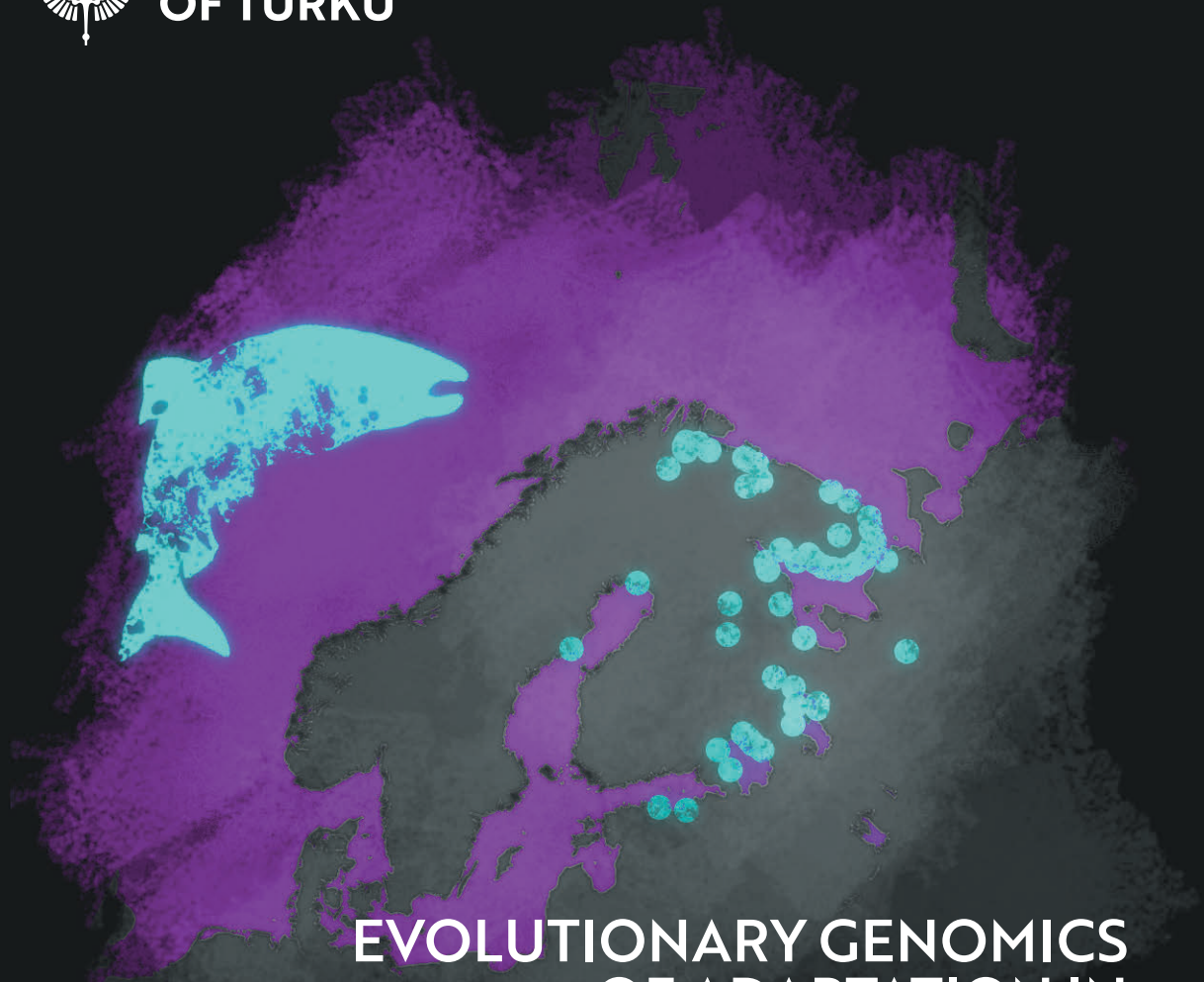
UNIVERSITY OF TURKU



EVOLUTIONARY GENOMICS OF ADAPTATION IN ATLANTIC SALMON FROM NORTHERN EUROPE

Ksenia J. Zueva

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. AII OSA – TOM. 376 | BIOLOGICA – GEOGRAPHICA – GEOLOGICA | TURKU 2021

UNIVERSITY OF TURKU

EVOLUTIONARY GENOMICS OF ADAPTATION IN ATLANTIC SALMON FROM NORTHERN EUROPE

Ksenia J. Zueva

TURUN YLIOPISTON JULKAISUJA – ANNALES UNIVERSITATIS TURKUENSIS
SARJA – SER. AII OSA – TOM. 376 | BIOLOGICA – GEOGRAPHICA – GEOLOGICA | TURKU 2021

University of Turku

Faculty of Science and Engineering
Department of Biology
Laboratory of Genetics
Doctoral Programme in Biology, Geography and Geology (BGG)

Supervised by

Professor Craig R. Primmer
University of Helsinki
Finland

Reviewed by

Docent Emily Knott
University of Jyväskylä
Finland

Professor Lisa Seeb
University of Washington
USA

Opponent

Senior Researcher Dorte Bekkevold
Technical University of Denmark
National Institute of Aquatic Resources
Denmark

Cover Image: Matthieu Bruneaux, Ksenia Zueva

ISBN 978-951-29-8327-8 (PRINT)
ISBN 978-951-29-8328-5 (PDF)
ISSN 0082-6979 (Print)
ISSN 2343-3183 (Online)
Painosalama Oy, Turku, Finland 2021

かたつぶり
そろそろ登れ
富士の山

小林一茶

Snail
Little by little climb
Mount Fuji

Kobayashi Issa

UNIVERSITY OF TURKU
Faculty of Science and Engineering
Department of Biology
Laboratory of Genetics
KSENIA J. ZUEVA: Evolutionary genomics of adaptation in Atlantic salmon from northern Europe
Doctoral Dissertation, 97 pp.
Doctoral Programme in Biology, Geography and Geology (BGG)
January 2021

ABSTRACT

Insight into genomic mechanisms of phenotypic variation and adaptation is essential for understanding evolutionary and population dynamics in wild populations. Knowledge about whether it is the same genetic architecture that underlies adaptation over different geographical scales and regions, and what role population history plays, is paramount for the consequent development of efficient conservation practices for the species.

Salmonid fishes are commonly characterised by a wide geographic range, distinct population structure, and high incidence of local adaptation, which makes them a great target for studies exploring both the genomic basis of adaptation and the comparative significance of loci involved in adaptation within and between species. The Atlantic salmon (*Salmo salar*) populations of northern Europe are particularly interesting: they include the least disturbed populations left in the wild, belong to several distinct phylogeographic lineages, and exhibit astonishing natural variation in response to a salmon ectoparasite, *Gyrodactylus salaris*, ranging from near resistance in the landlocked and Baltic salmon to high susceptibility with devastating effect in Atlantic Ocean salmon.

In this study, I used genome-wide approaches to further characterize the population structure and phylogeographic history of northern European Atlantic salmon (Chapters I-III). I explored the mechanisms behind the observed variation in the levels of susceptibility to *G. salaris*, by searching for genes playing a key role in the response to the parasite (Chapters I and II). Subsequently, I broadened my work to search for genomic regions involved in local adaptation in general. I examined whether the identified selection targets were similar over a broad geographic range and independent studies, and thus whether there are patterns of adaptive divergence that could be universal across Atlantic salmon populations (Chapter III).

To achieve this, I used a large collection of Atlantic salmon samples and applied two SNP arrays of varying density to individual and pooled-per-population DNA samples. I looked for genomic signatures of directional selection in response to specific selective pressures, including *G. salaris* presence (Chapters I and II). I also looked for loci that may underly local adaptation in general by examining signatures of divergent directional selection among three geographically and genetically distinct sets of populations (Chapter III). To overcome the challenge of correlated environmental traits and the confounding effects of neutral evolution I used a careful

methodological strategy, taking into account the phylogeographic relationships of populations and considering only repeated lines of evidence over multiple analyses.

Several genomic regions, genes, and single SNP outliers were identified in relation to the observed variation in susceptibility to *G. salaris*, and to other potential selective pressures. Analyses of gene functions and comparison to other research suggest that the detected loci under *G. salaris*-mediated selection are participating in control of both innate and acquired immune systems. As there were few genes involved uniquely in immunity among the parasite-related candidates, my results highlight that the immune response in Atlantic salmon may be mediated by a large number of multi-functional loci (Chapters I and II). When examining for locally adaptive candidates in general, seventeen haploblocks were repeatedly found as candidates for divergent selection within different population groups. Several of these genomic regions contained loci known to be of large effect and to be associated with life-history traits and, interestingly, immunity (Chapter III).

Overall, this thesis provides evidence that diversification in Atlantic salmon is driven both by multiple loci acting in specific population groups, and by few large-effect loci acting over a wide geographic range. Exploring the effect of these loci on salmon fitness would help to validate the importance of identified genes and help to assess the long-term viability of northern European salmon.

KEYWORDS: Atlantic salmon, adaptive diversification, natural selection, *Gyrodactylus salaris*

TURUN YLIOPISTO
Luonnontieteiden ja tekniikan tiedekunta
Biologian laitos
Genetiikan laboratorio
KSENIA J. ZUEVA: Pohjois-Euroopan lohen sopeutumisen evoluutiogenettinen tausta
Väitöskirja, 97 s.
Biologia, maantieteiden ja geologian tohtoriohjelma (BGG)
Tammikuu 2021

TIIVISTELMÄ

Jotta ymmärtäisimme luonnonpopulaatioiden evoluutio- ja populaatiodynamiikkaa, on ensisijaisen tärkeää selvittää fenotyyppisen vaihtelun sekä adaptaation takana olevat geneettiset mekanismit. Suunniteltaessa tehokkaita suojelutoimia on tärkeää tietää, ovatko maantieteellisesti eri etäisyyksillä ja alueilla esiintyvissä sopeumissa taustalla samat geneettiset rakenteet. Lisäksi on erityisen tärkeää ymmärtää, mikä osuus populaation historialla on.

Lohi-kalat ovat levittäytyneet maantieteellisesti laajalle alueelle, populaatioiden rakenteet ovat erilaisia ja paikallinen adaptaatio eli sopeutuminen on huomattavaa. Nämä piirteet tekev ät lohesta hyvän tutkimuskohteen sekä sopeutumisen geneettisen taustan selvittämiseen että sopeumien taustalla olevien lokusten vertailuun, niin lajien sisällä kuin niiden välillä. Pohjois-Euroopan lohi (*Salmo salar*) on erityisen mielenkiintoisia useasta syystä: populaatiot muodostavat useita erillisiä fylogeografisia linjoja, ja osaa populaatioista ihmisen toiminta on häirinnyt vain vähän. Lisäksi, vasteena *Gyrodactulus salaris* -ulkoloiselle, Pohjois-Euroopan lohipolulaatioissa esiintyy hämmästyttävä määrä luonnollista muuntelua täysin vastustuskykyisistä sisävesien ja Baltian alueen populaatioista hyvin alttiisiin populaatioihin, joissa loisen vaikutukset ovat tuhoisia.

Tässä tutkimuksessa käytin koko genomin kattavia menetelmiä selvittääkseni Pohjois-Euroopan lohien populaatiorakennetta sekä fylogeografista historiaa (Kappaleet I-III). Tutkin *G. salaris* -alttiuteen vaikuttavia mekanismeja etsimällä geenejä, jotka ovat keskeisiä loisten aiheuttamille vasteille (Kappaleet I ja II). Laajensin tutkimusta etsimällä genomista alueita, jotka yleisesti liittyvät paikalliseen sopeutumiseen. Tutkin, olivatko tunnistamani alueet samoja maantieteellisesti laajalla alueella ja olivatko erilliset itsenäiset tutkimukset havainneet samoja alueita, eli sisältävätkö eri lohipopulaatioissa sopeutumisen seurauksena eriytyneet alueet universaaleja piirteitä (Kappale III).

Saavuttaakseni em. tavoitteet käytin laajaa lohinäytteiden kokoelmaa; sovelsin kahta eri tiheyksistä SNP-sirua DNA-näytteisiin, jotka oli otettu joko yksilöistä tai poolattu populaatioista. Etsin genomista merkkejä suuntaavasta valinnasta vasteena tiettyihin valintapaineisiin, mukaanlukien *G. salaris* -loisen läsnäolo (Kappaleet I ja II). Lisäksi etsin lokuksia, jotka voivat yleisesti olla paikallisen adaptaation takana; etsin suuntaavan valinnan merkkejä kolmesta maantieteellisesti ja geneettisesti erillisestä populaatiosta (Kappale III). Hallitakseni korreloivia ympäristötekijöitä

sekä neutraalin evoluution haittaavia vaikutuksia käytin analysoinnissa tarkkaa metodologista strategiaa: otin huomioon populaatioiden fylogeografiset suhteet sekä käsittelin ainoastaan tuloksia, jotka toistuivat eri analyyseissä.

Tunnistin useita genomin alueita, geenejä sekä yksittäisiä poikkeavia SNP:tä, jotka liittyivät havaittuun alttiuteen *G. salaris* -loiselle sekä myös muihin mahdollisiin valintapaineisiin. Geenien toiminnan analysointi sekä vertailu aikaisempaan tutkimukseen osoitti, että nyt tunnistetut *G. salaris* -valintapaineen alaiset lokukset osallistuvat sekä sisäsyntyiseen että hankittuun immuniteettiin. Koska vain muutamat näistä kandidaattigeeneistä liittyivät yksinomaan immuniteettiin, tulokseni korostavat, että lohen immuunivaste saattaa olla usean monitoimisen lokuksen aikaansaama (Kappaleet I ja II). Vertaillessani yleisesti paikalliseen sopeutumiseen liittyviä kandidaattigeenejä tunnistin toistuvasti kaikkiaan seitsemäntoista haploblokkia eri populaatioissa. Useat näistä genomin alueista sisälsivät lokuksia, joiden tiedetään olevan vaikutukseltaan merkittäviä sekä liityvän niin lajin elinhistoriaan kuin, kiinnostavaa kyllä, immuniteettiin (Kappale III).

Kokonaisuudessaan tämä väitöstyö osoittaa, että lohen erilaistuminen on sekä useiden erilaisten, yksittäisissä populaatioissa vaikuttavien lokuksien että muutaman suurella maantieteellisellä alueella vaikuttavan lokuksen tulos. Tutkimalla näiden lokuksien vaikutusta lohen elinkelpoisuuteen voidaan varmistaa tunnistettujen geenien tärkeys sekä arvioida Pohjois-Euroopan lohien elinkelpoisuutta pitkällä aikavälillä.

ASIASANAT: Lohi, sopeutumisesta johtuva erilaistuminen, luonnonvalinta, *Gyrodactylus salaris*

Table of Contents

List of Original Publications

This dissertation is based on the following original publications, which are referred to in the text by their Roman numerals:

I Zueva K. J., Lumme J., Veselov A. E., Kent M. P., Lien S., Primmer C. R. Footprints of directional selection in wild Atlantic salmon populations: evidence for parasite-driven evolution? *PLoS One*, 2014; 9(3): e91672.

II Zueva K. J., Lumme J., Veselov A. E., Kent M. P., Primmer C. R. Genomic signatures of parasite-driven natural selection in north European Atlantic salmon (*Salmo salar*). *Marine Genomics*, 2018; 39: 26–38.

III Zueva K. J., Lumme J., Veselov A. E., Primmer C. R., Pritchard V. L. Population genomics reveals repeated signals of adaptive divergence in the Atlantic salmon of north-eastern Europe. *Journal of Evolutionary Biology*, 2020; 00:1-13. https://doi.org/10.1111/jeb.13732

Author contributions to the original publications:

	I	II	III
Original idea and research design	KZ, CP	KZ, CP	CP, VP, KZ
Data (sample) collection and contribution	AV, JL, CP	AV, JL, MK, CP	AV, JL, CP
Performing experiments & lab work	MK, SL	KZ, MK	KZ
Data analysis	KZ, CP	KZ	VP, KZ
Writing of the manuscript	KZ, CP	KZ, CP	KZ, VP

KZ= Ksenia Zueva, CP= Craig Primmer, VP= Victoria Pritchard, JL= Jaakko Lumme, AV= Alexey Veselov, MK= Matthew Kent, SL= Sigbjørn Lien.

1 Introduction

1.1 Overview

Identifying the genomic architecture of phenotypic variation and local adaptation across populations and species is a prime focus of evolutionary biology. This involves questions such as what types of genetic variation underly traits of interest, how this genetic variation is distributed across the genome, and whether it differs among populations across time and space. Identifying the molecular basis of adaptive and/or phenotypic divergence is an essential step in understanding how species adapt to their environment and, consequently, in predicting population dynamics in changing environments and in designing effective management practices (Bekkevold et al., 2020; Shaw & Etterson, 2012; Waldvogel et al., 2020). From a broader evolutionary perspective, it is also essential to understand if the same genomic architecture underlies adaptation across different geographic scales and regions, what role the history of population demography plays, and what is the relative impact of loci of small and large effect on evolutionary processes (Deagle et al., 2013; Rissler, 2016; Timpson et al., 2018).

The first step in identifying the molecular basis of adaptive and/or phenotypic divergence is to detect genes that are evolving under selection pressure, by distinguishing them from genes evolving only under background neutral genomic processes.

1.2 Detecting directional selection

1.2.1 Theoretical basis of tests for directional selection

The availability of large genomic resources that have flourished in the last 15-20 years has made it possible to use so-called ‘genome scans’ to study genetic variation across entire genomes in a search for regions subject to natural selection. Signatures that selection leaves in genomes depend on selection type, strength, and time frame (Oleksyk et al., 2010). For instance, strong positive selection leads to increased frequency or even fixation of beneficial alleles in a population, with simultaneous reduction of standing genetic variation in linked neutral sites (processes known as

'selective sweep' and 'genetic hitchhiking') (Manel et al., 2016; Nielsen et al., 2005). If selection pressure differs across populations, the genetic divergence between the populations would increase at selected loci and linked neutral loci, compared to the non-selected loci.

Several approaches have been used to identify signals of genetic divergence and potential selection. The 'outlier' methods are based on detecting loci with elevated interpopulation differentiation (estimated using F_{ST} or similar) when compared to an empirical or neutral distribution, and they are widely used to infer local adaptation in wild populations and non-model species (Oleksyk et al., 2010). However, the results can be difficult to interpret, as those methods give no information as to which environmental or life history constraints 'outlier' loci are related to. Another main type of genome scan methods utilises prior knowledge about environmental or phenotypic characteristics distinguishing the populations, and is based on the assumption that allele frequencies of selected loci should be unusually correlated with the environmental variable exerting selective pressure (Coop et al., 2010).

Even though the mentioned approaches are likely to miss weakly selected loci (Whitlock & Lotterhos, 2015), they are efficient in detecting strong selection, and are widely used to infer selection and adaptation in domestic and wild populations (Haasl & Payseur, 2016), including in salmonid fishes (Elmer, 2016).

1.2.2 Methods to assess candidate loci validity

Genome scan methods aim at separating footprints of positive selection from baseline variation shaped by neutral processes such as gene flow, inbreeding, and genetic drift (Gautier, 2015). These mentioned demographic processes can affect allele frequencies in a manner similar to natural selection, and distinguishing between the two can be quite challenging, especially in populations with small effective population size (Oleksyk et al., 2010). Apart from demography, footprints of directional selection can be mimicked by other selective forces acting on a genomic region, e.g. balancing selection in regions of low recombination (Matthey-Doret & Whitlock, 2019; Weigand & Leese, 2018). Moreover, in populations varying in their standing genetic variation the same selective pressure may result in different loci being associated with the adaptive phenotype, which may complicate the detection of selected loci in populations with different phylogeographic histories (Przeworski et al., 2005). Another challenge in inferring the genomic basis of local adaptation to a particular agent of selection originates from correlated environmental pressures commonly occurring in wild populations. For example, variation in pathogen communities affecting the hosts is often associated with

variation in food regimes and habitat, which complicates identification of parasite-specific genetic divergence (Karvonen & Seehausen, 2012).

Due to such challenges, and given the fact that large datasets lead to a large number of comparisons being performed in a single analysis, genome scans for selection may result in a substantial number of false positives (Weigand & Leese, 2018). A widely used tactic to strengthen the validity of detected loci is to consider only overlapping results after using several alternative methods of loci detection, and/or analysing several independent datasets (Hoban et al., 2016; Rellstab et al., 2015). Loci detected with parallel and independent lines of evidence are less prone to type I error and are more likely to be truly important mediators of adaptive response. One important step in candidate loci validation is assessing the repeatability of results over multiple independent studies of the same populations, and of different populations across varying phylogeographic histories. The latter can improve our understanding of global evolutionary dynamics, for instance of the genetic architecture behind adaptation and diversification across a broad geographic scale (Turner et al., 2018; Yeaman et al., 2018).

In Atlantic salmon (*Salmo salar*), while a few loci have been repeatedly identified as underlying life-history variation or local adaptation in independent studies (Ayllon et al., 2015; Barson et al., 2015; Bourret et al., 2013; Pritchard et al., 2018), the majority of candidates remain unique to a single study. It is therefore paramount to design further research in a manner that would maximise the chance to detect new and validate already identified loci of adaptive importance, in order to infer their evolutionary significance at the scale of the whole species.

1.3 Atlantic salmon (*Salmo salar*) as a study system

1.3.1 Overview

Salmonid fishes, including Atlantic salmon, are species with pronounced genetic population structuring and widespread local adaptations, which arise from the precise homing behaviour that drives fish to return to their natal river for spawning, as well as from geographic isolation in different water basins (rev. by Fraser et al., 2011). Population genetic structure is usually temporally stable (Ozerov et al., 2013; Palstra & Ruzzante, 2010; Primmer, 2011) and local adaptation is prominent at various geographic scales, ranging from hundreds of kilometres between rivers to just a few kilometres between single tributaries (Fraser et al., 2011; Primmer, 2011). Salmonids exhibit great ecological, physiological, and life history diversity, and local adaptation can arise as quickly as in a few generations (rev. in Primmer 2011). This makes salmonids the perfect target for studies looking for the genetic basis of

local adaptation and aiming to answer broad evolutionary questions such as what genomic architecture underlies adaptive divergence, and parallel and convergent evolution.

Atlantic salmon (*Salmo salar*) is found in northern Atlantic Ocean and rivers flowing into it, and is a socially, economically and culturally important species with high value for commercial and recreational fishing, aquaculture, and local communities (Myrvold et al., 2019). Understanding its biology and the genomics behind its local adaptation is of utmost importance for species management and conservation, especially given the alarming global decline in wild salmon numbers, and the fact that this species is endangered or extinct in certain parts of its range (NASCO Report, 2019).

In my thesis I focus on wild Atlantic salmon from northeastern Europe, and use a large collection of wild Atlantic salmon samples exceptional both in the number of populations studied and in the geographic range covered, as it includes subarctic stocks from the Barents and White Seas, Baltic Sea, and freshwater Ladoga and Onega lakes. This region contains the last relatively undisturbed rivers of the Russian Northwest and is characterised by different post-glacial histories and a variety of potentially strong selective pressures, which makes it a unique system to study natural adaptation processes.

1.3.2 Phylogeographic history of northern European Atlantic salmon

The modern basins of the Baltic, White and Barents Seas and Russian landlocked lakes were formed gradually, following the retreat of the Scandinavian Ice Sheet after the Last Glacial Maximum (~20,000 years ago), and were subsequently colonized or re-colonized at different times by salmon from different phylogenetic lineages (Patton et al., 2017; Stroeven et al., 2016). Lake Onega formed first, followed by Lake Ladoga and later by the Baltic Sea (Björck, 1995; Saarnisto & Saarinen, 2001). These basins were colonized by individuals from an eastern freshwater refugium that was isolated from an Atlantic Ocean influence for at least 130,000 years, with some additional gene flow from the eastern Atlantic Ocean into the Baltic Sea (Koljonen et al., 1999; Kudersky et al., 2003; Kuusela et al., 2009; Säisä et al., 2005, Figure1). The Barents and White Seas areas were deglaciated later, and were colonized by salmon from the eastern Barents Sea refugium and the Atlantic Ocean (Asplund et al., 2004; Bourret, Kent, et al., 2013; Tonteri et al., 2005). This history resulted in strong genetic divergence between the Baltic and the Barents & White Seas lineages (e.g. Rougemont & Bernatchez, 2018), as well as in high divergence between lakes Onega and Ladoga due to their prolonged isolation (Ozerov et al., 2010; Tonteri et al., 2007).

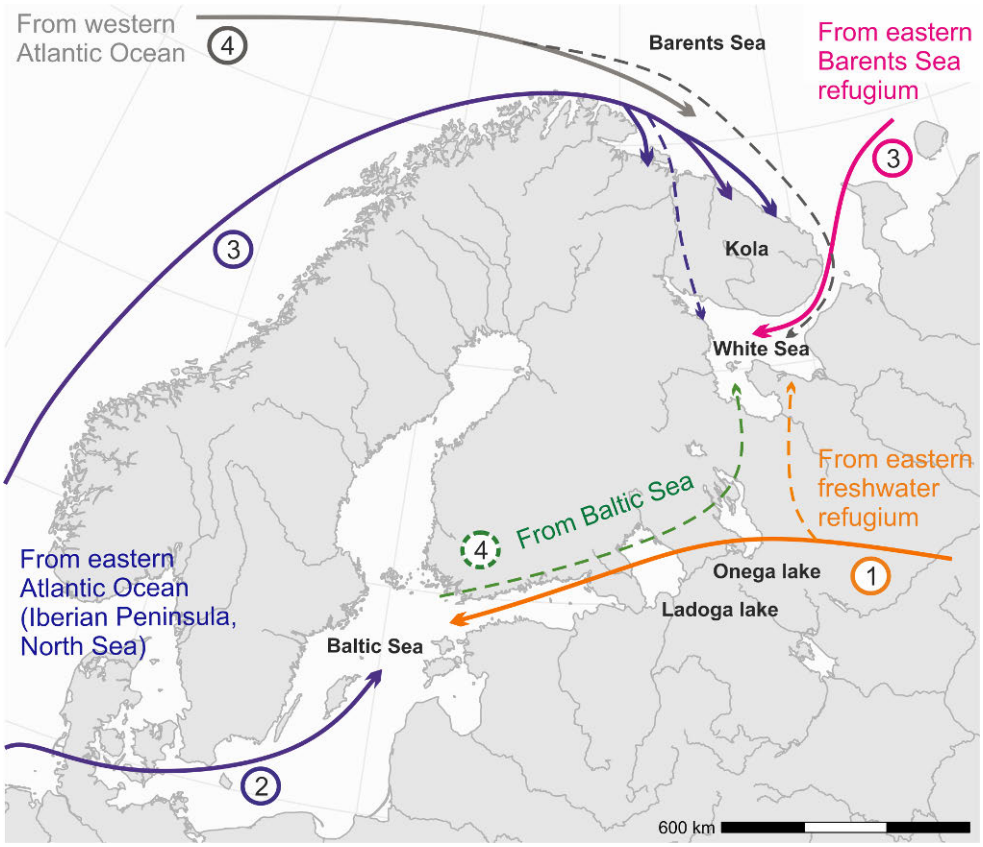


Figure 1. Main directions of post-glacial colonization for Atlantic salmon in northern Europe. Solid arrows indicate consensus colonization directions, while dashed lines show routes hypothesized only by some studies. Colours highlight different areas of salmon origin. The graph does not represent all existing hypotheses, but only those discussed in the thesis. Numbers indicate the approximate order in which colonization events took place.

The genetic structure and phylogeographic history of the Barents & White Seas lineage are particularly interesting, not least because these last relatively undisturbed Atlantic salmon populations in Europe face increasing threats from changing environment and growing anthropogenic pressure (Ozerov et al., 2012). There is a clear phylogeographic division between salmon rivers opening into the Barents Sea and those opening into the White Sea (Asplund et al., 2004; Ozerov et al., 2017; Tonteri et al., 2009; Wennevik et al., 2019). Most studies agree that the western Barents Sea (northern shore of the Kola Peninsula) was colonized by salmon from the eastern Atlantic Ocean (Iberian peninsula, North Sea) (Consuegra et al., 2002; Säisä et al., 2005; Tonteri et al., 2005, 2009), with influences from the

western Atlantic Ocean (North American coast; Asplund et al., 2004; Bradbury et al., 2015; Makhrov et al., 2005).

The phylogeographic structure of the salmon populations in rivers opening into the White Sea is not so clear cut. Many studies identify southern and eastern Kola Peninsula salmon as one group distinct from the rest of the White Sea populations (Asplund et al., 2004; Tonteri et al., 2009), but others separate Kola rivers into multiple separate clusters within the White Sea group (Ozerov et al., 2017). Most studies agree that White Sea populations were re-colonized from a glacial refugium in the Eastern Barents Sea (Asplund et al., 2004; Kazakov & Titov, 1991; Tonteri et al., 2005, 2009). However, some studies suggest that additional refugia, such as eastern Atlantic Ocean (Asplund et al., 2004; Makhrov et al., 2005), western Atlantic Ocean (Makhrov et al., 2005), and the Baltic Sea (Kazakov & Titov, 1991; Makhrov et al., 2005) might have participated in recolonization of the White Sea.

Altogether, while there is a consensus about major directions of salmon recolonization in the region, additional research is required to reach a consensus about finer genetic structure and phylogeographic history of the White Sea salmon.

1.3.3 Threat posed by *Gyrodactylus salaris*

Northeastern European salmon populations vary in a number of abiotic and biotic traits, but one of the most striking differences is the response to a potentially very dangerous parasite, *Gyrodactylus salaris*. *G. salaris* is a small monogenean flatworm that feeds on the skin and mucus of salmon while the fish is in the freshwater habitat (river or lake).

Baltic lineage Atlantic salmon naturally coexists with the parasite: low-level infections are observed in only 1% of fish from lakes Onega and Ladoga, and in 20% of fish from rivers draining to the Baltic Sea (Kuusela et al., 2009); additionally, the parasite has little or no negative effect on the infected fish (Bakke et al., 1990). These low levels of susceptibility are thought to be a result of a long co-evolutionary history dating back to when salmon and *G. salaris* co-occurred in the eastern freshwater refugium (~130,000 years ago) (Kudersky et al., 2003; Kuusela et al., 2007, 2009). Salmon from rivers draining to the Atlantic Ocean, including the Barents and White Seas, are not exposed to *G. salaris* naturally. However, if the parasite is introduced, the fish mortality rates reach up to 95% (Johnsen & Jensen, 1991). In 1970s, the parasite was accidentally introduced to numerous populations in Norway and to one location in the White Sea via stocking of infected Baltic salmon, which caused the subsequent annihilation of these stocks (rev. in Harris et al., 2011).

Despite the potentially devastating effect of *G. salaris* and the everexisting danger of further spread of the parasite to the susceptible areas (NASCO, 2018), the

biological and genetic mechanisms of the response of *S. salar* to *G. salaris* are not fully understood. One study has estimated the heritability of survival after *G. salaris* infection in challenge-tests ($h^2 = 0.32 \pm 0.1$ on the liability scale) (Salte et al., 2010), a few studies have examined differential gene expression profiles in challenge experiments (Gilbey et al., 2003; Kania et al., 2010; Matejusová et al., 2006), one study using 39 microsatellites markers identified several quantitative trait loci (QTL) influencing parasite resistance (Gilbey et al., 2006), and one study looked for an association between immune-relevant microsatellites under elevated selection pressure in northern Europe and *G. salaris*-induced mortality rates, but the authors did not find a significant correlation (Tonteri et al., 2010). While there is a continuous discussion about possibilities for selective breeding for resistance and about implications of ongoing natural selection in affected Norwegian rivers (Karlsson et al., 2020), the exact genomic basis of the differential immune response exhibited by salmon of different origin when exposed to *G. salaris* remains unclear.

Parasites are one of the major selective forces acting on host populations (Eizaguirre & Lenz, 2010; Wilson et al., 2019) and driving divergent evolution and adaptive radiation (Karvonen & Seehausen, 2012). Due to the importance of Atlantic salmon in aquaculture and commercial fisheries, the molecular basis of its response to various hatchery pathogens has been extensively studied (e.g. Holm et al., 2017; Moen et al., 2007, 2015; Reyes-Lopez et al., 2015). Studies of the genetic basis of co-adaptation between wild salmon and its pathogens, on the other hand, are scarce. However, thanks to the strength of selective pressure that *G. salaris* exerts, looking for signals of parasite-mediated directional selection at the genome scale is a promising approach to identify the genetic basis of resistance/tolerance to this parasite.

1.3.4 Regional variation in other environmental variables

Rivers, lakes and marine environments inhabited by Atlantic salmon populations in northern Europe also vary in a number of abiotic and biotic features that potentially exert strong selection on populations.

Most obviously, the salinity of the basins that salmon use for their feeding migrations differs drastically, ranging from the truly marine Barents and White Seas, to the brackish Baltic Sea, and the freshwater landlocked lakes. This is reflected in variations in fish physiology and the smolting process (McCormick et al., 2019; Nilsen et al., 2003, 2008), and differences in diet (Jacobsen & Jacobsen, 2001; Salminen et al., 2001). Water temperature regimes also vary among the freshwater and marine environments used by northeastern European salmon; and given that temperature is known to affect, among other things, metabolism and

development (Brown et al., 2004) and overall parasite diversity (Adlard et al., 2015), it is also likely to be a strong selective force. In addition, the Barents and White Sea salmon exhibit regional variations in average smolt age, age of sexual maturity, and timing of the return spawning run (Ponomareva, 2007; Potutkin et al., 2007).

Wild salmon populations are expected to be under multiple simultaneous selection pressures, and it may be challenging to single out the genomic basis of the response to one particular selective element. Northeastern European salmon present a great opportunity to overcome this obstacle and isolate the genomic basis of differential response to *G. salaris*, by comparing signals of selection present in populations from different parts of the region and belonging to different lineages. From a broader perspective, the same approach allows to explore whether the emerging patterns are consistent with parallel or divergent evolution and to explore how universal the genetic architecture behind adaptive diversification is. It would also facilitate the development of genetic resources for this species, and increase our understanding of the genetic-ecological interactions vital for stock management and balance between wild and aquaculture salmon.

2 Aims of the Thesis

This dissertation is organized around four different aims:

1. Identifying the genomic basis of the remarkable adaptation gradient exhibited by Atlantic salmon populations to the parasite *Gyrodactylus salaris*, while accounting for the confounding effects of genetic drift, phylogeographic history and correlated environmental factors (Chapters I and II)
2. Identifying the genomic signatures of directional selection exerted by water salinity, temperature, and other abiotic factors varying among the studied populations (Chapters I and III)
3. Examining the repeatability of genomic searches for signals of adaptive divergence: loci emerging as locally adaptive across different geographic scales, and across independent studies – are they the same? (Chapter III)
4. Characterizing population genetic structure in relation to existing hypotheses of phylogeographic history of salmon populations from northeastern Europe (Chapters I, II, and III)

3 Materials and Methods

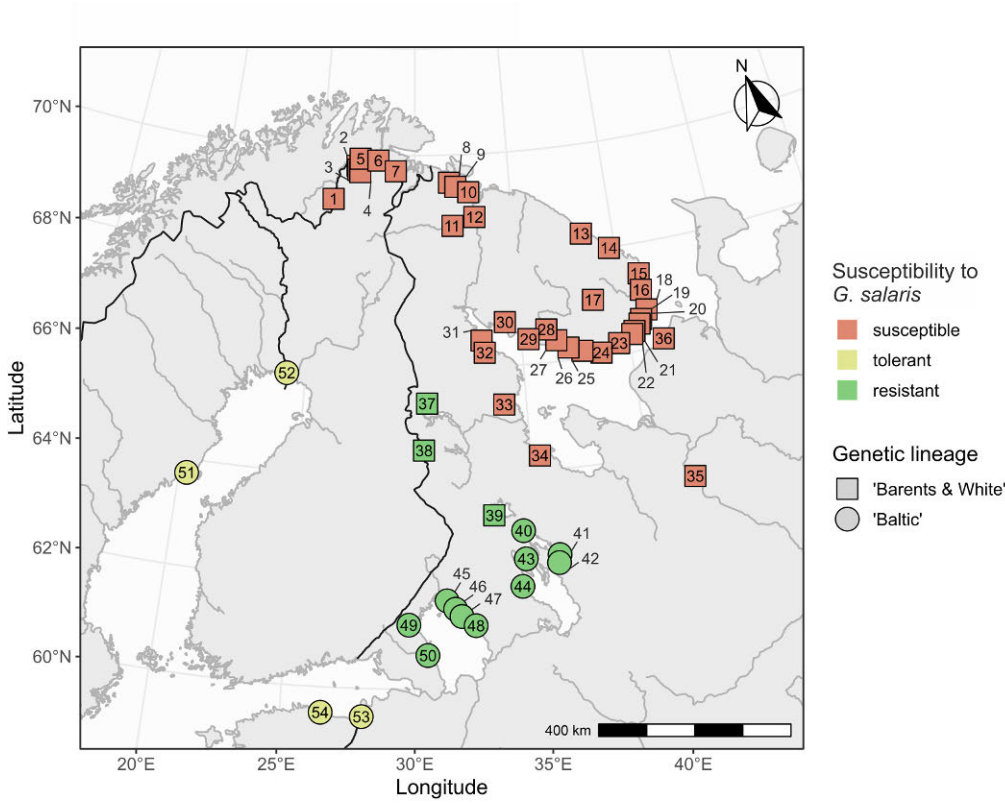
3.1 Sampling and molecular techniques

Samples used throughout this thesis were primarily fin clips from wild juvenile salmon caught by electrofishing between 1996 and 2008. The Näätämö river and Teno river system samples were scales collected from adult fish during spawning migration (Aykanat et al., 2015; Pritchard et al., 2016). Altogether, 54 sampling locations and more than 2700 individuals were analysed across all Chapters (Figure 2). Throughout the thesis, I use the terms 'population', 'river', and 'sampling location' interchangeably.

Specifics of the genomic DNA extraction and sample preparation for genotyping and allelotyping are described in detail in the respective chapters. In **Chapter I**, 472 individual samples from 12 salmon populations were genotyped with the Illumina iSelect SNP array that assayed 6176 SNP markers (as in Bourret et al., 2013). This allowed moderate coverage of 1 SNP per every 500 kb of the genome on average. In **Chapters II** and **III**, I pooled equal amounts of DNA from all available individuals on a per-river basis and applied an allelotyping approach to estimate population allele frequencies for each SNP, which enabled a substantial increase in the number of studied populations. Population pools were allelotyped using a custom 220,000 SNP Affymetrix Axiom array (unpublished), which significantly improved the density of SNP coverage, with 1 SNP per every 15kb of salmon genome on average.

The quality control and data pre-processing steps are detailed in respective chapters. Briefly, for individual genotype data in **Chapter I** these steps included eliminating SNPs with > 10% missing data and with minor allele frequency (MAF) < 0.05 across all populations. For the allelotyping data in **Chapters II** and **III** I corrected the relative frequency of B allele to account for different relative intensities of the A and B allele probe signals among different SNPs, controlled for high noise among pooling replicates, and removed SNPs with technical genotyping problems and SNPs with MAF across all populations < 0.05.

Unless stated otherwise, all data formatting, quality control, and statistical analyses were performed either in R environment (R Core Team, 2019) or using Unix text processing command line tools.



Usage of sampling locations per chapter

1.	○●●	Teno Inarijoki (Teno2)	20.	○●●	Sosnovka	39.	○○●	Luzhma
2.	○○●	Teno Tsarsjoki	21.	○●●	Babya	40.	○●●	Kumsa
3.	○○●	Teno Kevojoki	22.	○●●	Lihodeevka	41.	●●●	Pyalma
4.	○○●	Teno Utsjoki	23.	○●●	Ust' Pyalka	42.	○●●	Tuba
5.	○●○	Teno Mainstem (Teno1)	24.	○●●	Strelna	43.	○●●	Lizma
6.	○○●	Teno Pulmankijoki	25.	○●●	Chavanga	44.	○●●	Shuya
7.	○●●	Näätämö	26.	○●●	Indera	45.	●●●	Sysky
8.	○●●	Titovka	27.	○●●	Varzuga	46.	○●●	Uuksa
9.	○●●	Zapadnaya Litsa	28.	●●●	Yapoma	47.	○●●	Tulema
10.	○●●	Ura	29.	○●●	Olenitsa	48.	○●●	Vidlitsa
11.	●●●	Tuloma	30.	○●●	Umba	49.	○●●	Hiitola
12.	○●●	Kola	31.	○●●	Nilma	50.	○●●	Taipale
13.	○●●	Drozdovka	32.	○●●	Pulonga	51.	●○●	Vindel
14.	○●●	Yokanga	33.	●●●	Pongoma	52.	●○○	Tornio
15.	○●●	Kachkovka	34.	●●●	Suma	53.	○○●	Narva
16.	○●●	Ponoi	35.	●●●	S. Dvina Emtsa	54.	●○●	Kunda
17.	●●●	Ponoi Lebyazia	36.	○●●	Megra			
18.	○●●	Danilovka	37.	●○●	Pistojoki			
29.	○●●	Sneznitsa	38.	○○●	Kamennaya			

Figure 2. Map of the sampling locations. Colours represent different levels of susceptibility to *G. salaris*, while shapes reflect the major phylogeographic lineages. The three small dots next to location name represent Chapters I to III, and usage of sampling locations in a given chapter is shown with grey colour.

3.2 Characterisation of population genetic structure

Chapter I aims at identifying the genetic basis of differences in parasite response in salmon from three distinct geographic regions: Atlantic Ocean (the Barents and White Seas), the Baltic Sea, and freshwater lakes. To evaluate the amount of genetic structure present at the population and region levels I used two traditional methods implemented in the *Arlequin3.5* software (Excoffier & Lischer, 2010). First, I estimated the pairwise F_{ST} (Weir & Cockerham, 1984), to evaluate the genetic differentiation among all studied populations. Further, I used the differential hierarchical analysis of molecular variance (AMOVA) (Excoffier et al., 1992) to confirm that overall population genetic structure followed geographic regions.

In **Chapters II** and **III** I explored regional population genetic structure using principal component analysis (PCA) of the population allele frequencies, implemented in the '*stats*' package and the *PCAdapt* 4.1.0. (Luu et al., 2017) package within the R-environment (R Core Team, 2019). PCA can be applied to pooled data, and the top principal components (PCs) are viewed as continuous axes of variation that reflect genetic variation due to ancestry in the sample. Contrary to AMOVA, PCA does not require any grouping of populations prior to the analysis, and this approach is widely used to identify and adjust for relatedness among sample individuals and/or populations.

In **Chapter III** I further investigated the phylogeographic history of the northeastern European salmon populations, using the *TreeMix* software (Pickrell & Pritchard, 2012). TreeMix infers patterns of population splits and mixing events in the history of the given population set, and represents the ancestral relationship between populations as a bifurcating graph with cross-connections indicating mixing events.

3.3 Genomic signatures of directional selection and adaptation

The various approaches that were used in this dissertation to detect signals of directional selection are summarised in Table 1. The approaches differ in the way they detect selection signals. They include methods based on detecting reduced genetic diversity, on detecting elevated population differentiation, or testing for environmental association of the allele frequencies. Those approaches also differ by whether or not SNP position along the genome is taken into consideration.

Table 1. Overview of the methods used in the thesis to detect signatures of directional selection

Method, software & reference	Test statistics	Method details	Chapters in which is used
METHODS BASED ON REDUCED DIVERSITY			
Kernel-smoothing moving average (as in Hohenlohe et al. 2010)	Reduced denetic diversity (H_E)	Identifies groups of adjacent markers showing selection signature	I
METHODS BASED ON INCREASED DIVERGENCE			
Kernel-smoothing moving average (as in Hohenlohe et al. 2010)	Elevated F_{ST}	Identifies groups of adjacent markers showing selection signature	I
Arlequin 3.5.0 (Excoffier & Lischer, 2010)	F_{CT}, p-value	Accounts for hierarchical population structure, user defines population groups	I
Bayenv 2.0 (Günther & Coop, 2013) BayPass 2.1 (Gautier 2015)	Absolute X^TX	Account for neutral population structure by computing variance-covariance matrix of population allele frequencies	II III
BayeScan 2.1 (Foll & Gaggiotti, 2008) BayScEnv 1.1 (de Villemereuil & Gaggiotti, 2015)	Alpha parameter, qvalue	Estimate the posterior probability of a given locus to be under selection by defining two alternative models, one that includes the effect of selection and another that excludes it	II III
PCAdapt (Luu et al. 2017)	Mahalanobis distances, p-value	Performs PCA and tests for outliers based on the correlations between genetic variation and the first K principal components	III
ENVIRONMENTAL ASSOCIATION			
LFMM (Frichot et al., 2013)	z-score, p-value	Latent factor mixed model. Correlations between environmental variables and allele frequencies are estimated simultaneously with inferring population structure	I, III
BayScEnv 1.1 (de Villemereuil & Gaggiotti, 2015)	g, qvalue	Tests if values of F_{ST} increase with environmental differentiation, model allows to compute F_{ST} for each population	III
BayPass 2.1 (Gautier 2015)	Absolute Pearson's correlation coefficient, r	Annotates footprints of selection by quantifying their association with population-specific covariates	III

In **Chapter I** I applied the genome scan approach based on hitchhiking mapping, looking for groups of adjacent SNPs that are characterised by reduced diversity within and increased divergence between studied populations. As opposed to methods focusing on single SNP outliers which are used further in **Chapters I – III**, this approach takes into account the physical location of SNPs along the genome and allows to identify groups of adjacent markers that deviate from the chromosome-wide average levels of the statistic being examined. I used locus-specific expected heterozygosity (H_E), calculated using *PowerMarker 3.25* (Liu & Muse, 2005), as a measure of genetic diversity (GD); and locus-specific F_{ST} (Weir & Cockerham, 1984) calculated using *Arlequin 3.5*. as a measure of inter-population divergence. Smoothed chromosome-wide distributions of GD and F_{ST} were generated using the "*locpoly*" function included in the *KernSmooth* R-package (Wand & Jones, 1995). Briefly, the contribution of the F_{ST} or GD statistics to the kernel-smoothed average was estimated using local polynomials and a bandwidth of the half-length of the estimated linkage disequilibrium in the dataset. 10,000 permutations were used to test whether the observed smoothed curves were significantly ($P \leq 0.01$) higher or lower than expected by chance within a local genome region, and such regions were considered to be under selection.

Throughout **Chapters I – III** several 'single-locus outlier' methods, based on detecting SNPs that were extremely differentiated between populations compared to the rest of genome, were used.

Arlequin 3.5 was used in **Chapter I** to detect signals of selection on a regional scale, by estimating locus-specific coefficient of differentiation among groups of populations (F_{CT}). The hierarchical island-model implemented in the program leads to a reduction in the number of potential false positives and is advantageous when some of the sampled populations share recent common ancestry. Prior to analysis the user defines the population groups, and the software uses coalescent simulations to estimate the p-values of locus-specific F_{CT}-statistics, conditioned on observed levels of heterozygosity.

Bayenv 2.0 (Günther & Coop, 2013), used in **Chapter II**, calculates a per-SNP population differentiation statistics X^TX, which accounts for underlying population structure. In contrast to the user-defined hierarchical population relationships in *Arlequin 3.5*, *Bayenv 2.0* models neutral population structure directly from allele frequencies by estimating a genome-wide population covariance matrix. Similar to the well-known F_{ST}, SNPs with elevated X^TX are considered to be candidates for directional local selection. A similar approach was chosen in **Chapter III**, but I estimated X^TX using a different software, *BayPass 2.1* (Gautier, 2015), which is more user-friendly and may provide improved estimation accuracy of the population covariance matrix.

BayeScan 2.1 (Foll & Gaggiotti, 2008), as well as its extension, *BayScEnv 1.1* (de Villemereuil & Gaggiotti, 2015) were used in **Chapters II** and **III** respectively, as alternative approaches to outlier detection. These methods use a Bayesian approach assuming an island model of migration to separate F_{ST} coefficients into a population-specific component, shared by all loci and a locus-specific component, shared by all populations. Departure from neutrality at a given locus is assumed when the locus-specific component is necessary to explain the observed pattern of diversity. This leads to two alternative models for each locus, one that includes effect of selection, and one that does not. The software then implements a MCMC algorithm to estimate the posterior probability of these models. The method has been suggested to be robust when dealing with complex demographic scenarios for neutral genetic differentiation (Foll & Gaggiotti, 2008).

PCAdapt 4.1.0 (Luu et al., 2017), used in **Chapter III**, provided yet another take on identifying loci under selection. PCAdapt assumes that markers excessively related to population structure are linked to candidate locally adaptive loci. First, a PCA is performed on the centered and scaled genotype matrix. Second, test statistics and p-values are computed based on the correlations between SNPs and the first K principal components (PCs), though it is also possible to perform one genome scan for each principal component. The approach can handle data sets containing admixed individuals, does not require *a priori* grouping of the samples, and is most powerful in scenarios of population divergence and range expansion.

Another widely applied type of genome-wide approaches looking for signatures of directional selection is environmental association methods, which test the association between allele frequencies and environmental variables.

The *LFMM* (Frichot et al., 2013) method was used in **Chapters I** and **III**. Latent factor mixed models are statistical regression models to test associations between allele frequencies and variables representing environmental or phenotypic traits, and include unobserved variables, called latent factors, that correct the model for confounding effects due to population structure and other hidden causes. LFMM software is effective in accounting for random effects due to population history and isolation-by-distance patterns, and does not require a control data set of *a priori* neutral loci.

BayScEnv 1.1 and *BayPass 2.1*, used in **Chapter III**, were already mentioned above. Apart from performing 'outlier analysis', these Bayesian approaches also allow identification of SNP-specific effects driven by environmental variables. BayPass (as well as LFMM) takes into account allele frequency correlations across populations to minimize the possible spatial correlations in allele frequencies. BayeScEnv, on the other hand, assumes that all populations are independent (de Villemereuil & Gaggiotti, 2015).

The environmental variables of interest were estimates of *G. salaris* - induced mortality rate, surface salinity of the basin, mean surface water temperature, and population coordinates (**Chapter I**); as well as upstream catchment area (**Chapter III**).

These various methods outperform each other under different scenarios of isolation by distance, hierarchical population structures, and situations when an environmental selective gradient is confounded with population structure.

3.4 Methodological approaches to assess the validity of detected loci

I used three major techniques to strengthen the validity of detected candidate loci and genomic regions.

The first approach is to perform independent tests for selection and/or environmental association using various statistical methods that utilise and bring out different aspects of the data, and to subsequently accept only overlapping results. I have used this approach in all three chapters (see **Chapters I-III**, and section 3.3. for details on the used statistical approaches).

The second approach is to use the same statistical method to test multiple population groups or population comparisons that are similar in terms of underlying traits or structure composition; and to subsequently compare the results for overlap. For example, in **Chapter I** I looked for genomic regions with elevated F_{ST} between populations from the Barents Sea and freshwater lakes Onega and Ladoga (design 2, see original publication). Six similar, but independent pair-wise comparisons were made, in which one population belonged to the Barents Sea, and the other to the lake basin. A given genomic region was considered to be under selection only when elevated F_{ST} was observed in at least two comparisons out of six, and only if both Ladoga and Onega populations were represented in the significant comparisons. This reduced the chance that observed signals of selections were due to high genetic drift in a single landlocked population. I applied a similar logic also in **Chapter III**, where I performed the identical analyses independently for populations from different geographic regions, and candidate locally-selected genome regions shared by at least two out of three geographic regions were discussed further.

Lastly, in order to disentangle the effects of correlated environmental traits, population structure, and neutral genomic processes, I used the following logic: I performed several tests with populations from varying and often contrasting environments and phylogeographic histories, and then selected loci detected as outliers in some of the performed tests, but absent in the others. For example, in **Chapter I** (design 4) I tried to disentangle the selective pressure due to the presence of *G. salaris* parasite from the pressure due to varying salinity levels. I identified highly

differentiated SNPs in the pairwise comparisons between three groups of populations: Atlantic Ocean (Barents and White Seas) vs. landlocked lakes, Atlantic Ocean vs. Baltic Sea, Baltic Sea vs. landlocked lakes. The Atlantic group is susceptible to *G. salaris*, while populations from the Baltic Sea and lakes can, to some extent, tolerate the pathogen. Therefore, only outliers common for both Atlantic Ocean vs. lakes and Atlantic Ocean vs. Baltic Sea tests, but not found in the Baltic Sea vs. lakes comparison, were considered to be the result of parasite-driven selection occurring in landlocked and Baltic populations (rather than due to salinity). The same logic was applied in **Chapter II** that also aimed at identifying signatures of parasite-driven selection. The final set of candidate genes was obtained by identifying genes that were detected by both Atlantic Ocean vs. Ladoga lake and Atlantic Ocean vs. Onega lake comparisons, but that were not present among outliers in the Ladoga lake vs. Onega lake test. Since there has been a prolonged isolation of the landlocked lakes from each other, this approach allowed me to exclude genomic regions that are likely to exhibit elevated levels of differentiation due to genetic drift rather than directional selection.

3.5 Functional annotation of detected loci and related analyses

Over the duration of this thesis the genomic resources available for Atlantic salmon have improved drastically, which was reflected in the methods used for functional annotation of the detected candidate loci and related analyses.

In **Chapter I**, SNPs were annotated to specific gene ontology (GO) terms by performing *tblastx* and *blastx* searches (Camacho et al., 2009) of SNP flanking regions against the nucleotide and protein NCBI databases (www.ncbi.nlm.nih.gov), and consequent retrieval of the corresponding human GO identifiers from the GO database (www.geneontology.org, all resources accessed on 3.04.2012).

By the time of **Chapters II** and **III**, the salmon genome build (ICSASG_v2, the latest as of 09.2020 RefSeq accession number: GCF_000233375.1) had become publicly available, along with the NCBI *Salmo salar* Annotation Release 100 (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Salmo_salar/100/). I mapped SNPs to the respective genes with the help of the *closest* function in *BEDTools 2.29.0* software (Quinlan & Hall, 2010); a gene was assigned with a SNP if the SNP's position in the genome fell within the gene margins (**Chapter II**), or if it was either overlapping (SNP within gene margins) or the closest downstream protein coding gene on either strand (**Chapter III**). To get gene and GO annotations for the ICSASG_v2 build I have used the *Ssa.RefSeq.db* R package (Grammes, 2017) (**Chapter II**) or the files of the NCBI *Salmo salar* Annotation Release 100 directly (**Chapter III**).

In **Chapter III** I had access to a dataset of individually genotyped fish from the Teno river system (combined data from Barson et al., 2015; Pritchard et al., 2016,

2018) and thus was able to additionally assess the possibility that the detected SNPs of interest occurred within the same haploblock, defined as a physically contiguous set of SNPs exceeding a specified linkage disequilibrium threshold. I used 883 individually genotyped fish to infer haploblocks for all SNPs used in Chapter III, using *PLINK 1.96* (Chang et al., 2015) and following the procedure detailed in Pritchard *et al.* (2018). Each haploblock that included one of the retained SNPs was annotated with the overlapping NCBI coding genes or documented for Atlantic salmon structural variants (Bertolotti et al., 2020) using the *intersect* function of *BEDTools 2.29.0.*

To investigate whether the detected sets of candidate SNPs (**Chapter I**) and genes (**Chapter II**) were significantly enriched or depleted for particular GO terms, I performed a GO enrichment analysis, using the *Cytoscape 2.8.3*. (Shannon et al., 2003) software and its plugin *BiNGO 2.44* (Maere et al., 2005) in **Chapter I**, and the *weight01* algorithm in the *topGO* package in R (Alexa & Rahnenfuhrer, 2016) in **Chapter II**. For the salinity- and parasite-related SNPs identified in **Chapter I,** I also assessed the functional relatedness of GO terms included in both lists of outliers, with the help of another *Cytoscape 2.8.3.* plugin, *ClueGo1.7.1* (Bindea et al., 2009).

4 Results and Discussion

4.1 Population structure and phylogeographic history of northern European Atlantic salmon

Both 7K (**Chapter I**) and 220K (**Chapters II** and **III**) SNP chips inferred a similar population genetic structure, which is largely in concordance with phylogeographic patterns identified previously.

First, patterns of genetic diversity and divergence were consistent across the different analyses. The diversity was the lowest, and inter-population divergence the highest, for the Onega and Ladoga lakes, followed by the Baltic Sea, the White Sea and the Barents Sea. This observation is in line with the suggested patterns of prolonged population isolation of Onega and Ladoga (Ozerov et al., 2010; Tonteri et al., 2007), and higher incidence of contemporary migration among populations in the Barents Sea (Ozerov et al., 2012).

Second, in **Chapters I** and **III**, I observed a clear division into two major lineages: the 'Baltic' clade, including salmon from the Baltic Sea and the freshwater lakes Onega and Ladoga, and the 'Barents-White' clade, including fish from the Barents and White Seas (Figure 3). This division was repeatedly suggested before and is a current consensus (Rougemont & Bernatchez, 2018; Säisä et al., 2005; Tonteri et al., 2005).

The freshwater Karelian populations that hydrologically belong to the White Sea basin (rivers Pisto, Luzhma, Kamennaya) also phylogeographically clustered with the 'Barents-White' clade (**Chapter III**), as was suggested previously (Bourret et al., 2013; Tonteri et al., 2005). However, there is also some previous evidence for the genetic closeness of Karelian stocks with the freshwater Onega and Ladoga lakes from the 'Baltic' cluster (Ozerov et al., 2013). The TreeMix analysis did not suggest gene flow to Karelian landlocked stocks from Onega and Ladoga but did infer possible migration from the Baltic Sea populations (Figure 3). Altogether, my results contribute to the existing theory of Karelian populations originating later in time than populations in the big freshwater lakes and having been re-colonised from the eastern Barents refugia, similarly to the anadromous White Sea populations (Lumme et al.,

2015); but they also add evidence for the possible influence of the 'Baltic' lineage salmon on the Karelian populations.

Tree Mix showed no inference of gene flow between the Baltic salmon and the White Sea fish, so I have no support for the previously suggested phylogeographic connection between these two regions (Kazakov & Titov, 1991; Makhrov *et al.*, 2005). There was, however, an unexpected evidence for gene flow events from the Baltic lineage salmon to the Teno river system (Figure 3); the robustness of this result and its implications for phylogeographic history of the region should be explored by future studies.

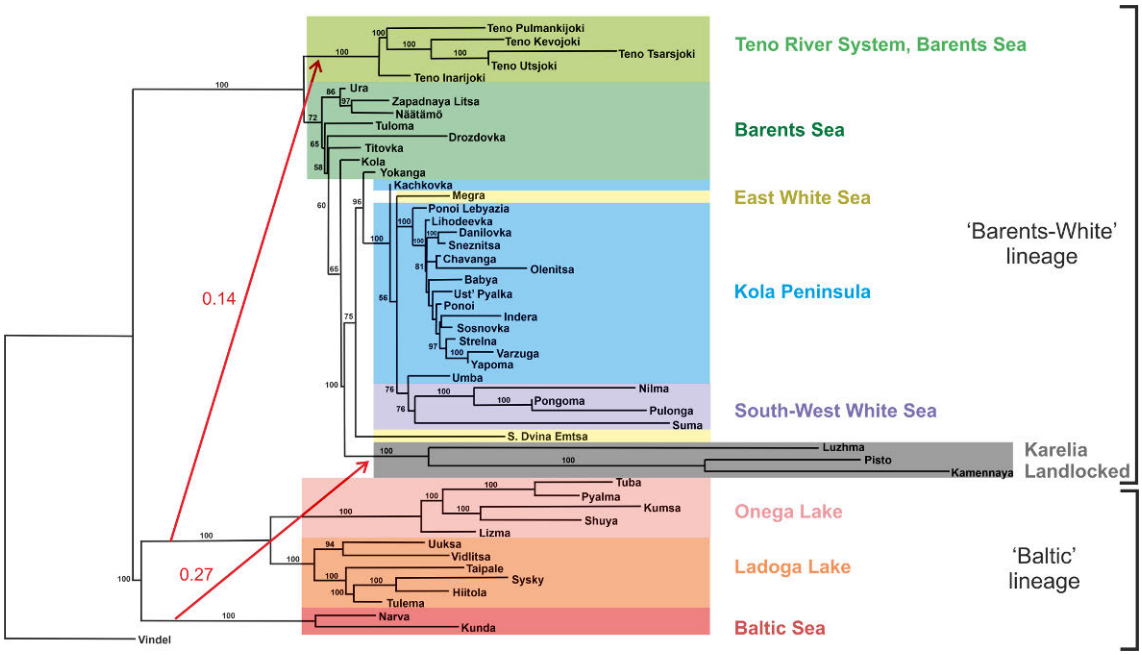


Figure 3. Population tree inferred by TreeMix. Block colours show geographic regional groupings. Branch numbers indicate nodal support based on 100 bootstrap trees without migration, Vindel population is used as an outgroup. Red arrows show long-distance gene flow events inferred by TreeMix, with numbers indicating migration weight.

Third, in **Chapter III** I looked in detail into the genetic structure of the Barents and White Sea populations. As found in previous studies, I observed a clear genetic transition between the Barents Sea and the White Sea, aligned along the geographical border separating these two basins. The White Sea populations were divided into a cluster of Kola peninsula rivers, and rivers opening into the South-West and East White Sea (Figure 2 in Chapter III). I did not observe further sub-structuring of the 'Kola' cluster, suggested recently by Ozerov et al. (2017), possibly due to the fact that the Ponoi river system, forming a separate cluster in their study, was represented

by only two tributaries in the current study. The SW and E White populations were strongly differentiated both from Kola cluster and from each other, and it is likely that isolation, strong genetic drift, and anthropogenic influence drove the identified structure. However, sampling additional populations from the east White Sea and east Barents Sea could increase the robustness of the identified population structure. The observed genetic transition between Barents and White Sea salmon is in concordance with the concept of secondary contact between eastern and western salmon lineages colonizing this area. In **Chapter III** I investigated the genetic architecture underlying this regional division, and identified several genomic regions of elevated differentiation between the two salmon groups. Such differentiated segments may contain locally adaptive loci, or be a consequence of other barriers of gene flow, for example purifying selection acting on areas of reduced recombination, e.g. inversions (Lotterhos, 2019; Wu, 2001). Overall, this is an interesting topic, and continued research on the nature of gene flow barriers between Barents and White Sea Atlantic salmon could improve our understanding of the recent evolutionary history of the region.

4.2 Evidence for parasite-driven natural selection

A primary goal of this thesis was to investigate the genomic basis of differences in susceptibility to the parasite *Gyrodactylus salaris* that are observed in northern European Atlantic salmon populations. A number of single SNP outliers and three genomic regions potentially affected by *G. salaris*-mediated selection were identified in **Chapter I**, and 57 candidate genes were detected in **Chapter II**. Altogether these loci were distributed among 25 out of 29 Atlantic salmon chromosomes and were identified as strong candidates based on the combined outcomes of several analyses.

Functional network analysis of SNP outliers detected in **Chapter I** indicated their involvement in two main functional groups: translation initiation and long-chain fatty-acyl-CoA metabolism. The candidate translation initiation factors are involved in several immune processes, including T-lymphocyte activation. However, they also control a variety of stress responses, including response to pathogen presence, osmotic and temperature stresses, and nutrient starvation (rev. in Chapter I). These translation initiation pathways are quite conserved among many distant taxa, thus their precise role in the studied Atlantic salmon populations is still open for discussion. The second suggested functional group was involved in fatty acids synthesis and elongation. Substantial evidence shows that fatty acids play a crucial role in the regulation of inflammation and in the balance of cytokines secretion, thus modulating innate immune response in vertebrates (rev. in Chapter I).

Genes detected in **Chapter II** were enriched for several GO terms, including lymph node development, response to virus, microtubule organization, and phospholipase-related activity. Among the particularly interesting genes are three copies of *mx* (myxovirus)-like gene (Figure 3 in Chapter II), as *mx* genes are part of the interferon-mediated innate immune response and are activated in response to a variety of viruses (Mitchell et al., 2013). Other promising candidates are T-cell leukemia homeobox protein 1 (*TLX1*)-like gene, involved in the development of teleost spleen and the maturation of lymphocytes; and nuclear receptor ROR-alpha-like gene (*RORα)*, which has diverse functions including regulation of fatty acids metabolism and regulation of inflammation cytokines (rev. in Chapter II). Several loci potentially involved in the formation of focal and cell-cell adhesions and cell signaling were also detected. These include talin-like locus (*LOC106561152*), talin being crucial during phagocytosis and for adhesion of natural killer cells and T-lymphocytes to the extra-cellular matrix and to target cells; two loci with phospholipase activity (*LOC106608623* and *LOC106588883*), phospholipases being involved in talin stabilisation, signal transduction in leukocytes, and other inflammation processes; and sphingomyelin phosphodiesterase 3-like gene (*LOC106560916*), potentially regulating a crucial part of the innate immune system, the Toll-signalling pathway (rev. in Chapter II). Taken together, the results of **Chapter II** present evidence that the set of candidate genes driven by adaptation to *G. salaris* is involved in cell-signalling and regulation during both innate and adaptive immune responses.

There was little overlap between **Chapters I** and **II** in terms of the candidate genes that were detected: only one candidate gene identified in Chapter II fell within a candidate region from Chapter I. However, the overall functional patterns of detected candidate loci were greatly similar: there was evidence for processes involved in innate immunity, such as fatty acids metabolism (Chapter I), regulation of cytokines and inflammation (Chapters I and II); and in adaptive immunity, such as lymphocyte maturation and T-cell activation (Chapters I and II). It is notable that differential cytokine production during the initial stage of the response to *G. salaris* presence is the main difference between susceptible and tolerant Atlantic salmon in infection experiments (Kania et al., 2010), suggesting that regulation of the first stage of innate immune response may be crucial in controlling parasite abundance and fish survival. Experimental research on controlled *G. salaris* infection has been limited, and future studies are needed to explore the relative role of innate and acquired immune systems in salmon response to the parasite.

Many studies support the concept of multiple loci being involved in immunity and response to pathogens in Atlantic salmon (Andresen et al., 2019; Matejusová et al., 2006; Moore et al., 2017; Tadiso et al., 2011). Taken together, my findings

suggest that *G. salaris* susceptibility and resistance/tolerance in Atlantic salmon also has a complex and polygenic basis, and that the observed differences in parasite susceptibility levels are mediated by natural selection acting on the regulatory mechanisms of both innate and adaptive immune systems.

4.3 Local adaptation to other environmental traits

In **Chapter I** I aimed to separate signals of positive selection in response to the *G. salaris* parasite from responses to other potential selective forces, such as salinity and summer temperature of the water basins that fish migrate to, and overall geographic location. In **Chapter III** these likely agents of local adaptation were complemented by the catchment area of the home rivers, used as a proxy for expected river flow at the sampling site.

Three genomic regions potentially affected by salinity-induced selection were identified in **Chapter I** (Figure 4 in Chapter I) along with several single SNP outliers. One of the identified 'salinity'-mediated regions overlaps with previously identified QTL that encompasses a calcium-sensor receptor, *CaSR*, involved in osmoregulation in several salmonid fishes including Atlantic salmon (Norman et al., 2012). Functional network analysis showed that the GO terms linked with 'salinity' outlier SNPs were associated with renal absorption and protein kinase B signalling, which, among other processes, is activated by cellular stress including hyperosmolarity (Konishi et al., 1997). Several molecular components are involved in salinity acclimation in salmonids, including Na^+/K^+ ATP-ase, cortisol and thyroid hormones, agents modulating intracellular calcium levels, and multiple other proteins (rev. in Chapter I). The salinity-tolerance mechanisms are quite complex, and loci identified in Chapter I open discussion on a potential role of salinity-mediated stress signalling in northern European Atlantic salmon from marine, brackish and freshwater environment.

Only 2 outlier SNPs were uniquely associated with water temperature, and only 5 with latitude of the sampling location. These are surprisingly small numbers, especially since temperature is regularly associated with population genetic diversity in salmonids (rev. by Olsen et al., 2010). In turn, a southward latitudal gradient and an associated rise in water temperature are often accompanied by an increase of pathogen biodiversity, so one could expect a correlated variation of genetic diversity of immune-related loci (Dionne et al., 2007). It is possible however, that mean sea/lake surface temperatures do not act as a selective pressure strong enough to affect survival of adult salmon in studied regions, and natal river temperatures could be used instead in future research.

In **Chapter III** I tested for associations between genomic variation and upstream catchment area in three regional groups of salmon, given that river landscape plays a prominent role in salmon survival (Armstrong et al., 2003), and catchment area in particular was shown to be linked to allelic variation in Teno salmon (Pritchard et al., 2018). Several candidate catchment-associated haploblocks were identified, but only one was shared between two out of three regional groups examined, Kola and Teno. This candidate haploblock contained genes *numa1* and *zfhx3*, which were previously shown to be highly differentiated between northern and southern Norwegian salmon (Kjærner-Semb et al., 2016), and to co-vary with seasonal migration timing in the Teno river system (Pritchard et al., 2018). Interestingly, this haploblock is also adjacent to the parasite-affected region on chromosome 11 identified in Chapter I.

In conclusion, I have identified several loci associated with various environmental characteristics of northern European salmon habitats. Apart from salinity-related loci, the overall number of these loci is quite small however, which may be due to properties of the used datasets or may suggest that the studied environmental factors do not create strong selective pressures.

4.4 Repeated signals of adaptive divergence across independent studies and geographic regions

In the final chapter of the thesis (**Chapter III**) I aimed to investigate how repeatable are candidate loci identified across different geographic regions, and how they compare with other independent studies.

The same 17 haploblocks containing candidate differentially selected loci were discovered in independent analyses of populations in two or three geographically distinct regions studied in **Chapter III**, Teno, Kola, and Barents (Figure 4). Three of these haploblocks (on chromosomes 9, 12 and 25) were also detected as containing candidate selection targets in other studies of Atlantic salmon. The haploblock on chromosome 25 harbours *vgll3* and *akap11* genes of large effect that impact age at sexual maturity across Atlantic salmon populations in Europe (Ayllon et al., 2015; Barson et al., 2015) and potentially in North America (Kusche et al., 2017). The haploblock on chromosome 9 encompassed the *six6* gene, involved in eye and brain development, co-varying with age at maturity and timing of return spawning migration, and being a candidate for differential local selection throughout the Atlantic salmon range (rev. in Pritchard et al. 2018). Finally, the third candidate region included *major histocompatibility complex II*, a well known actor of the immune system, exhibiting signals of divergent selection in Atlantic salmon (Dionne

et al., 2009; Hillestad et al., 2020) and other salmonids (e.g. Larson et al., 2014, 2019).

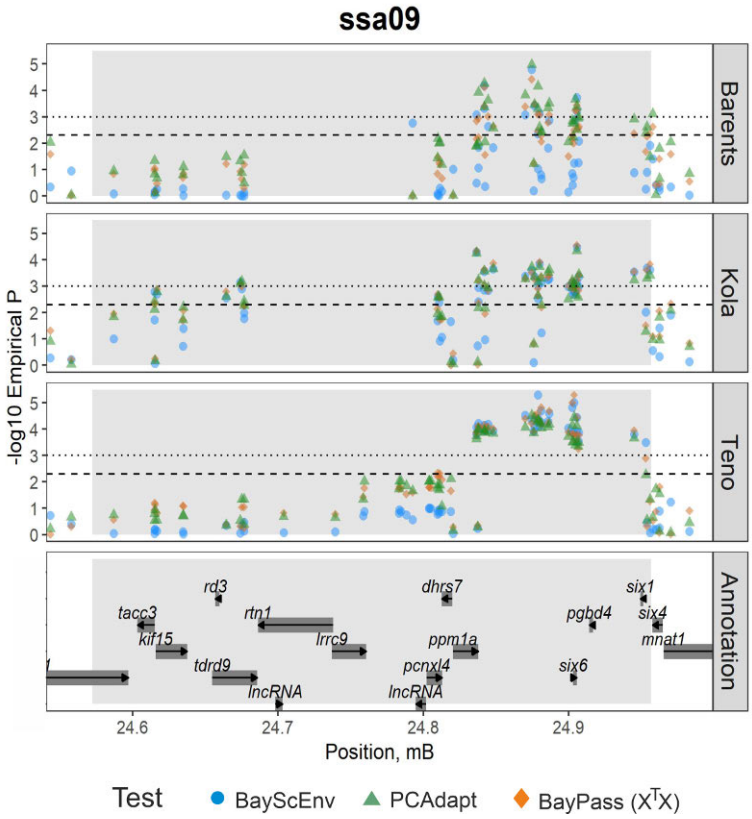


Figure 4. Example of the detected candidate haploblocks identified in Chapter III: a haploblock on chromosome 9, encompassing, among others, the *six6* gene (haploblock boundaries are marked with gray rectangle). Levels of SNP differentiation detected by the 'outlier' tests are presented for the three geographical regions studied (Barents, Kola, Teno), and each dot symbolizes one SNP. Dashed and dotted lines indicate empirical $p < 0.005$ and $p < 0.001$ respectively (where empirical p = SNP rank / total number of tests).

While traditionally the control of complex traits is thought to be polygenic, growing evidence across different taxonomic groups suggests that single loci (or blocks of linked genes) of major effect are widespread targets of selection (Oomen et al., 2020). The results of Chapter III contribute to this, as the same gene clusters of large effect were identified in salmon population groups from two distinct phylogeographic lineages and across a broad geographic scale. These independent repeated signals of adaptive divergence may be indicative of parallel evolution processes across northern European salmon populations, and are great targets for

experimental exploration of their effect on fitness and phenotypic variation in different populations. Moreover, theory predicts that variation in factors such as migration, selection, genetic drift, and degree of gene pleiotropy may play an important role in forming the genetic architecture of local adaptation; for example adaptation with migration tends to lead to fewer, larger, and more tightly linked divergent alleles (Dittmar et al., 2016; Yeaman & Whitlock, 2011). Thus, identifying the relative role of loci with small or large effect may in turn help to understand the background population processes and to direct management practices (Prince et al., 2017).

4.5 Challenges and perspectives

The candidate adaptive targets identified over the course of this dissertation are based on the overlap of various analyses and methodological designs, which give high confidence in the detected loci. However, certain aspects of the study system should be kept in mind when interpreting the results.

First, there are some technical limitations. The methodological approaches implemented here are designed to detect strong signals of directional selection. It is therefore possible that some potentially interesting loci under weaker selection were not identified, especially given my tactic of taking only overlapping results across several methodological approaches (Whitlock & Lotterhos, 2015). Variation between datasets in number of studied populations and in density of SNP coverage also may have affected how much statistical power I had for separating signals of selection from neutral population processes. In addition, approximately 10% of the Atlantic salmon genome retain residual tetrasomy (Lien et al., 2016), and current SNP arrays do not allow detection of regions of potential adaptive importance from this part of the genome.

In **Chapters II** and **III** I used the allelotyping of the pooled genomic DNA approach to infer population-specific allele frequencies. DNA-pooling provides a cost-effective alternative to individual genotyping, and it allowed me to drastically increase the number of studied populations in Chapters II and III, compared to Chapter I. Still, there are some common problems associated with the allelotyping of DNA pools, including potentially high error rate when estimating allele frequencies, and challenges in estimating linkage disequilibrium in the dataset (Ozerov et al., 2013). The first challenge can be solved by applying rigorous quality control and allele frequency correction techniques, which allows an accurate estimation of population allele frequencies (Ozerov et al., 2013; Pritchard et al., 2016). Further, complementing the 'allelotyping' data with a smaller dataset of individually genotyped samples opens additional opportunities, for example a

possibility to reliably estimate linkage disequilibrium and to infer haploblocks existing in the genome (Chapter III).

Another challenge, relevant for all studies of the genomics of adaptation in wild populations, is correlated environmental and life history traits. I have applied sophisticated methodological designs contrasting populations by the response to the parasite, freshwater vs. marine environment, and phylogeographic lineages, to disentangle selection acting upon multiple environmental factors. However, some of the loci detected in relation to a particular selective pressure are likely to be linked to additional environmental traits (reviewed in Chapters I and II). This may be due to methodological limitations, but also due to gene pleiotropic effects. For instance, *vgll3,* the genomic region of major effect for sea age at maturity, is also likely to affect parr maturation and multiple reproduction strategy, traits varying across geographic regions (Aykanat et al., 2019; Lepais et al., 2017). Thus, gene pleiotropy may complicate the interpretation of evolutionary trajectories observed across the Atlantic salmon range.

5 Conclusion

The main findings of this dissertation are twofold.
Firstly, my research suggests a polygenic basis of the variable response of Atlantic salmon to the *G. salaris* parasite, observed in different populations of northern Europe. Taken together, the detected candidate genes strongly suggest that the regulation of both innate and acquired immune responses contribute to parasite resistance and/or tolerance in salmon from freshwater lakes and the Baltic Sea, when contrasted against susceptible Barents Sea and White Sea salmon. My results demonstrate that a wide range of loci, likely having multiple additional functions, can be immunologically relevant in wild populations. To further validate the involvement of detected genes in the *G. salaris* response, one promising tactic could be to assess their expression profiles in controlled challenge experiments, as was done, for instance, in a study challenging brown trout with a live parasitic nematode (Haarder et al., 2013).

Secondly, I demonstrate that across a broad geographic area inhabited by Atlantic salmon, the same few haploblocks repeatedly emerge as targets of local selection, and there is strong evidence for the encompassed loci to be genes of major adaptive importance and likely of large effect. The detected candidate haploblocks were identified in salmon groups of different phylogeographic lineages and varying in such prominent life-history traits as age at maturity and timing of spawning migration, which may attest to parallel evolution processes taking place in salmon in northern Europe.

To summarise, my study presents evidence for polygenic basis of local adaptation in relation to certain selection pressures (e.g. *G. salaris* and salinity), and for loci of major effect driving diversification on a wide geographic scale. Experimental confirmation of the influence of these loci on populations' fitness will help to further understand the evolutionary dynamics of wild salmon of northern Europe and to develop measures for its conservation.

Acknowledgements

This project has been a large part of my professional and personal life for quite a while, and the number of people who directly or indirectly helped me during these years is incredible and inspiring. I am deeply grateful to all of you for making this thesis possible.

Most importantly, this journey would not have happened without my supervisor, Craig Primmer. Craig, you have an amazing ability to give guidance and at the same time freedom for independent thinking, and it was a pleasure and a privilege to be in your research group and to learn from you. Thank you for your patience, for always having time for me, and for pushing me forward. You have always been supportive and encouraging, from the very beginning when I came to the Genetics lab as an exchange student asking whether 'there is a project I can do', till the end, when you were nudging me to 'finish it already and go on with your life'. You have never doubted me, and words cannot express how grateful I am for this alone.

I would not be able to complete (or, for that matter, start) this work without great effort from my co-authors, collaborators, and everybody involved in sample collection and analysis. First of all, huge thanks to my dear friend Victoria Pritchard. Vicky, your sharpness and vast knowledge have always impressed me! But most importantly, thank you for constant encouragement, our chats, night walks with koira, and for always telling me exactly what you think. I would not finish my PhD without you. Thank you Jaakko Lumme, for guiding me in the world of mysterious and wonderful creature *G. salaris*, and for giving an example of pure scientific curiosity that I admire. My deep gratitude to Jaakko and Alexey Veselov, as well as other people who enabled salmon collection, this thesis would not be possible without you. Big thanks to Matthew Kent, Sigbjørn Lien, and everybody involved in the SNP arrays creation and analysis, and huge thanks to Mikhail Ozerov and Anni Tonteri, whose amazing DNA extraction skills saved me a lot of time and effort. Also, Misha's knowledge about Russian salmon populations was invaluable. And even though the *G. salaris* – related project did not make it into this dissertation, I am deeply grateful to Ene Saadre from the Põlula Fish Rearing Center, Heidi Aisala, and Tiina Sävilammi for helping me on the path of parasite transcriptomics.

During my studies, I was fortunate to be part of two astonishing research groups, PnP and EvolConGen. Fellow PnP-ers, our work together, coffee breaks with table football, night movie watching in the lab, and all the extra activities were the best times! Thank you Paula, Anti, Spiros, Misha, Siim, Ilkka, Hannu, Olaf (also Higgins and Kira), Susan J, Susan B, Roland, Erica, Fernando, Silva, Hannu, Marru, Shihab, Sanna, Dima, Tanzeela, Tiina, Megha, Tutku, Yann, and others. Heidi Viitaniemi, Veronica Laine and Pavel Matos, you were the best officemates one can ever wish for, thank you for all the laughs we shared and for being an example for me (see, I finally made it)! Special award also goes to Meri Lindqvist and Ville Auke, for solving with calm confidence every 'Houston, we have a problem' moment that ever occurred!

When the group moved to Helsinki and evolved into EvolConGen, I had the pleasure to meet a whole new bunch of amazing people. Andrew, Marion, Antti and Eirik, your bright energy is positively contagious! Paul, Johanna, Jackie, Nico, Suvi, Outi, Annuka, Yann, Kenyon, J-P, Nikolai, and everybody else – thank you for your work-related and not so work-related comments and encouragements, for great time during meetings, parties, and spawning samplings! Only one thing is there to ask: 'How much is the fish?' ;)

I am very grateful to the past and present people working in the Turku University Genetics division, namely Tatjana, Raja, Katja, Elsi, and Sanna: you made the place so welcoming and pleasant to work in! Special thank you to Irma Saloniemi, for warm words that encouraged me during my exchange studies back in 2010 and also later. I am very lucky to get to know many wonderful researchers from the Biology Department in Turku, especially Chiara, Bineet, Elina, Suvi, Glenda, and Maria. Chats with you about life, dogs, and studies kept me sane. Suvi, you deserve a medal for suffering Mira's smell when I took her to the office right from the shelter!

All this time, my two dearest friends were by my side, Megha Dash and Gayathri Murukesan. Thank you for believing in me and for keeping telling me 'Ksenia, you can do it'. I am extremely lucky to have you both in my life, to share thoughts, aspirations, life milestones with you, or just to chat about everything and nothing. I also want to say arigatou to ladies from the Turku kendo group, Kaisa, Agnieszka, Heta, Iiris, and Elina. Nothing works better after a long day, than sword-fighting and a loud kiai. While started with kendo, friendship with you is so much more now! As always, fighto! Nitin and Mahdy, thank you for taking my mind off work when I need it, you are awesome! Also, huge thanks to Dasha for warmth and encouragement from Russia!

My family gives me unwavering support and love. My mom, a biologist, and dad, a great fisherman, were the first to teach me how to be curious about the world and how to love and appreciate nature; my grandma had always been a role model for me; and my sister, the funniest and the wisest, is a source of immense inspiration.

Thank you! Мама, папа, и Катюха, огромное вам спасибо за поддержку, участие, и вопросы «что ты там делаешь и ну когда уже закончишь?» Без вас ничего бы не вышло :) I am happy I will be able to show you this book and say ‘that’s what I’ve been doing all this time’.

Matthieu, no words can express how much our partnership means or describe the depth of my gratitude to you. You give me strength and support when I need it, you push me in the right direction if I get lost, you tell that everything will be alright if I am down, and you are there laughing with me. Your integrity and sincerity as a human being and as a scientist are extraordinary, thank you for making me a better person too.

I am extremely grateful to Docent Emily Knott and Professor Lisa Seeb for reviewing and commenting on this thesis, and it is an honour to have Dr. Dorte Bekkevold as the opponent for my defence. Also, I sincerely thank Professor Päivi Onkamo for acting as a custos of the public defence.

Vicky, thank you for the language check of this thesis, and thank you Heidi and Heta for translating the Abstract into Finnish. I warmly thank Jon Brommer, Hanna Tuomisto, Raja Rouhiainen, Sanna Ranto, and Säri Jarvi for helping me with all the administrative things, at one point or another.

Lastly, I would like to express my gratitude to the Population Genetics Graduate School, Turku University Foundation, and the Academy of Finland for supporting this work.

January 2021
Ksenia Zueva

List of References

Adlard, R. D., Miller, T. L., & Smit, N. J. 2015. The butterfly effect: parasite diversity, environment, and emerging disease in aquatic wildlife. *Trends in Parasitology*, *31*(4), 160–166.

Alexa, A. & Rahnenfuhrer, J. 2016. *topGO: Enrichment Analysis for Gene Ontology. R package version 2.40.0.*

Andresen, A. M. S., Lutfi, E., Ruyter, B., Berge, G., & Gjøen, T. 2019. Interaction between dietary fatty acids and genotype on immune response in Atlantic salmon (*Salmo salar*) after vaccination: A transcriptome study. *PLoS ONE*, *14*, e0219625.

Armstrong, J. D., Kemp, P. S., Kennedy, G. J. A., Ladle, M., & Milner, N. J. 2003. Habitat requirements of Atlantic salmon and brown trout in rivers and streams. *Fisheries Research*, *62*(2), 143–170.

Asplund, T., Veselov, A., Primmer, C. R., 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 Zoologici Fennici*, *41*, 465–475.

Aykanat, T., Johnston, S. E., Orell, P., Niemelä, E., Erkinaro, J., & Primmer, C. R. 2015. Low but significant genetic differentiation underlies biologically meaningful phenotypic divergence in a large Atlantic salmon population. *Molecular Ecology*, *24*, 5158–5174.

Aykanat, T., Ozerov, M., Vähä, J., Orell, P., Niemelä, E., Erkinaro, J., & Primmer, C. R. 2019. Co-inheritance of sea age at maturity and iteroparity in the Atlantic salmon *vgll3* genomic region. *Journal of Evolutionary Biology*, *32*(4), 343–355.

Ayllon, F., Kjærner-Semb, E., Furmanek, T., Wennevik, V., Solberg, M. F., Dahle, G., Taranger, G. L., Glover, K. A., Almén, M. S., Rubin, C. J., Edvardsen, R. B., & Wargelius, A. 2015. The vgll3 locus controls age at maturity in wild and domesticated Atlantic Salmon (*Salmo salar* L.) males. *PLoS Genetics*, *11*, e1005628.

Bakke, T. A., Jansen, P. A., & Hansen, L. P. 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.

Barson, N. J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G. H., Fiske, P., Jacq, C., Jensen, A. J., Johnston, S. E., Karlsson, S., Kent, M., Moen, T., Niemelä, E., Nome, T., Næsje, T. F., Orell, P., Romakkaniemi, A., Sægrov, H., Urdal, K., … Primmer, C. R. 2015. Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*, *528*, 405–408.

Bekkevold, D., Höjesjö, J., Nielsen, E. E., Aldvén, D., Als, T. D., Sodeland, M., Kent, M. P., Lien, S., & Hansen, M. M. 2020. Northern European *Salmo trutta* (L.) populations are genetically divergent across geographical regions and environmental gradients. *Evolutionary Applications*, *13*, 400–416.

Bertolotti, A. C., Layer, R. M., Gundappa, M. K., Gallagher, M. D., Pehlivanoglu, E., Nome, T., Robledo, D., Kent, M. P., Røsæg, L. L., Holen, M. M., Mulugeta, T. D., Ashton, T. J., Hindar, K., Sægrov, H., Florø-Larsen, B., Erkinaro, J., Primmer, C. R., Bernatchez, L., Martin, S. A. M., … Macqueen, D. J. 2020. The structural variation landscape in 492 Atlantic salmon genomes. *Nature Communications*, *11*, 5176.

Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z., & Galon, J. 2009. ClueGO: a Cytoscape plug-in to decipher functionally

grouped gene ontology and pathway annotation networks. *Bioinformatics Application Note*, *25*, 1091–1093.

Björck, S. 1995. A Review of the history of the Baltic Sea, 13.0-8.0 ka BP. *Quaternary International*, *27*, 19–40.

Bourret, V., Dionne, M., Kent, M. P., Lien, S., & Bernatchez, L. 2013. Landscape genomics in atlantic salmon (*Salmo salar*): Searching for gene-environment interactions driving local adaptation. *Evolution*, *67*(12), 3469–3487.

Bourret, V., Kent, M. P., Primmer, C. R., Vasemägi, A., Karlsson, S., Hindar, K., McGinnity, P., Verspoor, E., Bernatchez, L., & Lien, S. 2013. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Molecular Ecology*, *22*, 532–551.

Bradbury, I. R., Hamilton, L. C., Dempson, B., Robertson, M. J., Bourret, V., Bernatchez, L., & Verspoor, E. 2015. Transatlantic secondary contact in Atlantic Salmon, comparing microsatellites, a single nucleotide polymorphism array and restriction-site associated DNA sequencing for the resolution of complex spatial structure. *Molecular Ecology*, *24*, 5130–5144.

Brown, J. H., Gillooly, J. F., Allen, A. P., Savage, V. M., & West, G. B. 2004. Toward a metabolic theory of ecology. *Ecology*, *85*(7), 1771–1789.

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. 2009. BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 1–9.

Chang, C. C., Chow, C. C., Tellier, L. C., Vattikuti, S., Purcell, S. M., & Lee, J. J. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience*, *4*(7), 1–16.

Consuegra, S., Garcia De Leaniz, C., Serdio, A., Gonzalez Morales, M., Straus, L. G., Knox, D., & Verspoor, E. 2002. Mitochondrial DNA variation in Pleistocene and modern Atlantic salmon from the Iberian glacial refugium. *Molecular Ecology*, *11*, 2037–2048.

Coop, G., Witonsky, D., Di Rienzo, A., & Pritchard, J. K. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics*, *185*, 1411–1423.

de Villemereuil, P., & Gaggiotti, O. E. 2015. A new F_{ST}-based method to uncover local adaptation using environmental variables. *Methods in Ecology and Evolution*, *6*, 1248–1258.

Deagle, B. E., Jones, F. C., Absher, D. M., Kingsley, D. M., & Reimchen, T. E. 2013. Phylogeography and adaptation genetics of stickleback from the Haida Gwaii archipelago revealed using genome-wide single nucleotide polymorphism genotyping. *Molecular Ecology*, *22*, 1917–1932.

Dionne, M., Miller, K. M., Dodson, J. J., & Bernatchez, L. 2009. MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*, 1555–1565.

Dionne, M., Miller, K. M., Dodson, J. J., Caron, F., & Bernatchez, L. 2007. Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. *Evolution; International Journal of Organic Evolution*, *61*(9), 2154–2164.

Dittmar, E. L., Oakley, C. G., Conner, J. K., Gould, B. A., & Schemske, D. W. 2016. Factors influencing the effect size distribution of adaptive substitutions. *Proceedings of the Royal Society B: Biological Sciences*, *283*, 20153065.

Eizaguirre, C. & Lenz, T. L. 2010. Major histocompatibility complex polymorphism: dynamics and consequences of parasite-mediated local adaptation in fishes. *Journal of Fish Biology*, *77*, 2023–2047.

Elmer, K. R. 2016. Genomic tools for new insights to variation, adaptation, and evolution in the salmonid fishes: a perspective for charr. *Hydrobiologia, 783,* 191–208.

Excoffier, L., & Lischer, H. E. L. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, *10*(3), 564–567.

Excoffier, L., Smouse, P. E., & Quattro, J. M. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. *Genetics*, *131*(2), 479–491.

Foll, M. & Gaggiotti, O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics*, *180*, 977–993.

Fraser, D. J., Weir, L. K., Bernatchez, L., Hansen, M. M., & Taylor, E. B. 2011. Extent and scale of local adaptation in salmonid fishes: Review and meta-analysis. *Heredity*, *106*, 404–420.

Frichot, E., Schoville, S. D., Bouchard, G., & François, O. 2013. Testing for associations between loci and environmental gradients using latent factor mixed models. *Molecular Biology and Evolution*, *30*(7), 1687–1699.

Gautier, M. 2015. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics*, *201*, 1555–1579.

Gilbey, J., Matejusova, I., Sorsa-Leslie, T., Collins, C. M., Cunningham, C. O., Verspoor, E., Noble, L. R., & Jones, C. S. 2003. *The genetic basis of Gyrodactylus salaris resistance in Atlantic salmon (Salmo salar).* International Society for Developmental and Comparative Immunology (ISDCI). 9th International Congress, 29 June – 4 July 2003, University of St. Andrews, Scotland, UK.

Gilbey, J., Verspoor, E., Mo, T. A., Sterud, E., Olstad, K., 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.

Grammes, F. 2017. *Ssa.RefSeq.db: Annotation package for the Salmo salar RefSeq (ICSASG_v2). R package version 1.3*.

Günther, T., & Coop, G. 2013. Robust identification of local adaptation from allele frequencies. *Genetics*, *195*(1), 205–220.

Haarder, S., Kania, P. W., Bahlool, Q. Z. M., & Buchmann, K. 2013. Expression of immune relevant genes in rainbow trout following exposure to live *Anisakis simplex* larvae. *Experimental Parasitology*, *135*(3), 564–569.

Haasl, R. J. & Payseur, B. A. 2016. Fifteen years of genomewide scans for selection: trends , lessons and unaddressed genetic sources of complication. *Molecular Ecology*, *25*, 5–23.

Harris, P. D., Bachmann, L., & Bakke, T. A. 2011. The Parasites and Pathogens of Atlantic Salmon: Lessons from *Gyrodactylus salaris*. In Ø. Aas, S. Einum, A. Klemetsen, & J. Skurdal (Eds.), *Atlantic Salmon Ecology*. Wiley-Blackwell Ltd.

Hillestad, B., Kristjánsson, Ó. H., Makvandi-Nejad, S., & Moghadam, H. K. 2020. Genome-wide association study confirms previous findings of major loci affecting resistance to Piscine myocarditis virus in Atlantic salmon (*Salmo salar* L.). *Genes*, *11*, 1–9.

Hoban, S., Kelley, J. L., Lotterhos, K. E., Antolin, M. F., Bradburd, G., Lowry, D. B., Poss, M. L., Reed, L. K., Storfer, A., & Whitlock, M. C. 2016. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *The American Naturalist*, *188*(4), 379–397.

Holm, H. J., Skugor, S., Bjelland, A. K., Radunovic, S., Wadsworth, S., Koppang, E. O., & Evensen, Ø. 2017. Contrasting expression of immune genes in scaled and scaleless skin of Atlantic salmon infected with young stages of *Lepeophtheirus salmonis*. *Developmental and Comparative Immunology*, *67*, 153–165.

Jacobsen, J. A., & Hansen, L. P. 2001. Feeding habits of wild and escaped farmed Atlantic salmon, *Salmo salar* L., in the Northeast Atlantic. *ICES Journal of Marine Science*, *58*, 916–933.

Johnsen, B. O. & Jensen, A. J. 1991. The Gyrodactylus story in Norway. *Aquaculture*, *98*, 289–302.

Kania, P. W., Evensen, O., Larsen, T. B., & Buchmann, K. 2010. Molecular and immunohistochemical studies on epidermal responses in Atlantic salmon *Salmo salar* L. induced by *Gyrodactylus salaris* Malmberg, 1957. *Journal of Helminthology*, *84*, 166–172.

Karlsson, S., Bolstad, G. H., Hansen, H., Jansen, P., Moen, T., & Noble, L. R. 2020. The potential for evolution of resistance to *Gyrodactylus salaris* in Norwegian Atlantic salmon. In *Norwegian Institute for Nature Research Report*.

Karvonen, A. & Seehausen, O. 2012. The role of parasitism in adaptive radiations-when might parasites promote and when might they constrain ecological speciation? *International Journal of Ecology*, *2012*, 1–20.

Kazakov, R. V. & Titov, S. F. 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.

Kjærner-Semb, E., Ayllon, F., Furmanek, T., Wennevik, V., Dahle, G., Niemelä, E., Ozerov, M., Vähä, J.-P., Glover, K. A., Rubin, C. J., Wargelius, A., & Edvardsen, R. B. 2016. Atlantic salmon populations reveal adaptive divergence of immune related genes - a duplicated genome under selection. *BMC Genomics*, *17*, 610.

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.

Konishi, H., Matsuzaki, H., Tanaka, M., Takemura, Y., Kuroda, S., Ono, Y., & Kikkawa, U. 1997. Activation of protein kinase B (Akt/RAC-protein kinase) by cellular stress and its association with heat shock protein Hsp27. *FEBS Letters*, *410*, 493–498.

Kudersky, L., Ieshko, E., & Schulman, B. 2003. Distribution and range formation history of the monogenean *Gyrodactylus salaris* Malmberg, 1957 — a parasite of juvenile Atlantic salmon *Salmo salar* Linnaeus. In *Atlantic salmon Biology, Conservation and Restoration* (pp. 77–83), Petrozavodsk.

Kusche, H., Côté, G., Hernandez, C., Normandeau, E., Boivin-Delisle, D., & Bernatchez, L. 2017. Characterization of natural variation in North American Atlantic Salmon populations (Salmonidae: *Salmo salar*) at a locus with a major effect on sea age. *Ecology and Evolution*, *7*, 5797–5807.

Kuusela, J., Holopainen, R., Meinilä, M., Anttila, P., Koski, P., Ziętara, M. S., Veselov, A., Primmer, C. R., & Lumme, J. 2009. Clonal structure of salmon parasite *Gyrodactylus salaris* on a coevolutionary gradient on Fennoscandian salmon (*Salmo salar*). *Annales Zoologici Fennici*, *46*, 21–33.

Kuusela, J., Zietara, M. S., & Lumme, J. 2007. Hybrid origin of Baltic salmon-specific parasite *Gyrodactylus salaris*: a model for speciation by host switch for hemiclonal organisms. *Molecular Ecology*, *16*, 5234–5245.

Larson, W. A., Dann, T. H., Limborg, M. T., McKinney, G. J., Seeb, J. E., & Seeb, L. W. 2019. Parallel signatures of selection at genomic islands of divergence and the MHC in ecotypes of sockeye salmon across Alaska. *Molecular Ecology*, *28*, 2254–2271.

Larson, W. A., Seeb, J. E., Dann, T. H., Schindler, D. E., & Seeb, L. W. 2014. Signals of heterogeneous selection at an MHC locus in geographically proximate ecotypes of sockeye salmon. *Molecular Ecology*, *23*, 5448–5461.

Lepais, O., Manicki, A., Glise, S., Buoro, M., & Bardonnet, A. 2017. Genetic architecture of threshold reaction norms for male alternative reproductive tactics in Atlantic salmon (*Salmo salar* L.). *Scientific Reports*, *7*, 1–13.

Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., Hvidsten, T. R., Leong, J. S., Minkley, D. R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A., Walenz, B., Hermansen, R. A., von Schalburg, K., Rondeau, E. B., Di Genova, A., Samy, J. K. A., … Davidson, W. S. 2016. The Atlantic salmon genome provides insights into rediploidization. *Nature*, *533*, 200–205.

Liu, K. & Muse, S. V. 2005. PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics*, *21*(9), 2128–2129.

Lotterhos, K. E. 2019. The effect of neutral recombination variation on genome scans for selection. *G3: Genes, Genomes, Genetics*, *9*, 1851–1867.

Lumme, J., Ozerov, M. Y., Veselov, A. E., & Primmer, C. R. 2015. The Formation of Landlocked Populations of Atlantic Salmon. In T. Vladiĉ & E. Petersson (Eds.), *Evolutionary Biology of the Atlantic salmon* (pp. 26–43). CRC Press.

Luu, K., Bazin, E., & Blum, M. G. B. 2017. pcadapt: an R package to perform genome scans for selection based on principal component analysis. *Molecular Ecology Resources*, *17*(1), 67–77.

Maere, S., Heymans, K., & Kuiper, M. 2005. BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics Application Note*, *21*, 3448–3449.

Makhrov, A. A., Verspoor, E., Artamonova, V. S., & O'Sullivan, M. 2005. Atlantic salmon colonization of the Russian Arctic coast: pioneers from North America. *Journal of Fish Biology*, *67*, 68–79.

Manel, S., Perrier, C., Pratlong, M., Abi-Rached, L., Paganini, J., Pontarotti, P., & Aurelle, D. 2016. Genomic resources and their influence on the detection of the signal of positive selection in genome scans. *Molecular Ecology*, *25*, 170–184.

Matejusová, I., Felix, B., Sorsa-Leslie, T., Gilbey, J., Noble, L. R., Jones, C. S., & Cunningham, C. O. 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.

Matthey-Doret, R. & Whitlock, M. C. 2019. Background selection and F_{ST}: Consequences for detecting local adaptation. *Molecular Ecology*, *28*(17), 3902–3914.

McCormick, S. D., Regish, A. M., Ardren, W. R., Björnsson, B. T., & Bernier, N. J. 2019. The evolutionary consequences for seawater performance and its hormonal control when anadromous Atlantic salmon become landlocked. *Scientific Reports*, *9*(968), 1–10.

Mitchell, P. S., Emerman, M., & Malik, H. S. 2013. An evolutionary perspective on the broad antiviral specificity of MxA. *Current Opinion in Microbiology*, *16*, 493–499.

Moen, T., Sonesson, A. K., Hayes, B., Lien, S., Munck, H., & Meuwissen, T. H. E. 2007. Mapping of a quantitative trait locus for resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*): comparing survival analysis with analysis on affected/resistant data. *BMC Genetics*, *8*(53), 1–13.

Moen, T., Torgersen, J., Santi, N., Davidson, W. S., Baranski, M., Ødegård, J., Kjøglum, S., Velle, B., Kent, M., Lubieniecki, K. P., Isdal, E., & Lien, S. 2015. Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic Salmon. *Genetics*, *200*, 1313–1326.

Moore, L. J., Jarungsriapisit, J., Nilsen, T. O., Stefansson, S., Taranger, G. L., Secombes, C. J., Morton, H. C., & Patel, S. 2017. Immune gene profiles in Atlantic salmon (*Salmo salar* L.) post-smolts infected with SAV3 by bath-challenge show a delayed response and lower levels of gene transcription compared to injected fish. *Fish and Shellfish Immunology*, *62*, 320–331.

Myrvold, K. M., Mawle, G. W., Andersen, O., & Aas, Ø. 2019. The Social, Economic and Cultural values of wild Atlantic salmon. A review of literature for the period 2009-2019 and an assessment of changes in values. *NINA Report 1668. Norwegian Institute for Nature Research*, 94.

NASCO. 2018. Report of the Meeting of the Working Group on *Gyrodactylus salaris* in the North-East Atlantic Commission Area. *North Atlantic Salmon Conservation Organisation Report*, 1–57.

NASCO. 2019. State of North Atlantic Salmon. *North Atlantic Salmon Conservation Organisation Report*, 1–30.

Nielsen, R., Williamson, S., Kim, Y., Hubisz, M. J., Clark, A. G., & Bustamante, C. 2005. Genomic scans for selective sweeps using SNP data. *Genome Research*, *15*, 1566–1575.

Nilsen, T. O., Ebbesson, L. O. E., Kiilerich, P., Björnsson, B. T., Madsen, S. S., McCormick, S. D., & Stefansson, S. O. 2008. Endocrine systems in juvenile anadromous and landlocked Atlantic salmon (*Salmo salar*): seasonal development and seawater acclimation. *General and Comparative Endocrinology*, *155*, 762–772.

Nilsen, T. O., Ebbesson, L. O. E., & Stefansson, S. O. 2003. Smolting in anadromous and landlocked strains of Atlantic salmon (*Salmo salar*). *Aquaculture*, *222*, 71–82.

Norman, J. D., Robinson, M., Glebe, B., Ferguson, M. M., & Danzmann, R. G. 2012. Genomic arrangement of salinity tolerance QTLs in salmonids: A comparative analysis of Atlantic salmon (*Salmo salar*) with Arctic charr (*Salvelinus alpinus*) and rainbow trout (*Oncorhynchus mykiss*). *BMC Genomics*, *13*(1), 420.

Oleksyk, T. K., Smith, M. W., & O'Brien, S. J. 2010. Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *365*, 185–205.

Olsen, J. B., Crane, P. A., Flannery, B. G., Dunmall, K., Templin, W. D., & Wenburg, J. K. 2010. Comparative landscape genetic analysis of three Pacific salmon species from subarctic North America. *Conservation Genetics*, *12*, 223–241.

Oomen, R. A., Kuparinen, A., & Hutchings, J. A. 2020. Consequences of single-locus and tightly linked genomic architectures for evolutionary responses to environmental change. *The Journal of Heredity*, *111*(4), 319–332.

Ozerov, M., Vähä, J.-P., Wennevik, V., Niemelä, E., Svenning, M.-A., Prusov, S., Diaz Fernandez, R., Unneland, L., Vasemägi, A., Falkegård, M., Kalske, T., & Christiansen, B. 2017. Comprehensive microsatellite baseline for genetic stock identification of Atlantic salmon (*Salmo salar* L.) in northernmost Europe. *ICES Journal of Marine Science*, *74*(8), 2159–2169.

Ozerov, M., Vasemägi, A., Wennevik, V., Niemelä, E., Prusov, S., & Kent, M. 2013. Cost-effective genome-wide estimation of allele frequencies from pooled DNA in Atlantic salmon (*Salmo salar* L.). *BMC Genomics*, *14*, 1–9.

Ozerov, M., Veselov, A. E., Lumme, J., & Primmer, C. R. 2013. Temporal variation of genetic composition in Atlantic salmon populations from the Western White Sea Basin: influence of anthropogenic factors? *BMC Genetics*, *14*(88), 1–15.

Ozerov, M. Y., Veselov, A. E., Lumme, J., & Primmer, C. R. 2012. "Riverscape" genetics: river characteristics influence the genetic structure and diversity of anadromous and freshwater Atlantic salmon (*Salmo salar*) populations in northwest Russia. *Canadian Journal of Fisheries and Aquatic Sciences*, *69*, 1–12.

Ozerov, M. Y., Veselov, A. J., Lumme, J., & Primmer, C. R. 2010. Genetic structure of freshwater Atlantic salmon (*Salmo salar* L.) populations from the lakes Onega and Ladoga of northwest Russia and implications for conservation. *Conservation Genetics*, *11*, 1711–1724.

Palstra, F. P. & Ruzzante, D. E. 2010. A temporal perspective on population structure and gene flow in Atlantic salmon (*Salmo salar*) in Newfoundland, Canada. *Canadian Journal of Fisheries and Aquatic Sciences*, *67*(2), 225–242.

Patton, H., Hubbard, A., Andreassen, K., Auriac, A., Whitehouse, P. L., Stroeven, A. P., Shackleton, C., Winsborrow, M., Heyman, J., & Hall, A. M. 2017. Deglaciation of the Eurasian ice sheet complex. *Quaternary Science Reviews*, *169*, 148–172.

Pickrell, J. K. & Pritchard, J. K. 2012. Inference of population splits and mixtures from genome-wide allele frequency fata. *PLoS Genetics*, *8*(11), e1002967.

Ponomareva, E. V. 2007. Population structure of Atlantic salmon (*Salmo salar* L.) from European North of Russia (RUS). *Doctoral Dissertation*, 1–120.

Potutkin, A. G., Veselov, A. J., Zubchenko, A. V., & Kaliuzhin, S. M. 2007. Anadromous migration of Atlantic salmon from the river Varzuga in the White Sea. *Ecology. Experimental Genetics and Physics (RUS).*, *11*, 100–108.

Primmer, C. R. 2011. Genetics of local adaptation in salmonid fishes. *Heredity, 106*, 401–403.

Prince, D. J., O'Rourke, S. M., Thompson, T. Q., Ali, O. A., Lyman, H. S., Saglam, I. K., Hotaling, T. J., Spidle, A. P., & Miller, M. R. 2017. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Science Advances*, *3*, e1603198.

Pritchard, V. L., Erkinaro, J., Kent, M. P., Niemelä, E., Orell, P., Lien, S., & Primmer, C. R. 2016. Single nucleotide polymorphisms to discriminate different classes of hybrid between wild Atlantic salmon and aquaculture escapees. *Evolutionary Applications*, *9*, 1017–1031.

Pritchard, V. L., Mäkinen, H., Vähä, J. P., Erkinaro, J., Orell, P., & Primmer, C. R. 2018. Genomic signatures of fine-scale local selection in Atlantic salmon suggest involvement of sexual maturation, energy homeostasis and immune defence-related genes. *Molecular Ecology*, *27*, 2560–2575.

Przeworski, M., Coop, G., & Wall, J. D. 2005. The signature of positive selection on standing genetic variation. *Evolution*, *59*(11), 2312–2323.

Quinlan, A. R. & Hall, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, *26*(6), 841–842.

R Core Team. 2019. *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.*

Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M., & Holderegger, R. 2015. A practical guide to environmental association analysis in landscape genomics. *Molecular Ecology*, *24*, 4348–4370.

Reyes-Lopez, F. E., Romeo, J. S., Vallejos-Vidal, E., Reyes-Cerpa, S., Sandino, A. M., Tort, L., Mackenzie, S., & Imarai, M. 2015. Differential immune gene expression profiles in susceptible and resistant full-sibling families of Atlantic salmon (*Salmo salar*) challenged with infectious pancreatic necrosis virus (IPNV). *Developmental and Comparative Immunology*, *53*, 210–221.

Rissler, L. J. 2016. Union of phylogeography and landscape genetics. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 8079–8086.

Rougemont, Q. & Bernatchez, L. 2018. The demographic history of Atlantic salmon (*Salmo salar*) across its distribution range reconstructed from approximate Bayesian computations. *Evolution*, *72*(6), 1261–1277.

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.

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.

Salminen, M., Erkamo, E., & Salmi, J. 2001. Diet of post-smolt and one-sea-winter Atlantic salmon in the Bothnian Sea, Northern Baltic. *Journal of Fish Biology*, *58*, 16–35.

Salte, R., Bentsen, H. B., Moen, T., Tripathy, S., Bakke, T. A., Ødegård, J., Omholt, S., & Hansen, L. P. 2010. Prospects for a genetic management strategy to control *Gyrodactylus salaris* infection in wild Atlantic salmon (*Salmo salar*) stocks. *Canadian Journal of Fisheries and Aquatic Sciences*, *67*, 121–129.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research*, *13*, 2498–2504.

Shaw, R. G. & Etterson, J. R. 2012. Rapid climate change and the rate of adaptation: Insight from experimental quantitative genetics. *New Phytologist*, *195*(4), 752–765.

Stroeven, A. P., Hättestrand, C., Kleman, J., Heyman, J., Fabel, D., Fredin, O., Goodfellow, B. W., Harbor, J. M., Jansen, J. D., Olsen, L., Caffee, M. W., Fink, D., Lundqvist, J., Rosqvist, G. C., Strömberg, B., & Jansson, K. N. 2016. Deglaciation of Fennoscandia. *Quaternary Science Reviews*, *147*, 91–121.

Tadiso, T. M., Krasnov, A., Skugor, S., Afanasyev, S., Hordvik, I., & Nilsen, F. 2011. Gene expression analyses of immune responses in Atlantic salmon during early stages of infection by salmon louse (*Lepeophtheirus salmonis*) revealed bi-phasic responses coinciding with the copepod-chalimus transition. *BMC Genomics*, *12*, 141.

Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J., & Richards, J. B. 2018. Genetic architecture: The shape of the genetic contribution to human traits and disease. *Nature Reviews Genetics*, *19*, 110–124.

Tonteri, A., Titov, S., Veselov, A., Zubchenko, A., Koskinen, M. T., Lesbarrères, D., Kaluzchin, S., Bakhmet, I., Lumme, J., & Primmer, C. R. 2005. Phylogeography of anadromous and non-anadromous Atlantic salmon (*Salmo salar*) from northern Europe. *Annales Zoologici Fennici*, *41*, 1–22.

Tonteri, A., Vasemägi, A., Lumme, J., & Primmer, C. R. 2010. Beyond MHC: signals of elevated selection pressure on Atlantic salmon (*Salmo salar*) immune-relevant loci. *Molecular Ecology*, *19*, 1273–1282.

Tonteri, A., Veselov, A. J., Titov, S., Lumme, J., & Primmer, C. R. 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*, 381–398.

Tonteri, A., Veselov, A. J., Zubchenko, A. V, Lumme, J., & Primmer, C. R. 2009. Microsatellites reveal clear genetic boundaries among Atlantic salmon (*Salmo salar*) populations from the Barents and White seas, northwest Russia. *Canadian Journal of Fisheries and Aquatic Sciences*, *66*, 717–735.

Turner, C. B., Marshall, C. W., & Cooper, V. S. 2018. Parallel genetic adaptation across environments differing in mode of growth or resource availability. *Evolution Letters*, *2*(4), 355–367.

Waldvogel, A., Feldmeyer, B., Rolshausen, G., Exposito-Alonso, M., Rellstab, C., Kofler, R., Mock, T., Schmid, K., Schmitt, I., Bataillon, T., Savolainen, O., Bergland, A., Flatt, T., Guillaume, F., & Pfenninger, M. 2020. Evolutionary genomics can improve prediction of species' responses to climate change. *Evolution Letters*, *4*(1), 4–18.

Wand, M. P. & Jones, M. C. 1995. Kernel smoothing. In *Monographs on Statistics and Applied Probability*. Chapman&Hall.

Weigand, H. & Leese, F. 2018. Detecting signatures of positive selection in non-model species using genomic data. *Zoological Journal of the Linnean Society*, *184*(2), 528–583.

Weir, B. S. & Cockerham, C. C. 1984. Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358–1370.

Wennevik, V., Quintela, M., Skaala, Ø., Verspoor, E., Prusov, S., & Glover, K. A. 2019. Population genetic analysis reveals a geographically limited transition zone between two genetically distinct Atlantic salmon lineages in Norway. *Ecology and Evolution*, *00*, 1–21.

Whitlock, M. C. & Lotterhos, K. E. 2015. Reliable detection of loci responsible for local adaptation: inference of a null model through trimming the distribution of F_{ST}. *The American Naturalist*, *186*, 26–36.

Wilson, K., Fenton, A., & Tompkins, D. 2019. Wildlife Disease Ecology. In *Basic and Applied Ecology*. Cambridge University Press.

Wu, C. I. 2001. The genic view of the process of speciation. *Journal of Evolutionary Biology*, *14*(6), 851–865.

Yeaman, S., Gerstein, A. C., Hodgins, K. A., & Whitlock, M. C. 2018. Quantifying how constraints limit the diversity of viable routes to adaptation. *PLoS Genetics*, *14*(10), 1–25.

Yeaman, S. & Whitlock, M. C. 2011. The genetic architecture of adaptation under migration-selection balance. *Evolution*, *65*(7), 1897–1911.

UNIVERSITY OF TURKU

Painosalama Oy, Turku, Finland 2021

ISBN 978-951-29-8327-8 (PRINT)
ISBN 978-951-29-8328-5 (PDF)
ISSN 0082-6979 (Print)
ISSN 2343-3183 (Online)