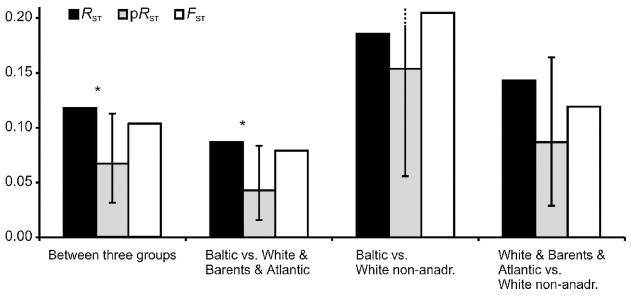

Figure 4. Results of the allele size permutation test including global estimates of R_{ST}, pR_{ST}, and F_{ST}. Populations were grouped according to the three clusters revealed by the phylogenetic tree (see Figure 3c for details). The 95% confidence intervals are given for pR_{ST}, and cases where global R_{ST} was significantly larger ($0.05 > P > 0.01$) than the permuted null distribution of pR_{ST} are indicated with an asterix. (Chapter **I**)

Both the population phylogram (Figure 3c) and the allele size permutation test (Figure 4) indicate that the populations from the White, Barents, and Atlantic basins most likely originate from different glacial refugia than the Baltic populations (Chapter **I**). This refugium, also suggested by Kazakov & Titov (1991) and Asplund *et al.* (2004), could potentially have been the eastern Barents Sea, which is likely to have been free of ice during the Late Weichselian period (Svendsen *et al.* 2004) thus enabling survival of refugial Atlantic salmon populations. In addition, the consistent rare occurrence of a western Atlantic allozyme allele **80* at the locus *ESTD-2** in several populations from the White and Barents Sea basins, including two non-anadromous populations, suggest immigration also from the western Atlantic Ocean, although neither the phylogenetic tree nor the allele size permutation test lend strong support to this theory (Chapter **I**).

With more thorough sampling of Atlantic salmon populations from the northwest Russian region the relationships between populations and hence the post-glacial colonization of the area became clearer (Chapter **II**). The principal component analysis divided the populations into four clusters corresponding well to the geographical sampling regions (Figure 5, Chapter **II**). The populations from the northeast Atlantic Ocean and western Barents Sea formed a clearly separate group. Likewise most the populations from eastern and southern Kola Peninsula grouped tightly together, as did the majority of the populations from western coast of the White Sea. The fourth group was formed by populations from the eastern White and Barents Sea with the exception of the river Megra population, which grouped together with the geographically closer populations from the eastern Kola Peninsula.

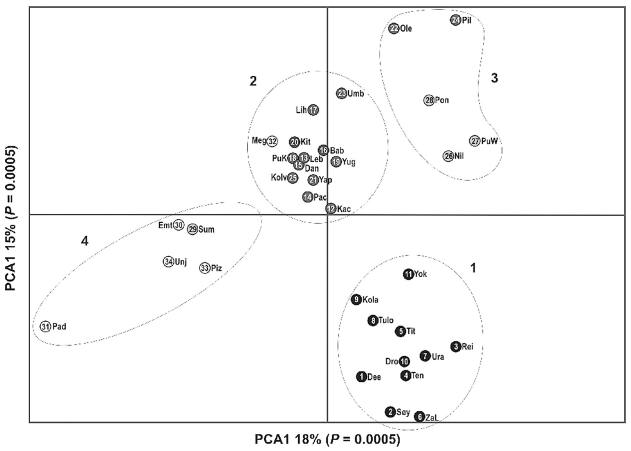

Figure 5. Result of the principal component analysis. Black circles depict the populations from the northeast Atlantic Ocean and western Barents Sea, dark grey circles the populations from the eastern and southern Kola Peninsula, light grey circles the populations from the western White Sea, and white circles the populations from the eastern White and Barents Seas. (Chapter **II**)

This clear separation of the populations into northeastern Atlantic and western Barents Sea vs. White and eastern Barents Sea populations implies that these regions were colonized from different refugia, and indicates colonization of the western Barents Sea from the west rather than from the east (Chapter **II**), which is also supported by findings of earlier studies (e.g. Asplund *et al.* 2004, Makhrov *et al.* 2005, Chapter **I**). This is further corroborated by allele frequencies at the locus *Ssa197*, in which large alleles (293-329 bp) were found among nine of the 11 populations forming the northeastern Atlantic and western Barents Sea PCA group but not among populations further to the east. The White and eastern Barents Seas in turn were most likely colonized from the east as suggested earlier (Asplund *et al.* 2004, Chapter **I**) and supported by the current observation of distinctiveness of the populations from the region. Moreover, the absence of large alleles at *Ssa197* in all of the White Sea and eastern Barents Sea populations suggests that there has been no immigration from the eastern Atlantic Ocean into the area.

Although significant isolation by distance was found when all the studied populations were included in the analysis (Mantel's r_{XY} = 0.39, P = 0.008; Figure 6, Chapter **II**), group-wise significant IBD was observed only among populations forming PCA groups 1 and 2 while in the other two PCA clusters no sign of IBD was detected (Figures 5, 6, Chapter **II**). This indicates that the spatial scale at which IBD analyses are carried out matters, as has also been noted in earlier studies (e.g. Slatkin 1993, Rousset 1997, Ehrich & Stenseth 2001, Castric & Bernatchez 2003).

When IBD was examined on various geographical scales a great deal of variation in the slope of the regression was found if only populations separated by less than 100 km were considered (Figure 7, Chapter **II**). When population pairs separated by more than 100 km were included in the analysis the IBD regression slope decreased as a function of increasing geographical distance until approximately 160 km after which the slope plateaued out (Figure 7, Chapter **II**). A similar pattern of IBD fading has been detected in other studies and has been given various explanations (Ehrich & Stenseth 2001, Castric & Bernatchez 2003). In this study, it is unlikely that the fading of IBD is due to F_{ST} reaching its upper limit, which has been suggested to be one of the causes for such a finding (Castric & Bernatchez 2003), as the regression slopes were also calculated using estimates of G'_{ST} and again a decreasing and finally leveling off pattern of IBD regression slopes was obtained. Most likely not enough time has passed since the area was colonized and hence the populations may exhibit a non-equilibrium situation. It may be that Atlantic salmon from the White and Barents Seas are still going through a transitory phase towards equilibrium and IBD cannot be detected at large geographical distances.

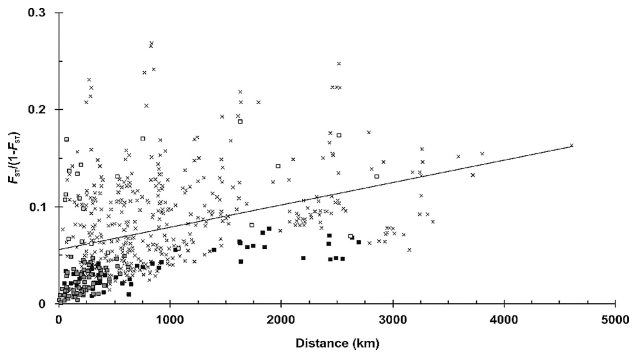

Figure 6. The relationship between the geographical (km) and genetic distance ($F_{ST}/(1-F_{ST})$) of north European Atlantic salmon populations. Squares represent interpopulation distances between populations among each PCA group (black, group 1; dark grey, group 2; light grey, group 3; white, group 4) while interpopulation distances between the four groups are represented with crosses. The line depicts the regression slope of all interpopulation comparisons (Mantel's r_{XY} = 0.39, P = 0.008). (Chapter **II**)

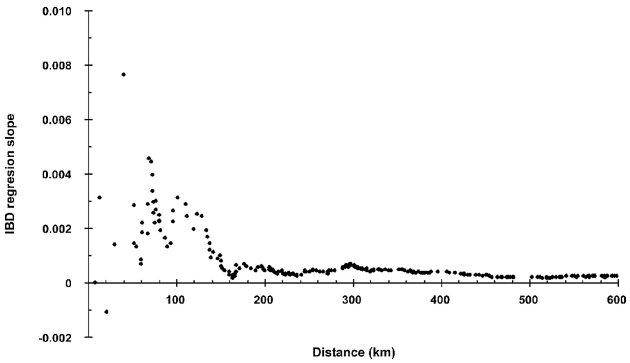

Figure 7. The relationship between the slope of isolation by distance regression and the geographic distance between the populations. Each slope was calculated by adding population pairs at increasing geographical distance into the analysis. For example, all the populations separated by 200 km or less were included into the calculation of the 200 km regression slope and by adding the next pair into the analysis a slope at 205 km was obtained. (Chapter **II**)

Suitability of microsatellite data to serve as reference for individual assignment

Despite the clear boundaries observed between populations from the White and Barents Seas (Figure 5, Chapter **II**), a hierarchical analysis of molecular variance indicated that only 2.6% or 3.1% of genetic variation could be explained by differences between the groups depending on whether the populations were grouped according to their geographical origin or the result of the PCA, respectively (Chapter **II**). This low level of between group variation is probably also reflected in the results of the assignment test as assignment success depends on the level of population differentiation (Cornuet *et al.* 1999).

The self-assignment efficiency varied considerably between populations (from 15.2% to 100%, median 69.0%, Table 1, Chapter **II**), while at a group level assignment success was relatively high (87.3% to 95.7%, Table 1, Chapter **II**). A common trend seemed to be that misassigned individuals were often assigned to a nearby population or to a population for which divergence from the source population was low. In a few cases the low success was most likely due to the numbers of sampled individuals being too low. Thus, to achieve a higher assignment success it is advisable to increase the number of sampled individuals in populations with low sample size and more importantly, increase the number of loci genotyped, as simulations have shown that when F_{ST} is between 0.05 and 0.1 the best result is obtained with 20-30 loci (Cornuet *et al.* 1999). Here, pairwise estimates of population divergence varied from 0.001 to 0.211 with a median of 0.064 (Chapter **II**) implying that population divergence was too low for reliable individual assignment at a population level. However, the present data as such may be used to fairly reliably assign fish to the correct group of origin.

Conservation value of northwest Russian Atlantic salmon populations

Considering diversity contained in different groups of Russian Atlantic salmon, with a view to using this information for prioritizing conservation efforts, populations from the eastern and southern Kola Peninsula (i.e. PCA group 2) retained the highest proportion of genetic diversity, 34.4%, while populations from the western White Sea (PCA group 3) retained 32.7% and populations from the eastern White and Barents Seas (PCA group 4) 31.9% of the overall genetic diversity (Figure 8, Chapter **II**). The proportion of genetic diversity maintained among the Russian Atlantic salmon populations from western Barents Sea (PCA group 1) was approximately 30% lower than among PCA groups 2, 3, and 4 (GD% = 23.5%; Figure 8, Chapter **II**). Overall, the pairs maintaining the greatest share of diversity from each PCA group together retained 53.3% of the total genetic diversity (Figure 8, Chapter **II**). This clearly indicates that conservation of populations from each of the observed

Table 1. Number of individuals assigned to different populations based on their multi-locus microsatellite genotype. Assignments within a PCA group are shaded in grey. (Chapter **II**)

Origin	Assigned to																																Total N	Correctly assigned (%)
	PCA group 1											PCA group 2													PCA group 3					PCA group 4						
	Dee	Søy	Rei	Ten	Tit	ZaL	Ura	Tul	Kola	Dro	Yok	Kac	Leb	Pac	Dan	Bab	Lih	PuK	Yug	Kit	Yap	Umb	Kolv	Meg	Ole	Pil	Nil	PuW	Pon	Sum	Emt	Pad	Piz	Unj		
Dee	**26**	2		3									1																				32	81.3
Søy	2	**21**	1		1				1		1																						27	77.8
Rei		1	**4**	2		1	3					1						1													1			14	28.6
Ten				**19**	3	1	5	3	2		2	2	1	1					3	1														43	44.2
Tit			2	4	**19**	3	1	1	1	1		1	3			2			1											1				40	47.5
ZaL				3	3	**28**	3	1	1											1														40	70.0
Ura			1	3	2	1	**28**	1	1	1	1						2																	41	68.3
Tul				1	1		3	**31**	8			1	2			1	1		1											1				51	60.8
Kola	1			1		1	2	1	**26**			1	1		1	2	1	2		1											1			42	61.9
Dro				1			1		1	**34**	1		1																1					40	85.0
Yok				1	1	1	1		2	1	**23**	4	1		1	1	1				2													40	57.5
Kac				2	1			3		1	4	**11**	2	2		3	4	1	5		1	1												41	26.8
Leb									1			2	**21**	3		5	4	2	3	3	1		1		1									47	44.7
Pac					1			1				4	8	**19**			2		2			1												38	50.0
Dan			1			1			1				2		**20**	5	1	4	3	2	2	1									1			44	45.5
Bab					1				1			3	3		4	**7**	3	7	6	4	3	2							1		1			46	15.2
Lih									1				3	1		5	**16**	9	6		5	1												47	34.0
PuK												1	4	2	2	10	7	**10**	2	1	3		1	1							2			47	21.3
Yug				2					1			3	1		3	4	5	3	**26**	4	2	1			1									56	46.4
Kit							1						3		3	2	2	4	4	**12**	3	2			1									37	32.4
Yap					1		1					2	3		1	5	4	3	1	3	**15**			1											40	37.5
Umb									2			2				3	1	1	1	1	1	**33**							2					47	70.2
Kolv													1			1		1	1			1	**36**											41	87.8
Meg															1	4		1	3		1			**30**											43	69.8
Ole																	1		1	1					**31**										34	91.2
Pil																						1				**32**			2						35	91.4
Nil																											**42**								42	100
PuW																												**43**							43	100
Pon												1				1	2			1									**45**							50	90.0
Sum					1																									**40**					41	97.6
Emt					1				2								1														**37**	1			42	88.1
Pad																															1	**47**			48	97.9
Piz																			1														**21**		22	95.5
Unj														1																			2	**8**	11	72.7
PCA gr. 1											**358**													47					1					4	410	87.3
PCA gr. 2											28													**533**					6					7	574	92.9
PCA gr. 3											0													9					**195**					0	204	95.6
PCA gr. 4											4													3					0					**157**	164	95.7

clusters is recommended as inclusion of just two populations from each group would retain considerably more of the entire region's genetic diversity than conserving all of the populations of a single group (Figure 8, Chapter **II**).

Populations contributing most to total allelic richness were partly the same as the populations retaining most genetic diversity: considering the three populations identified to retain the most genetic diversity within each of the four PCA groups (i.e. 12 populations in total) six out of these 12 populations were the same with both analysis methods (Figures 8 and 9, Chapter **II**). Apart from a few exceptions, genetic diversity was the main component contributing to allelic richness among PCA groups 1 and 2 while among PCA groups 3 and 4, genetic divergence was more influential (Figure 9, Chapter **II**). This emphasizes the importance of considering the C^{r}_{S} and C^{r}_{D} components and not just the total contribution of each population when making management plans based on this analysis. More importantly, it is advisable to base management plans on several prioritizing methods as different methods pinpoint different populations worthy of conservation.

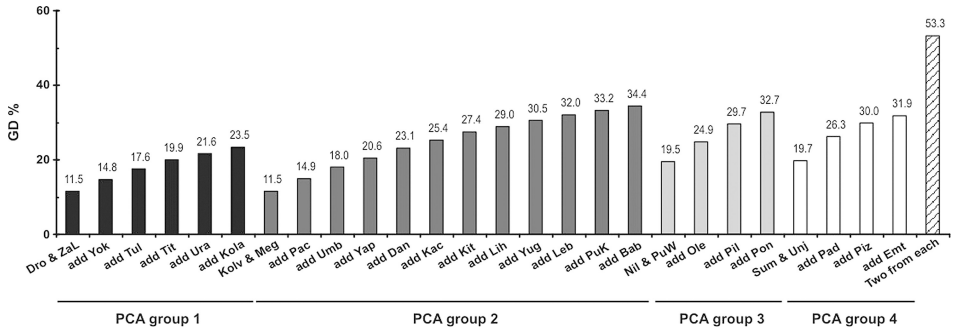

Figure 8. Proportion of genetic diversity (GD%) maintained in various groupings of the studied populations. For each PCA cluster, the pair retaining the greatest share of genetic diversity, and the increase in diversity with successive additions of populations are indicated. For example, in PCA group 1 Drozdovka, Zapadnaya Litsa, Yokanga and Tuloma together retain 17.6% of the total genetic diversity while adding Titovka to the group increased GD% to 19.9%. The last column depicts the proportion of genetic diversity preserved in the pairs with highest GD% from each PCA cluster together. (Chapter **II**)

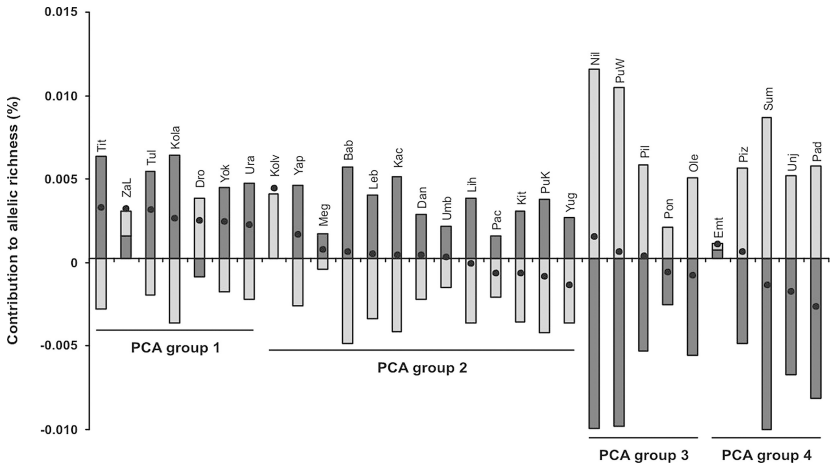

Figure 9. Each populations' contribution to total allelic richness. The total contribution (C^{r}_{T}) of a population is denoted with a filled circle and the two components of variation with bars (dark grey bar, C^{r}_{S}; light grey bar, C^{r}_{D}). Populations within a PCA group are ordered according to C^{r}_{T}. (Chapter **II**)

The effect of migratory behaviour on genetic diversity and population divergence

Overall, the level of genetic diversity of the anadromous populations was significantly higher (A_r = 5.7, H_o = 0.63, H_e = 0.63) than that of the non-anadromous populations (A_r = 3.7, H_o = 0.48, H_e = 0.46; all P values ≤ 0.001; Figure 10, Chapter **III**). In addition, the level of genetic divergence among the freshwater populations (F_{ST} = 0.31) was significantly greater than among the anadromous populations (F_{ST} = 0.12; all P values ≤ 0.001; Figure 10, Chapter **III**). The G'_{ST} estimate, which corrects for differences in variability between loci, was 0.34 for anadromous and 0.62 for freshwater populations. These results are congruent with those of DeWoody & Avise (2000) who found a significant difference in the number of microsatellite alleles and H_e in an interspecific comparison of anadromous (A = 11.3, H_e = 0.68) and freshwater fishes (A = 7.5, H_e = 0.46).

The linear regression indicates that the result obtained here holds even when accounting for differences in stock sizes: within a given population census size class, anadromous populations almost always had a higher level of genetic diversity than non-anadromous populations of the same size class (Figure 11, Table 2, Chapter **III**). This positive correlation between estimated census stock size and genetic diversity can be explained by basic population genetic theory, whereby the effect of genetic drift is expected to be higher in populations with small N_e than in populations with high N_e (e.g. Frankham *et al.* 2002).

The census size of the populations was not, however, the only factor found to affect the genetic diversity characteristics of the populations: the interaction between life history strategy and phylogeographic origin also contributed significantly to the level of genetic diversity observed in a population (Table 2, Chapter **III**). Indeed, anadromous populations from the Atlantic Ocean, White Sea and Barents Sea were found to be more variable than non-anadromous populations from

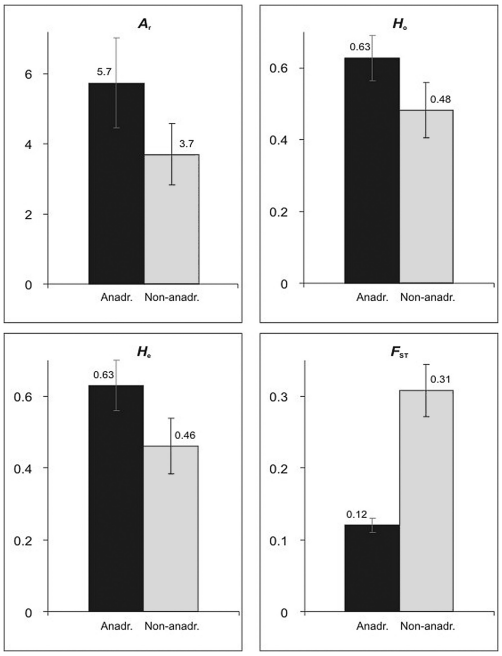

Figure 10. A comparison of genetic diversity indices and the level of genetic divergence between anadromous and non-anadromous Atlantic salmon populations. Error bars indicate standard deviation. A_r - allelic richness, H_o - observed heterozygosity, H_e - expected heterozygosity, Anadr. - anadromous populations, Non-anadr. – non-anadromous populations. (Chapter **III**)

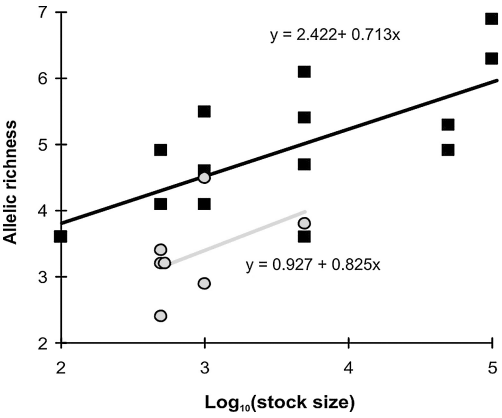

Figure 11. Linear regression depicting the association between genetic diversity (allelic richness) and stock size for populations with alternative migration behaviours. Anadromous populations are marked with black squares and non-anadromous with grey circles. (Chapter **III**)

the same area, while in the Baltic Sea basin the result was different as no differences in genetic diversity characteristics were found between anadromous and non-anadromous populations (Table 3, Chapter **III**). Among anadromous Atlantic salmon, more diversity was observed in populations from the Atlantic Ocean, and White and Barents Seas than in populations from the Baltic Sea, whereas heterozygosity levels of the Baltic non-anadromous populations were higher than of the non-anadromous populations from the White Sea (Table 3, Chapter **III**). All these population groups differ in the number of refugia they have been colonized from and in the time they have been isolated from other populations since colonization, which demonstrates how important it is to take the glacial history of the studied area and population phylogeography into consideration.

Table 2. The analysis of variance table for stock size, life history type (anadromous vs. non-anadromous), and basin of origin (Baltic Sea basin vs. Atlantic Ocean, White and Barents Sea basins) in explaining the variability of allelic richness (A_r), and expected (H_e) and observed heterozygosity (H_o). (Chapter **III**)

Source of variation	A_r (R^2 = 79%)				H_e (R^2 = 83%)				H_o (R^2 = 76%)			
	Sum of squares	df	*F*	*P* value	Sum of squares	df	*F*	*P* value	Sum of squares	df	*F*	*P* value
Log_{10} (stock size)	6.738	1	18.38	***	0.021	1	9.03	**	0.004	1	1.42	NS
Life history type	2.815	1	7.68	*	0.038	1	16.34	***	0.037	1	14.25	**
Basin of origin	0.064	1	0.18	NS	0.000	1	0.06	NS	0.001	1	0.32	NS
Interaction between life history type and basin of origin	2.889	1	7.88	*	0.045	1	19.57	***	0.042	1	16.20	***

* $0.01 < P \leq 0.05$.

** $0.001 < P \leq 0.01$.

*** $P \leq 0.001$.

NS, not significant

Table 3. A comparison of the least squares means (LSMEAN) of allelic richness (A_r), and expected (H_e) and observed heterozygosity (H_o) between anadromous and freshwater populations (in rows) and between populations from the Baltic Sea basin and basins of the Atlantic Ocean, White Sea and Barents Sea (in columns). Anadr. – anadromous populations. Non-an. – non-anadromous populations. (Chapter **III**)

Phylogeographic origin	A_r LSMEAN		*P* value	H_e LSMEAN		*P* value	H_o LSMEAN		*P* value
	Anadr.	Non-an.		Anadr.	Non-an.		Anadr.	Non-an.	
Atlantic Ocean, White and Barents Seas	5.041	3.268	***	0.640	0.427	***	0.644	0.437	***
Baltic	4.059	3.993	NS	0.540	0.540	NS	0.556	0.554	NS
P value	*	NS		**	**		*	**	

* $0.01 < P \leq 0.05$.

** $0.001 < P \leq 0.01$.

*** $P \leq 0.001$.

NS, not significant

Signatures of selection in immune relevant and random loci

In Chapter **IV**, 23 immune relevant marker loci, 62 markers linked to other genes and 14 anonymous microsatellites were screened to search for footprints of selection and to study if different groups of genetic markers showed different patterns of genetic differentiation. In pairwise comparisons of different environments (freshwater Lake Ladoga, brackish Baltic Sea, and marine Atlantic Ocean) the traditional hitchhiking methods found altogether 23 loci that deviated from neutral expectations (Figures 12 and 13, Chapter **IV**). Only one locus was supported by both of the employed methods and five additional loci by one test but in two different pairwise comparisons. The remaining outlier loci were identified as significant only once. Given that the observed number of outlier loci identified by the hitchhiking methods was lower than the expected number of false positives (30 out of the total 594 significance test performed) and the lack of support for candidate status of the loci by multiple tests, none of the loci identified here seem to be strong candidates for selection.

However, the marker group approach implies, that the immune relevant loci and anonymous microsatellites might have indeed experienced different selection pressures compared to random loci as they have more alleles correlated with latitude than what is expected if selection has affected all loci in a similar fashion (Table 4, Chapter **IV**). Conversely, in a comparison of the levels of genetic differentiation none of the marker groups stood out as more or less divergent than the others (Figure 14, Chapter **IV**). One possible explanation for why more pronounced differences between the immune relevant and random loci were not found is the fact that for many EST's no homologs were found in BLAST searches, possibly leading to misclassification of loci (i.e. grouping of non-immune relevant loci into the random locus group or vice versa) which makes the groups more similar in terms

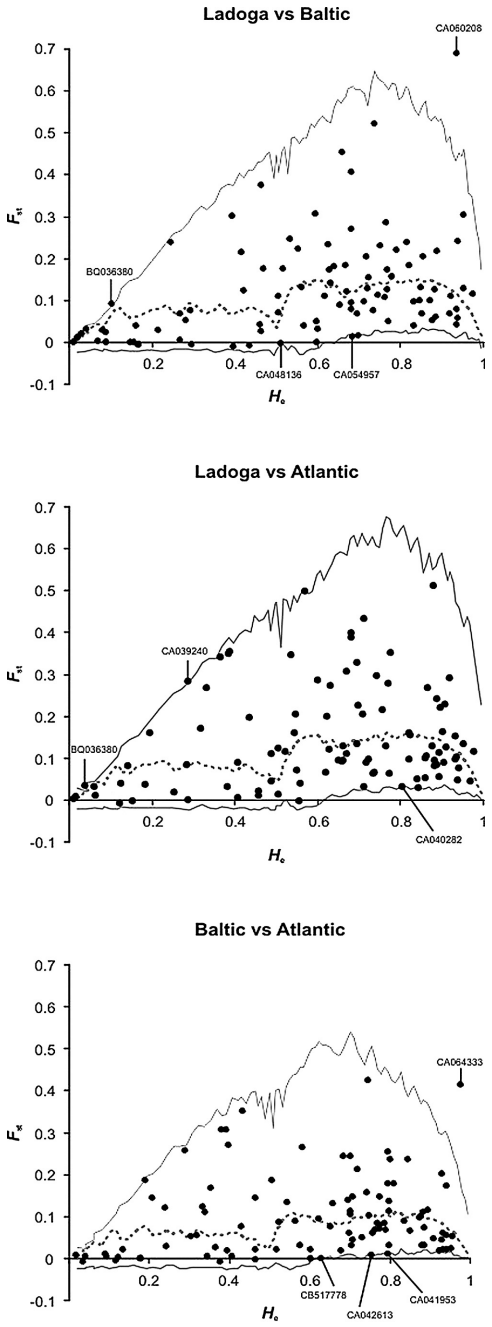

Figure 12. The F_{ST} estimate for each locus plotted against its heterozygosity in different pairwise comparisons (freshwater vs. brackish, freshwater vs. marine, and brackish vs. marine). The solid lines represent the 95% confidence intervals and the dotted line the median. Significant outlier loci are indicated. (Chapter **IV**)

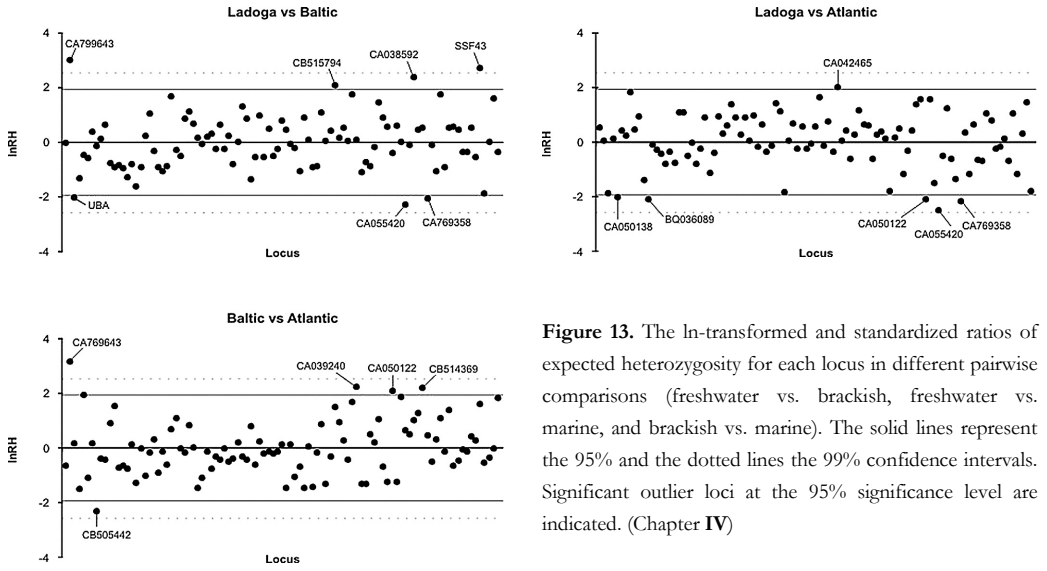

Figure 13. The ln-transformed and standardized ratios of expected heterozygosity for each locus in different pairwise comparisons (freshwater vs. brackish, freshwater vs. marine, and brackish vs. marine). The solid lines represent the 95% and the dotted lines the 99% confidence intervals. Significant outlier loci at the 95% significance level are indicated. (Chapter **IV**)

of selection pressure than what they are in reality. Furthermore, anonymous microsatellites are not always selectively neutral but may be linked to functional loci. In fact, three of the allegedly neutral microsatellite loci utilized in Chapter **IV** have been found to be associated with either innate (*Ssa85*) or acquired *G. salaris* resistance (*Ssa171*, *SSOSL311*; Gilbey *et al.* 2006).

Yet another complication in searching for immune relevant marker loci from genes that are differentially expressed after a pathogen challenge is that the gene itself may not have been under selection. If selection was targeted to a distant regulatory element or factor instead of the gene itself it is likely that it will not be identified by markers linked to the gene as hitchhiking methods can detect signs of selection only from markers linked to the selected genomic region.

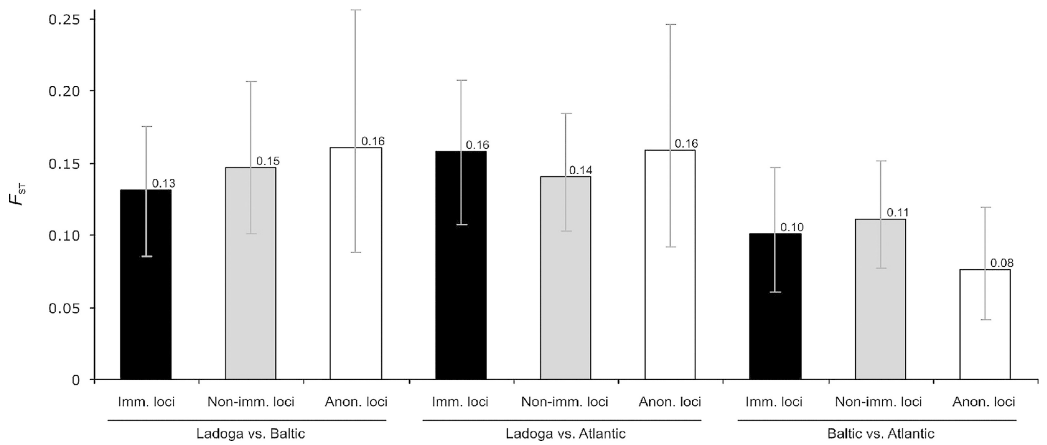

Figure 14. Global F_{ST} estimates for immune relevant loci, loci linked to other genes and anonymous microsatellites between freshwater and brackish, freshwater and marine, and brackish and marine populations. None of the comparisons revealed statistically significant differences. (Chapter **IV**)

Table 4. Number of alleles significantly correlated with mortality caused by *G. salaris*, basin salinity, and sampling location latitude and longitude for different groupings of the studied loci on the Bonferroni corrected significance level corresponding to $P = 0.05$. *P* values indicate the probability that the observed allele numbers follow the expectation accoding to the χ^2 test. (Chapter **IV**)

Marker type	Observed (obs.) and expected (exp.) number of alleles significantly correlated with an environmental variable								Total number of alleles
	G. salaris mortality		Salinity		Latitude		Longitude		
	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	Obs.	Exp.	
Immune relevant loci	10	8.8	12	9.4	19	13.1	11	8.2	167
Non-immune relevant loci	32	37.4	32	40.0	43	55.7	29	34.8	712
Anonymous loci	15	10.9	17	11.6	23	16.2	13	10.1	207
χ^2 *P* value	0.26		0.08		0.01		0.23		
EST's	42	46.1	44	49.4	62	68.8	40	42.9	879
Anonymous loci	15	10.9	17	11.6	23	16.2	13	10.1	207
χ^2 *P* value	0.15		0.07		0.05		0.30		
Immune relevant loci	10	8.8	12	9.4	19	13.1	11	8.2	167
Random loci	47	48.2	49	51.6	66	71.9	42	44.9	919
χ^2 *P* value	0.64		0.40		0.06		0.27		

CONCLUSIONS AND FUTURE DIRECTIONS

The main aim of this thesis was to elucidate the post-glacial history and population genetic structure of north European Atlantic salmon, both of which are aspects of great importance for the management and conservation of the species. In particular the northwest Russian populations, which are among the few last European populations still in their natural state, have not been studied in adequate detail before. Previous studies have employed markers with low level of polymorphism, whose ability to detect population structure at small geographical scales is weak (i.e. allozymes, Kazakov & Titov 1991, Makhrov *et al.* 2005), or markers which do not resolve the genetic structure of the whole species but of just one sex (i.e. mtDNA, Asplund *et al.* 2004). Thus, within the scope of this thesis highly polymorphic nuclear markers are employed for the first time to study the northwest Russian Atlantic salmon populations enabling firm population genetic conclusions to be drawn.

Chapters **I** and **II** significantly added to the current knowledge of post-glacial colonization of north European Atlantic salmon. An especially important discovery was the division of Atlantic salmon from the White and Barents Seas into four separate clusters, which has not been observed in previous studies employing nuclear markers although is supported by mtDNA studies. Based on the results of Chapters **I** and **II**, and supported by other studies, colonization of the White and Barents Seas from two glacial refugia is suggested. This also has important conservation implications as the results clearly indicate that conservation of populations from all of the observed clusters is warranted. Thus, it is advisable to consider each cluster as a separate management unit. Chapter **II** also illustrates the importance of genotyping an adequate number of microsatellite loci to succeed in individual assignment. To be able to accurately assign individuals caught in offshore fisheries to their population of origin utilization of a minimum of 20 to 30 loci is recommended.

Chapter **III** demonstrated how differences in population life histories within a species, migratory behaviour in this case, and in their phylogeographic origin affect the genetic characteristics of populations, namely diversity and divergence levels. The anadromous populations from the Atlantic Ocean, White Sea and Barents Sea possessed higher levels of genetic diversity than the anadromous populations form the Baltic Sea basin. Among the non-anadromous populations the result was the opposite: the Baltic freshwater populations were more variable. This emphasises the importance of taking the life history of a population into consideration when developing conservation strategies: due to the limited possibilities for new genetic diversity to be generated via gene flow, it is expected that freshwater Atlantic salmon populations would be more vulnerable to extinction following a population crash. Hence, high conservation status is warranted in order to ensure the long-term survival of the limited number of European populations with this life-history strategy.

Lately, it has been suggested that genetic variation of adaptive importance, not just selectively neutral, should be incorporated into management plans (e.g. Crandall *et al.* 2000). Genome scans, including hitchhiking approaches, provide the means to identify functionally important genetic variation (e.g. Vasemägi & Primmer 2005, Storz 2005). In Chapter **IV** signatures of selection were searched for among immune relevant loci and among loci of other functions or presumably no function at all. Indeed, the results imply that the immune relevant loci and anonymous microsatellites might have experienced different selection pressures compared to non-immune relevant EST-linked loci. One reason for why a clearer difference between immune relevant and random loci was not found is the currently limited availability of genomic data for Atlantic salmon, which might have hampered the

classification of marker loci into correct categories thus making the marker groups more similar in terms of selection pressure than what they are in reality. Hence, this finding of differential selection pressures needs to be verified by further studies, which should become possible in the future as genomic data on Atlantic salmon accumulates.

As we are moving from the population *genetics* era to the population *genomics* era, information about the function of different genes in various organisms will increase, also in those species previously described as non-model organisms. The method of finding molecular markers linked to genes of functional importance described in Chapter **IV** will likely be proven useful for conservation purposes. Large numbers of molecular markers may be developed and the genetic structure of populations inferred with markers linked to genes of different function. This way, information about adaptive genetic variation among populations can easily be incorporated into management plans.

ACKNOWLEDGEMENTS

Making of this thesis has been quite a roller coaster for me and possibly taken me through the whole spectrum of emotions that human beings are capable of expressing. Several people have been there for me during this journey, both in moments of great joy and grave despair, and I wish to express my gratitude to them.

First of all and most importantly I wish to thank my supervisors professors Craig Primmer and Jaakko Lumme. Craig, I want to thank you for the wealth of support and time that you always gave me when I needed them. Also, you have created wonderful working settings around you and I feel privileged to have been a member of a group with such a friendly and supporting spirit.

Jaska, many things I know of Atlantic salmon I learned from you. You helped to put my research into the right perspective, which is greatly appreciated. And thank you for welcoming me in Oulu when I was missing my *alma mater* and finding a spot for me to work in.

I would also like to thank Anti Vasemägi for his help especially during the last and no doubt the most critical phases of making this thesis. I have really enjoyed our discussions related to detecting signals of selection and learned a lot about it from you. You truly have a gift of and the patience for explaining things.

A warm thank goes also to all of my co-authors, all ten of you. Thank you for your input and comments, which no doubt made this thesis as excellent as it is. I'm especially grateful to Alexei Veselov and Sergey Titov and all my other Russian collaborators who provided me with the samples and the information that are the basis of this thesis. Without you and all of those expeditions to the beautiful rivers of northwest Russia I wouldn't be here now!

I also wish to thank professor Pekka Pamilo, the former head of the section of genetics at the University of Oulu for letting me work there during the first year of my post-graduate studies although being enrolled at another university. The University and the Department of Biology especially will always have a special place in my heart and the years I spent there are definitely among the best of my life so far. Thank you Pekka also for organizing all those interesting graduate school courses.

A big, big thank you goes also to all present, past and visiting PnP members. Heikki, Paula, JP, Laura, Anti, Albert, Kalle, Ville, Akarapong, Katharina, José, Erica, and Hanna especially - thank you guys so very much for being you! Many of my feelings I have shared with you and always received so much support. I feel that one of the most important things that the making of this thesis has given me is your friendship! Albert, my years long office mate, I especially want to thank you for letting me, a Mac person, use your computer with a Windows operating system so many times. Katharina, your input with the lab work helped me very much and I'm grateful to you for doing it so well. Leena and Ville, thank you both for your reliability and efficiency. You truly keep things working smoothly!

The best part of moving to Turku were the people at the section of genetics, Raija, Pirjo, Christina, Seppo, Satu, and Niina, who welcomed us there. Everybody at genetics I want to thank you – I have enjoyed working with you and I have always felt like at home at the department. In particular, I wish to thank Satu, Tatjana, Raija and Pirjo for your assistance and for keeping the laboratory working.

When I arrived to Turku I also got acquainted with an excellent bunch of people nowadays also known as the Peggy people. Meri, Ritu, Outi, Tapio, Päivi, Mirkka, Sanna, Markus, Kalle, Ilkka, everybody, it has been wonderful to get to know you and share all of those coffee cups with you.

During these years I also had the privilege of working at the Department of Biological and Environmental Sciences at the University of Helsinki. That time brought more wonderful

people into my life (and a few old friends from Oulu too) and it is always fun meeting you guys again. So, Hannu, Theresa, Cim, Cano, Tuomas, Henna, Jukka, Mike, Marja, Maria - the list is much, much longer than this, really - thank you for the good times.

I also want to thank the folks at the University of Oulu. After leaving I still I met many of you several times a year on graduate school courses and on other instances too, which always brought good memories back to me. I really appreciate that! I would especially like to thank Mari, Anne, Johanna, and Paula for the excellent discussions we have had. I'm sure the world got a little bit better place to live after every dinner we had.

While in Turku I partially stepped back into student life. During these a bit over three years in Turku I shared a student flat in the lovely Pilvilinna with some marvellous gals. Reetta, Miia, Tiina, and Terhi I really had a lot of fun living with you. I especially want to thank Tiina and Terhi, my latest flatmates, for listening and understanding during the last stretch of preparing the thesis and for providing me an internet connection at odd hours.

I've saved the best for last. Maija, Sanna, Krista, and Teppo, my dearest friends, there are no words to describe how important all of you have been for me during these years and how much I value the support you have given me throughout the making of this thesis. I don't know what I have done to deserve the friendship of such wonderful people as you but I'm grateful for it and will cherish it in my heart for the rest of my life. I love you guys!

And finally… *Isä, sinä olet ollut tukenani kaikki nämä vuodet sekä jakanut kanssani kaikki iloni ja suruni ja kiitän sinua siitä. Olet paras ystäväni ja minulle rakkain ihminen koko maailmassa. Tämä väitöskirja on omistettu sinulle, maailman parhaalle isille.*

This thesis was funded by the Finnish Academy, the Finnish Ministry of Agriculture and Forestry, and the Graduate Schools of Population Genetics and Biological Interactions. Dr. Jodie Painter is acknowledged for her thorough and very professional English revision of the thesis.

REFERENCES

Andrén E, Andrén T & Sohlenius G (2000). The Holocene history of the southwestern Baltic Sea as reflected in a sediment core from the Bornholm Basin. *Boreas* 29: 233-250.

Asplund T, Veselov A, Primmer CR, Bakhmet I, Potutkin A, Titov S, Zubchenko A, Studenov I, Kaluzchin S & Lumme J (2004). Geographical structure and postglacial history of mtDNA haplotype variation in Atlantic salmon (*Salmo salar* L.) among rivers of the White and Barents Sea basins. *Annales Zoologi Fennici* 41: 465-475.

Avise JC (1998). The history and purview of phylogeography: a personal reflection. *Molecular Ecology* 7: 371-379.

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

Bakke TA, Jansen PA & Hansen LP (1990). Differences in the host resistance of Atlantic salmon, *Salmo salar* L., stocks to the monogenean *Gyrodactylus salaris* Malmberg, 1957. *Journal of Fish Biology* 37: 577-587.

Beaumont M & Nichols R (1996). Evaluating loci for use in the genetic analysis of population structure. *Proceedings of the Royal Society of London B, Biological Sciences* 263: 1619-1626.

Berg OK (1985). The formation of non-anadromous populations of Atlantic salmon, *Salmo salar* L., in Europe. *Journal of Fish Biology* 27: 805-815.

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

Bernatchez L & Wilson CC (1998). Comparative phylogeography of Nearctic and Palearctic fishes. *Molecular Ecology* 7: 431-452.

Björck S (1995). A review of the history of the Baltic Sea, 130-8.0 ka BP. *Quaternary International* 27: 19-40.

Bourke EA, Coughlan J, Jansson H, Galvin P & Cross TF (1997). Allozyme variation in populations of Atlantic salmon located throughout Europe: diversity that could be compromised by introductions of reared fish. *ICES Journal of Marine Science* 54: 974-985.

Brinkmann B, Klintschar M, Neuhuber F, Hühne J & Rolf B (1998). Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *American Journal of Human Genetics* 62: 1408-1415.

Callen DF, Thompson AD, Shen Y, Phillips HA, Richards RI, Mulley JC & Sutherland GR (1993). Incidence and origin of "null" alleles in the (AC)n microsatellite markers. *American Journal of Human Genetics* 52: 922-927.

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

Cavalli-Sforza LL & Edwards AWF (1967). Phylogenetic analysis: models and estimation procedures. *American Journal of Human Genetics* 19: 233-257.

Collins CM, Olstad K, Sterud E, Jones CS, Noble LR, Mo TA & Cunningham CO (2007). Isolation of a FIP2-like gene from Atlantic salmon (*Salmo salar* L.), found upregulated following infection with the monogenean parasite *Gyrodactylus salaris* Malmberg, 1957. *Fish and Shellfish Immunology* 22: 282-288.

Consuegra S, García De Leániz C, Serdio A, Gonzalez Morales M, Straus LG, Knox D & Verspoor E (2002). Mitochondrial DNA variation in Pleistocene and modern Atlantic salmon from the Iberian glacial refugium. *Molecular Ecology* 11: 2037-2048.

Cornuet J, Piry S, Luikart G, Estoup A & Solignac M (1999). New methods employing multilocus genotypes to select or exclude populations as origins of individuals. *Genetics* 153: 1989-2000.

Crandall KA, Bininda-Emonds ORP, Mace GM & Wayne RK (2000). Considering evolutionary processes in conservation biology. *Trends in Ecology and Evolution* 15: 290-295.

Crespi BJ & Fulton MJ (2004). Molecular systematics of Salmonidae: combined nuclear data yields a robust phylogeny. *Molecular Phylogenetics and Evolution* 31: 658-679.

Crozier RH, Agapow PM & Pedersen K (1999). Towards complete biodiversity assessment: an evaluation of the subterranean bacterial communities in the Oklo region of the sole surviving natural nuclear reactor. *FEMS Microbiology Ecology* 28: 325-334.

DeWoody JA & Avise JC (2000). Microsatellite variation in marine, freshwater and anadromous fishes compared with other animals. *Journal of Fish Biology* 56: 461-473.

Di Rienzo A, Peterson AC, Garza JC, Valdes AM, Slatkin M & Freimer NB (1994). Mutational processes of simple-sequence repeat loci in human populations. *Proceedings of the National Academy of Sciences, USA* 91: 3166-3170.

Dietrich WF, Miller J, Steen R, Merchant MA, Damron-Boles D, Husain Z, Dredge R, Daly MJ, Ingalls KA, O'Connor TJ, Evans CA, DeAngelis MM, Levinson DM, Kruglyak L & Lander ES (1996). A comprehensive genetic map of the mouse genome. *Nature* 380: 149-152.

Ehrich D & Stenseth NC (2001). Genetic structure of Siberian lemmings (*Lemmus sibiricus*) in a continuous habitat: large patches rather than isolation by distance. *Heredity* 86: 716-730.

Estoup A & Angers B (1998). Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. In: *Advances in molecular ecology* (ed. Carvalho GR), pp. 55-85. IOS Press, Amsterdam.

Fitzsimmons NN (1998). Single paternity of clutches and sperm storage in the promiscuous green turtle (*Chelonia mydas*). *Molecular Ecology* 7: 575-584.

Frankham R, Ballou JD & Briscoe DA (2002). *Introduction to conservation genetics*. Cambridge University Press, United Kingdom.

Fraser DJ & Bernatchez L (2001). Adaptive evolutionary conservation: towards a unified concept for defining conservation units. *Molecular Ecology* 10: 2741-2752.

Gardner MG, Bull CM, Cooper SJB & Duffield GA (2000). Microsatellite mutations in litters of the Australian lizard *Egernia stokesii*. *Journal of Evolutionary Biology* 13: 551-560.

Gilbey J, Verspoor E, Mo TA, Sterud E, Olastad E, Hytterød S, Jones C & Noble L (2006). Identification of genetic markers associated with *Gyrodactylus salaris* resistance in Atlantic salmon *Salmo salar*. *Diseases of Aquatic Organisms* 71: 119-129.

Hansen MM, Villanueva B, Nielsen EE & Bekkevold D (2007). Investigating the genetics of populations. In: *The Atlantic salmon: genetics, conservation and management* (eds. Verspoor E, Strandmeyer L & Nielsen JL), pp. 86-109. Blackwell Publishing, Oxford, UK.

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

Hedrick PW (1999). Perspective: Highly variable loci and their interpretation in evolution and conservation. *Evolution* 53: 313-318.

Hedrick PW (2005). A standardized genetic differentiation measure. *Evolution* 59: 1633-1638.

Hewitt GM (2004). The structure of biodiversity - insights from molecular phylogeography. *Frontiers in Zoology* 1. *doi: 10.1186/1742-9994-1-4*.

Hey J (1997). Mitochondrial and nuclear genes present conflicting portraits of human origins. *Molecular Biology and Evolution* 14: 166-172.

Hutchings JA & Jones MEB. (1998). Life history variation and growth rate thresholds for maturity in Atlantic salmon, *Salmo salar*.

Canadian Journal of Fisheries and Aquatic Science 55: 22-47.

Joost S, Bonin A, Bruford MW, Després L, Conord C, Erhardt G & Taberlet P (2007). A spatial analysis method (SAM) to detect candidate loci for selection: towards a landscape genomics approach to adaptation. *Molecular Ecology* 16: 3955-3969.

Kauer MO, Dieringer D & Schlötterer C (2003). A microsatellite variability screen for positive selection associated with the "out of Africa" habitat expansion of *Drosophila melanogaster*. *Genetics* 165: 1137-1148.

Kayser M, Roewer L, Hedman M, Henke L, Henke J, Brauer S, Krüger C, Krawczak M, Nagy M, Dobosz T, Szibor R, de Knijff P, Stoneking M & Sajantila A (2000). Characteristics and frequency of germline mutations at microsatellite loci from the human Y chromosome, as revealed by direct observation in father/son pairs. *American Journal of Human Genetics* 66: 1580-1588.

Kazakov RV & Titov SF (1991). Geographical patterns in the population genetics of Atlantic salmon, *Salmo salar* L., on U.S.S.R. territory, as evidence for colonization routes. *Journal of Fish Biology* 39: 1-6.

King TL, Kalinowski ST, Schill WB, Spidle AP & Lubinski BA (2001). Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. *Molecular Ecology* 10: 807-821.

King TL, Eackles MS & Letcher BH (2005). Microsatellite DNA markers for the study of Atlantic salmon (*Salmo salar*) kinship, population structure, and mixed-fishery analyses. *Molecular Ecology Notes* 5: 130-132.

Koljonen M-L, Jansson H, Paaver T, Vasin O & Koskiniemi J (1999). Phylogeographic lineages and differentiation pattern of Atlantic salmon (*Salmo salar*) in the Baltic Sea with management implications. *Canadian Journal of Fisheries and Aquatic Sciences* 56: 1766-1780.

Kudersky LA, Ieshko E & Schulman B (2003). Distribution range formation history of the monogenean *Gyrodactylus salaris* Malmberg, 1957 - a parasite of juvenile Atlantic salmon *Salmo salar* Linnaeus, 1758. In: *Atlantic salmon: biology, conservation and restoration* (eds. Veselov AJ, Ieshko EP, Nemova NN, Sterligova OP & Shustov YA), pp. 77-83. Russian Academy of Science, Karelian Research Center, Institue of Biology, Petrozavodsk, Russia.

Levinson G & Gutman GA (1987). Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Molecular Biology and Evolution* 4: 203-221.

Litt M & Luty JA (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. *American Journal of Human Genetics* 44: 397-401.

MacCrimmon HR & Gots BL (1979). World distribution of Atlantic salmon, *Salmo salar*. *Journal of the Fisheries Research Board of Canada* 36: 422-457.

MacHugh DE (1996). *Molecular biogeography and genetic structure of domesticated cattle*. PhD thesis, University of Dublin.

Makhrov AA, Verspoor E, Artamonova VS & O'Sullivan M (2005). Atlantic salmon colonization of the Russian Arctic coast: pioneers from the North America. *Journal of Fish Biology* 67 (supplement A): 68-79.

Mangerud J, Astakhov VI, Murray A & Svendsen JI (2001). The chronology of a large ice-dammed lake and the Barents-Kara Ice Sheet advances, northern Russia. *Global and Planetary Change* 31: 321-336.

Mangerud J, Jakobsson M, Alexanderson H, Astakhov V, Clarke GKC, Henriksen M, Hjort C, Krinner G, Lunkka J-P, Möller P, Murray A, Nikolskaya O, Saarnisto M & Svendsen JI (2004). Ice-dammed lakes and rerouting of the drainage of northern Eurasia during the Last Glaciation. *Quaternary Science Reviews* 23: 1313-1332.

Martin SAM, Blaney SC, Houlihan DF & Secombes CJ (2006). Transcriptome response following administration of a live bacterial vaccine in Atlantic salmon (*Salmo salar*). *Molecular Immunology* 43: 1900-1911.

Matejusová I, Felix B, Sorsa-Leslie T, Gilbey J, Noble LR, Jones CS & Cunningham CO (2006). Gene expression profiles of some

immune relevant genes from skin of susceptible and responding Atlantic salmon (*Salmo salar* L.) infected with *Gyrodactylus salaris* (Monogenea) revealed by suppressive subtractive hybridisation. *International Journal for Parasitology* 36: 1175-1183.

Maynard Smith J & Haigh J (1974). The hitch-hiking effect revisited. *Genetical Research* 23: 23-35.

McConnell S, Hamilton L, Morris D, Cook D, Paquet D, Bentzen P & Wright J (1995a). Isolation of salmonid microsatellite loci and their application to the population genetics of Canadian east coast stocks of Atlantic salmon. *Aquaculture* 137: 19-30.

McConnell SK, O'Reilly P, Hamilton L, Wright JM & Bentzen P (1995b). Polymorphic microsatellite loci from Atlantic salmon (*Salmo salar*): Genetic differentiation of North American and European populations. *Canadian Journal of Fisheries and Aquatic Sciences* 52: 1863-1872.

Mills D (1989). *Ecology and management of Atlantic salmon.* Chapman and Hall, London, New York.

Moritz C (1994). Defining 'evolutionarily significant units' for conservation. *Trends in Ecology and Evolution* 9: 373-375.

Nilsson J, Gross R, Asplund T, Dove O, Jansson H, Kelloniemi J, Kohlmann K, Löytynoja A, Nielsen EE, Paaver T, Primmer CR, Titov S, Vasemägi A, Veselov A, Öst T & Lumme J (2001). Matrilinear phylogeography of Atlantic salmon (*Salmo salar* L.) in Europe and postglacial colonization of the Baltic Sea area. *Molecular Ecology* 10: 89-102.

Ohta T & Kimura M (1973). A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population. *Genetical Research* 22: 201-204.

Pamilo P & Nei M (1988). Relationships between gene trees and species trees. *Molecular Biology and Evolution* 5: 568-583.

Parrish DL, Behnke RJ, Gephard SR, McCormick SD & Reeves GH (1998). Why aren't there more Atlantic salmon (*Salmo salar*)? *Canadian Journal of Fisheries and Aquatic Sciences* 55 (supplement 1): 281-287.

Peeler E, Thrush M, Paisley L & Rodgers C (2006). An assessment of the risk of spreading the fish parasite *Gyrodactylus salaris* to uninfected territories in the European Union with the movement of live Atlantic salmon (*Salmo salar*) from coastal waters. *Aquaculture* 258: 187-197.

Petit RJ, El Mousadik A & Pons O (1999). Identifying populations for conservation on the basis of genetic markers. *Conservation Biology* 12: 844-855.

Phillips RB, Matsuoka MP, Konon I & Reed KM (2000). Phylogenetic analysis of mitochondrial and nuclear sequences supports inclusion of *Acantholingua ohridana* in the genus *Salmo*. *Copeia* 2000: 546-550.

Primmer CR, Ellegren H, Saino N & Møller AP (1996). Directional evolution in germline microsatellite mutations. *Nature Genetics* 13: 391-393.

Primmer CR, Saino N, Møller AP & Ellegren H (1998). Unraveling the processes of microsatellite evolution through analysis of germ line mutations in barn swallows *Hirundo rustica*. *Molecular Biology and Evolution* 15: 1047-1054.

Queller DC, Strassmann JE & Hughes CR (1993). Microsatellites and kinship. *Trends in Ecology and Evolution* 8: 285-288.

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

Roberge C, Paez DJ, Rossignol O, Guderley H, Dodson J & Bernatchez L (2007). Genome-wide survey of the gene expression response to saprolegniasis in Atlantic salmon. *Molecular Immunology* 44: 1374-1383.

Rousset F (1996). Equilibrium values of measures of population subdivision for stepwise mutation processes. *Genetics* 142: 1357-1362.

Rousset F (1997). Genetic differentiation and estimation of gene flow from F-statistics under isolation by distance. *Genetics* 145: 1219-1228.

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

Saarnisto M & Saarinen T (2001). Deglaciation chronology of the Scandinavian Ice Sheet from the Lake Onega Basin to the Salpausselkä End Moraines. *Global and Planetary Change* 31: 387-405.

Saarnisto M, Grönlund T & Ekman I (1995). Lateglacial of Lake Onega - contribution to the history of the eastern Baltic Basin. *Quaternary International* 27: 111-120.

Schlötterer C, Vogl C & Tautz D (1997). Polymorphism and locus-specific effects on polymorphism at microsatellite loci in natural *Drosophila melanogaster* populations. *Genetics* 146: 309-320.

Schlötterer C (2002). A microsatellite-based multilocus screen for the identification of local selective sweeps. *Genetics* 160: 753-763.

Shimoda N, Knapik EW, Ziniti J, Sim C, Yamada E, Kaplan S, Jackson D, De Sauvage F, Jacob H & Fishman MC (1999). Zebrafish genetic map with 2000 microsatellite markers. *Genomics* 58: 219-232.

Slatkin M (1993). Isolation by distance in equilibrium and non-equilibrium populations. *Evolution* 47: 264-279.

Slatkin M (1995). Hitchhiking and associative overdominance at a microsatellite locus. *Molecular Biology and Evolution* 12: 473-480.

Snoj A, Melkic E, Susnik S, Muhamedagic S & Dovc P (2002). DNA phylogeny supports revised classification of *Salmothymus obtusirostris*. *Biological Journal of the Linnean Society* 77: 399-411.

Storz JF (2005). Using genome scans of DNA polymorphism to infer adaptive population divergence. *Molecular Ecology* 14: 671-688.

Ståhl G (1987). Genetic population structure of Atlantic salmon. In: *Population genetics and fishery management* (eds. Ryman N & Utter F), pp. 121-140. University of Washington Press, Seattle.

Svendsen JI, Astakhov VI, Bolshiyanov DY, Demidov I, Dowdeswell JA, Gataullin V, Hjort C, Hubberten HW, Larsen E, Mangerud J, Melles M, Möller P, Saarnisto M & Siegert MJ (1999). Maximum extent of the Eurasian ice sheets in the Barents and Kara Sea region during the Weichselian. *Boreas* 28: 234-242.

Svendsen JI, Alexanderson H, Astakhov VI, Demidov I, Dowdeswell JA, Funder S, Gataullin V, Henriksen M, Hjort C, Houmark-Nielsen M, Hubberten HW, Ingolfsson O, Jakobsson M, Kjær KH, Larsen E, Lokrantz H, Lunkka JP, Lysså A, Mangerud J, Matiouchkov A, Murray A, Möller P, Niessen F, Nikolskaya O, Polyak L, Saarnisto M, Siegert C, Siegert MJ, Spielhagen RF & Stein R (2004). Late Quaternary ice sheet history of northern Eurasia. *Quaternary Science Reviews* 23: 1229-1271.

Säisä M, Koljonen M-L, Gross R, Nilsson J, Tähtinen J, Koskiniemi J & Vasemägi A (2005). Population genetic structure and postglacial colonization of Atlantic salmon (*Salmo salar*) in the Baltic Sea area based on microsatellite DNA variation. *Canadian Journal of Fisheries and Aquatic Sciences* 62: 1887-1904.

Tautz D (1989). Hypervariability of simple sequences as a source for polymorphic DNA markers. *Nucleic Acids Research* 17: 6463-6471.

Tsoi SCM, Ewart KV, Penny S, Melville K, Liebscher RS, Brown LL & Douglas SE (2004). Identification of immune-relevant genes from Atlantic salmon using suppression subtractive hybridization. *Marine Biotechnology* 6: 199-214.

Vähä J-P (2007). *Conservation genetics of Teno river Atlantic salmon (Salmo salar): genetic structure in space and time, and the effect of escaped farmed salmon.* PhD thesis, University of Turku.

Vasemägi A & Primmer CR (2005). Challenges for identifying functionally important genetic variation: the promise of combining complementary research strategies. *Molecular Ecology* 14: 3623-3642.

Vasemägi A, Nilsson J & Primmer CR (2005). Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo*

salar L.). *Molecular Biology and Evolution* 22: 1067-1076.

Verspoor E, McCarthy EM, Knox D, Bourke EA & Cross TF (1999). The phylogeography of European Atlantic salmon (*Salmo salar* L.) based on RFLP analysis of the ND1/16sRNA region of the mtDNA. *Biological Journal of the Linnean Society* 68: 129-146.

Verspoor E (2007). Introduction. In: *The Atlantic salmon: genetics, conservation and management* (eds. Verspoor E, Stradmeyer L & Nielsen JL), pp. 1-13. Blackwell Publishing, Oxford, UK.

Webb J, Verspoor E, Aubin-Horth N, Romakkaniemi A & Amiro P (2007). The Atlantic salmon. In: *The Atlantic salmon: genetics, conservation and management* (eds. Verspoor E, Stradmeyer L & Nielsen JL), pp. 17-46. Blackwell Publishing, Oxford, UK.

Weber JL & May PE (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* 44: 388-396.

Weber JL & Wong C (1993). Mutation of human short tandem repeats. *Human Molecular Genetics* 2: 1123-1128.

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

WWF (2001). *The status of wild Atlantic salmon: a river by river assessment*. WWF.

Xu X, Peng M, Fang Z & Xu X (2000). The direction of microsatellite mutations is dependent upon allele length. *Nature Genetics* 24: 396-399.