

Population genomics reveals repeated signals of adaptive divergence in northeastern European Atlantic salmon

Ksenia J. Zueva¹, Jaakko Lumme², Alexey E. Veselov^{3*}, Craig R. Primmer^{4,5} & Victoria
L. Pritchard^{4,6}.

¹ *Department of Biology, University of Turku, Turku, Finland. ksezue@utu.fi*

² *Department of Biology, University of Oulu, Oulu, Finland*

³ *Institute of Biology of the Karelian Research Centre, Petrozavodsk, Russia*

⁴ *Department of Biosciences, University of Helsinki, Helsinki, Finland*

⁵ *Institute of Biotechnology, University of Helsinki, Helsinki, Finland*

⁶ *Rivers and Lochs Institute, Inverness College, University of Highlands and Islands, Inverness, U.K.
victorialpritchard@gmail.com*

**The authors dedicate this work to the memory of our dear friend and colleague, Alexey E Veselov, who passed away
on 03.05.2020.*

Corresponding author:

Ksenia Zueva

Division of Genetics and Physiology

Department of Biology (Vesilinnantie 5)

FI-20014 University of Turku

FINLAND

Phone (358) 40 547 2780

Email: ksezue@utu.fi

This article has been accepted for publication and undergone full peer review but has not been
through the copyediting, typesetting, pagination and proofreading process, which may lead to
differences between this version and the [Version of Record](#). Please cite this article as [doi:
10.1111/jeb.13732](#)

This article is protected by copyright. All rights reserved.
This article is protected by copyright. All rights reserved.

Population genomics reveals repeated signals of adaptive divergence in the Atlantic salmon of northeastern Europe.

ABSTRACT

Our ability to examine genetic variation across entire genomes have enabled many studies searching for the genetic basis of local adaptation. These studies have identified numerous loci as candidates for differential local selection, however relatively few have examined the overlap among candidate loci identified from independent studies of the same species in different geographic areas or evolutionary lineages. We used an allelotyping approach with a 220K SNP array to characterize the population genetic structure of Atlantic salmon in northeastern Europe and ask whether the same genomic segments emerged as outliers among populations in different geographic regions. Genome-wide data recapitulated the phylogeographic structure previously inferred from mtDNA and microsatellite markers. Independent analyses of three genetically and geographically distinct groups of populations repeatedly inferred the same 17 haploblocks to contain loci under differential local selection. The most strongly supported of these replicated haploblocks had known strong associations with life history variation or immune response in Atlantic salmon. Our results are consistent with these genomic segments harbouring large-effect loci which have a major role in Atlantic salmon diversification and are ideal targets for validation studies.

Keywords: local adaptation, divergent selection, Atlantic salmon, allelotyping, Russia

INTRODUCTION

Technological advances enabling variation to be examined across entire genomes have enabled many studies searching for the genetic basis of local adaptation. These studies, which commonly examine a collection of populations for the genomic signatures of selective sweeps or the presence of polymorphisms that co-vary with environmental parameters, have identified a myriad of loci as ‘candidates’ to underly adaptive phenotypes (Haas & Payseur, 2015; Hoban *et al.*, 2016). Fewer

Accepted Article

studies, however, have investigated the extent of overlap among candidate loci across independent studies of the same species in different geographic areas or evolutionary lineages. Examining such overlap can provide important insights into the evolutionary process, in particular allowing us to ask whether the same genetic architecture repeatedly underlies adaptative diversification (e.g. Roesti *et al.*, 2012; Ravinet *et al.*, 2015; Turner 2017; Yeaman *et al.*, 2018), and whether this architecture includes loci of large effect (Oomen *et al.*, 2020). Further, from a technical point of view, the rate of false positive candidates in a single study can be high (Weigand *et al.*, 2018; Lotterhos & Whitlock, 2015; François *et al.*, 2016). Loci that are repeatedly identified as locally adaptive candidates in replicated or complementary analysis of independent datasets (e.g. Buehler *et al.*, 2014; Liu *et al.*, 2014; Rellstab *et al.*, 2015) are more likely to be true positives that are important in adaptive diversification for the species as a whole. Such loci are ideal targets for further research aimed at validating their adaptive importance by quantifying their phenotypic consequences and characterizing their effect on fitness in different environments, and ultimately understanding their population genetic history and evolutionary significance.

Salmonid fishes, including the anadromous and potadromous Atlantic salmon (*Salmo salar*), typically show strong population genetic subdivision due to their precise natal homing and/or isolation in different bodies of water (reviewed in Fraser *et al.*, 2011). They also exhibit a wide range of life history variation within and across populations, including differences in age at sexual maturity, migration timing and breeding location, and exploitation of ecological niches (e.g. Bernatchez *et al.*, 2010; Dodson *et al.*, 2013; Quinn *et al.*, 2016). These properties, together with the excellent genomic resources available for the taxon, have made salmonid fishes important models for examining the genetic basis of phenotypic variation and local adaptation. Recent studies have identified three independent large-effect loci that repeatedly underlie life-history variation across salmonid populations and species (Ayllon *et al.*, 2015; Barson *et al.*, 2015; Hess *et al.*, 2016; Prince *et al.*, 2017; Veale & Russello, 2017). A small number of additional loci have been confirmed as strong candidates to underly local adaptation across multiple populations (Pritchard *et al.*, 2018; Larson *et al.*, 2019). Nevertheless, most loci that have been identified as locally adaptive candidates in salmonids remain unique to a single study. There is a need to examine how many of these loci are of broader adaptive significance.

The native breeding range of Atlantic salmon spans the north Atlantic coast from eastern North America to western Russia. Populations are genetically structured at multiple spatial scales, which reflects long-term isolation and post-glacial recolonization processes. This structure ranges from extreme divergence between the North American and European lineages to fine-scale differentiation between populations in different tributaries of the same river (e.g. Vähä *et al.*, 2007; Lehnert *et al.*,

2019; Wennevik *et al.*, 2019). European Atlantic salmon have been separated into up to five major phylogenetic lineages, each with a distinct geographic range (Verspoor *et al.*, 1999; King *et al.*, 2001; Bourret *et al.*, 2013, Rougemont & Bernatchez 2018). The 'Barents-White Seas' lineage of Atlantic salmon is distributed across the contiguous subarctic regions of northeastern Norway, northern Finland and northwestern Russia (Fig. 1, Bourret *et al.*, 2013; Wennevik *et al.*, 2019). Rivers in this region support some of the world's most important wild Atlantic salmon populations, which face ongoing threats including legal and illegal fishing pressure, the local expansion of salmon aquaculture, and changing environmental conditions (Primmer *et al.*, 2006; Ozerov *et al.* 2012). Previous studies using allozymes, mtDNA and microsatellites have identified further substructure within the Barents-White Seas lineage. In particular, on the Russian Kola Peninsula, there is a clear genetic transition between Atlantic salmon spawning in rivers that open into the Barents Sea and those spawning in rivers that open into the White Sea (Asplund *et al.*, 2004; Tonteri *et al.*, 2009; Ozerov *et al.*, 2017; Wennevik *et al.*, 2019). This transition is associated with broad-scale differences in life history parameters between the two regional groups including age at sexual maturity and timing of the return spawning migration (Jensen, 1999; Potutkin *et al.*, 2007; Ponomareva, 2007). Many studies have also identified a second genetic transition between the White Sea populations that spawn in rivers on the Kola Peninsula and those spawning in rivers opening onto the southwestern White Sea in the Republic of Karelia (Asplund *et al.*, 2004; Tonteri *et al.*, 2009; but see Ozerov *et al.*, 2017). This regional sub-structuring is suggested to be the result of post-glacial recolonization from multiple refugia. There is a consensus that the Barents region was re-colonized from the eastern Atlantic - uniquely among European Atlantic salmon the Barents populations also contain genetic signals consistent with gene flow from North America (Asplund *et al.*, 2004; Makhrov *et al.*, 2005; Bradbury *et al.*, 2015). Salmon are suggested to have re-colonized the White Sea from a northeastern glacial refugium (Asplund *et al.*, 2004; Tonteri *et al.*, 2009; Bourret *et al.*, 2013) or alternatively from the Baltic via freshwater Lake Onega (Makhrov *et al.*, 2005).

Recently, Pritchard *et al.* (2018) used a SNP array to scan for genomic signatures of differential local selection among 10 genetically distinct populations in the Teno (Tana/Deatnu) River, a drainage which supports a large, subdivided, Barents Sea Atlantic salmon stock. They examined patterns of interpopulation differentiation, allelic association with upstream catchment area and distance from the ocean, and local haplotype homozygosity, and identified 32 genomic segments as particularly strong candidates to contain locally adaptive variants. Here, we use the same SNP array to characterize population structure among 30 additional salmon populations from northeastern Europe, and ask whether the same genomic segments emerge as locally adaptive candidates between the different geographic regions.

MATERIALS AND METHODS

Data formatting and statistical analyses were performed either in R 3.5.1. (R Core Team 2019), with assistance from functions in the *tidyverse* package (Wickham 2017), or in the Linux shell.

Sample collection & DNA extraction: We used samples of Atlantic salmon that had been collected from 29 rivers in northeastern Russia and one river in northwestern Finland/Norway (Näätämö/Neiden/Njeävdám) between 1999 and 2008 (median sample size = 44 fish, range = 25-240; Table 1, Fig. 1). All samples were obtained according to relevant national legislations. Most of these rivers are known to have temporally stable population genetic structure (Ozerov *et al.*, 2013). Due to the late expansion of Atlantic salmon farming into this region they are not expected to have contained genetic material from aquaculture strains at the time of sampling, however several of the rivers including Kola and Uмба have been subject to long-term supplementary stocking programs using locally caught broodstock (Makhrov *et al.* 2004). All Russian river samples were fin clips in ethanol from juvenile fish caught by electrofishing. Näätämö samples were dry scales taken from adult fish harvested in the river. DNA had previously been extracted from the samples using various methods (Ozerov *et al.*, 2012; Zueva *et al.*, 2014, 2018; Pritchard *et al.*, 2016) and stored at -20°C for several years. All archived DNA was qualitatively assessed for degradation using gel electrophoresis. For 618 degraded samples, DNA was freshly extracted from stored tissue using QIAamp DNA mini kits (Qiagen, Table 1).

DNA pool preparation and allelotyping: Each river was considered to represent a separate population. We used an allelotyping approach to estimate population allele frequencies as detailed in Pritchard *et al.* (2016) and Zueva *et al.* (2018). Briefly, all DNA extractions were fluorometrically quantified and adjusted to a final concentration of 10 ng/ul. Equal volumes of standardized DNA from each individual were combined into a pool, and the final concentration of each pool was verified as 10 ± 0.5 ng/ul. Each pool was independently generated four times ('pipetting replicates'). For Russian rivers, the pools included all fish samples available for the river. For historical reasons, the 240 Näätämö samples were divided by sex and age-at-maturity into four pools of 60 fish each.

All pools were allelotyped at the Centre for Integrative Genetics (CIGENE, Norway) using a custom 220,000 Atlantic salmon SNP array ('220K SNP chip', Affymetrix Axiom) on a GeneTitan genotyping platform. The SNPs on the array are a subset of 930,000 SNPs that were discovered in Norwegian aquaculture salmon and included on a previous array (unpublished). The 220K SNPs were selected for maximum informativeness of the basis of their SNPish performance (SNPish, V1.4, Affymetrix), minor allele frequency (MAF) in aquaculture strains, and physical distribution across the genome (Barson *et al.*, 2015). The aquaculture strains used for SNP discovery

were established in the 1970s from 41 Norwegian and Swedish wild populations (Glover *et al.*, 2017), thus ascertainment bias is expected to be low for northern European wild populations in general.

The 220K SNP array has previously been used to identify candidate loci underlying age-at-maturity, and migration timing, and responding to differential local selection at fine and broad scales (Barson *et al.*, 2015; Cauwelier *et al.*, 2017; Pritchard *et al.*, 2018; Sylvester *et al.*, 2018; Lehnert *et al.*, 2019, 2020), and is a workhorse for the genetic improvement of aquaculture salmon (e.g. Kijas *et al.*, 2017).

Data pre-processing: We calculated raw relative B allele frequency for each allelotyped pool at each SNP from the A and B intensity values returned by CIGENE. To identify any SNPs with unusually high noise among pipetting replicates, we estimated the standard deviation (SD) of B allele frequencies over the four replicates, took the 20% of SNPs with the highest SD, and then compared the identity of these SNPs among populations (as in Zueva *et al.* 2018). No SNP had consistently high SD across pools for all populations, and therefore no SNPs were excluded during this step. We then calculated the mean B allele frequency across all replicates within a river (4 for Russian rivers, 16 for Näämöö), and applied a polynomial probe correction algorithm (PPC) to adjust this B allele frequency estimate for SNP-specific variation in the A and B intensities. We used PPC parameters that had previously been estimated from a simultaneously genotyped and allelotyped sample of 610 Atlantic salmon from the Teno River, as in Pritchard *et al.* (2016). We note that population-specific polymorphisms in SNP flanking sequences, and incomplete representation of all three genotypes in the Teno dataset used to calculate PPC parameters, may mean that this correction is suboptimal for some SNPs in our current dataset. However, this is expected to add random noise to allele frequency estimates rather than generating systematic bias.

The following SNPs were permanently excluded from the initial 220,000 SNP dataset: 1,846 SNPs without a known position on the Atlantic salmon genome assembly (ICSASG_v2, GCF_000233375.1); 36 SNPs with known off-target variants; and 1,225 SNPs previously found to deviate from HWE at $p < 0.0001$ in either of two large single-population samples of individually genotyped fish (Inarijoki, $n = 120$; Teno mainstem $n = 268$; Aykanat *et al.*, 2015; Pritchard *et al.*, 2018). The latter SNPs were excluded because HW deviations were considered to indicate technical genotyping problems. Prior to each analysis, we also removed SNPs with $MAF < 0.05$ across the relevant set of populations. Where required by input file format, we converted estimated allele frequencies to allele counts. Based on results from Ozerov *et al.*, 2013, who found population allele frequencies to be accurately estimated from pools containing as few as 35 fish, we assumed a similar level of error for all pools and standardized the total number of alleles to 80 per population (40 diploid individuals).

Genetic variation & population genetic structure: We estimated genome-wide heterozygosity for each population as mean observed heterozygosity over all SNPs following MAF filtering over the entire 30-population dataset (186,107 SNPs). To explore regional population genetic structure, we performed a principal component analysis (PCA) on population allele frequencies using the R package *PCAdapt* 4.1.0 (Luu *et al.*, 2017). As we could not directly estimate linkage disequilibrium (LD) from our dataset, we approximated LD pruning by excluding the same SNPs that were removed in the LD pruning step for the Teno River dataset of Pritchard *et al.* (2018): 28,771 SNPs were kept. We initially ran the PCA allowing 29 PCs and selected the most suitable number to retain by examining the resulting scree plot.

This initial PCA (see below) identified two main clusters which separated along PC1 and included all Barents and Kola Peninsula populations (hereafter 'Barents' and 'Kola' regional groups). Thus, differentiation along PC1 corresponded to the previously observed genetic division between Barents and White Sea Atlantic salmon. The remaining six populations in South-West (SW) and East (E) White Sea separated from the two main clusters and also from one other along PCs 2-5 (see below). To examine how different parts of the genome contributed to this overall differentiation between the Barents and Kola clusters, while avoiding confounding effects from other outlier populations, we retained only the 24 populations in the two main clusters and used *PCAdapt* with $K=1$ to identify SNPs (out of 185,023 after MAF filtering) that disproportionately contributed to the single PC separating the two clusters. We transformed *PCAdapt* p-values into q-values using *qvalue* (Storey *et al.*, 2002) and defined SNPs with $q < 0.05$ as outliers. We annotated these outliers with the overlapping or closest downstream protein coding gene (on either strand) using the *closest* function of BEDTools 2.29.0 (Quinlan & Hall, 2010) in combination with NCBI *Salmo salar* Annotation Release 100. We also asked whether SNPs with high inter-cluster differentiation overlapped with recently documented structural variants in Atlantic salmon (including Barents and White Sea populations, Bertolotti *et al.*, 2020) using BEDTools *closest*.

Phylogeographic history of Atlantic salmon in northeastern Europe: To investigate the phylogeographic history of western Russian Atlantic salmon in a wider geographical context, we combined our data with two additional 220K population allelotyping datasets. The first dataset included Atlantic salmon populations from the Baltic Sea, Lake Ladoga, Lake Onega, and landlocked sites in the Republic of Karelia, and was generated simultaneously with the primary Russian dataset (Zueva *et al.*, 2018; Fig. 1, Table S1). The second dataset, from Pritchard *et al.* 2016 (an independent dataset from that of Pritchard *et al.* 2018), included salmon from five tributaries of the Teno River (Inarjoki, Kevojoki, Upper Pulmankijoki, Tsarsjoki & Utsjoki). Allele frequency estimation and quality control had been performed identically for all datasets; 196,621 SNPs were

retained after MAF filtering across all 52 populations. We reconstructed the bifurcating tree relationship among sampled populations, allowing possible historical mixing events, using TreeMix (Pickrell *et al.*, 2012). We specified the River Vindel as the outgroup based on previous phylogenetic studies of the region (Tonteri *et al.*, 2005), and accounted for linkage disequilibrium by performing the analysis on windows of 30 contiguous SNPs (average physical distance = 308kb). Following pilot runs that allowed up to 10 mixing events, we performed 8 replicate runs each with 0-6 possible mixing events and examined how sequential addition of migration events altered the tree structure, model likelihood and matrix of residuals (visualized using TreeMix plotting functions in R). We obtained nodal support for the tree without migration by generating 100 bootstrap replicates with TreeMix and combining them using the *consense* function of Phylip 3.697 (Felsenstein, 1981).

Genomic signatures of differential local selection: In order to avoid confounding loci that exhibit high inter-population divergence due to differential local selection with those that exhibit the same pattern due to regional phylogeographic history, we analysed the Barents and Kola regional population groups separately. We took two approaches. First, we identified SNPs that were unusually different among populations when examining F_{ST} or other measures of differentiation across the genome ('outlier analysis'). Second, we looked for associations between SNP allele frequencies and an environmental variable previously identified as a selective factor for Atlantic salmon: upstream catchment area, a surrogate for expected river flow volume at the sampling site (Pritchard *et al.*, 2018, 'environmental association analysis'). As all recorded collection locations were close to the river mouth, we used total river catchment area as our measurement (Table 1).

Catchment areas were obtained from the online database of the Federal Water Resources Agency of Russia (<http://textual.ru/gvr/>, last accessed 25th July 2020) and mean standardized for analyses.

We aimed to compare overlap between locally adaptive candidates identified in this study with those identified in the previous Teno River study of Pritchard *et al.* (2018), which used individually genotyped fish. That study applied some statistical approaches that are not possible with population-level allele frequency data. Therefore, in order to more properly compare overlap in candidate regions among studies, we also re-analysed (where applicable) the Pritchard *et al.* (2018) dataset using the statistical approaches below.

Total number of SNPs retained after MAF filtering across each regional group and therefore used in the following analyses were: Barents, 187,262; Kola, 178,521; Teno, 198,829. Of these SNPs, 83.6% were shared by all three groups and 93.3% by at least two of the three groups. We retained SNPs that were not shared between the three regional groups of populations, as SNPs below the 0.05

MAF threshold in one group may nevertheless label adaptively important genetic variation within another group.

PCAdapt: To identify SNPs disproportionately contributing to the PCs separating the individual populations (outlier analysis), we ran *PCAdapt* as described above, specifying K equal to ($n_{\text{populations}} - 1$): Barents: K=7; Kola: K=15, Teno K=9.

BayPass: BayPass 2.1 (Gautier, 2015) implements the statistical approach originally described in (Coop *et al.*, 2010; Günther & Coop, 2013), and estimates a measure of deviation from underlying genome-wide population structure ($X^T X$), and measures of environmental association, correcting for population structure, for each SNP. We generated genomic variance-covariance matrices using subsets of SNPs obtained through approximate LD pruning, as described above (Barents, $n=28,885$; Kola, $n=27,533$; Teno, $n=33,175$; parameters: -npilot 20, -pilotlength 1000, -nthreads 8, all others default). We then performed five replicate runs of BayPass (parameters: -npilot 25, -pilotlength 2000, -nval 2000, -nthreads 8, all others default), each with an independently generated variance-covariance matrix. We used a different random number seed to initialize each instance of matrix generation and BayPass analysis. We used the median value of $X^T X$ (outlier analysis) or Pearson correlation coefficient r (environmental association analysis) over the five runs as our indicative variable for each SNP.

BayScEnv: BayScEnv 1.1 (de Villemereuil & Gaggiotti, 2015), is a modification of the Bayesian approach implemented in BayeScan (Foll & Gaggiotti, 2008) that allows examination of the effects of putative environmental selective agents. BayScEnv can implement a single model that simultaneously identifies SNP-specific effects driven by a target environmental variable and discriminates them from SNP-specific effects driven by other processes (including other forms of local selection). However, to obtain results that were comparable with those from our other analyses, we ran two separate models: one without an environmental effect (outlier analysis, equivalent to the BayeScan model, parameters: -pr_pref 1, -pr_jump 0.005, -nbp 12, -threads 16), and one examining only effects associated with upstream catchment area (environmental association analysis, parameters: -pr_pref 0, -pr_jump 0.005, -nbp 12, -threads 16).

LFMM: LFMM 1.4 (Frichot *et al.*, 2013) examines SNP - environmental associations while correcting for population structure using latent factor mixed model (environmental association analysis). We specified number of latent factors (K) equal to the number of populations, performed five replicate runs for each regional group, and took the median z-score over the replicate runs.

Evidence from combined tests: For each regional group of populations (Barents, Kola, Teno), we identified the genomic segments most likely to harbour loci under differential local selection by

combining evidence from the different statistical approaches over all SNPs, following Pritchard *et al.* (2018). For the outlier analyses we ranked SNPs independently by the three different test statistics (PCAdapt: pvalue; BayPass: $X^T X$; BayScEnv: qvalue(alpha)) and retained the top-ranked 0.5% for each. We then compared these three sets of top-ranked SNPs, and retained only the SNPs present in all three of them. Hereafter, we call these final retained SNPs ‘Outlier SNPs’ - there is one set of Outlier SNPs each for Barents, Kola and Teno. We used an identical strategy to retain candidate SNPs for the environmental association analysis (hereafter ‘GEA SNPs’; SNPs ranked by - BayPass: absolute Pearson’s R, BayScEnv: qvalue(g); LFMM: absolute z-score).

Divergent selection on a locus is expected to leave a signature on multiple linked SNPs, and the possibility of detecting this within a set of populations depends on many factors including marker density and the local recombination landscape. We therefore assessed whether different Outlier SNPs or GEA SNPs could be tagging the same selected locus by asking whether they occurred within the same haploblock, defined as a physically contiguous set of SNPs exceeding a specified linkage disequilibrium threshold. As haploblocks cannot be estimated from population-level allele frequency data, we used Plink 1.96 (Chang *et al.*, 2015) to infer haploblocks for all SNPs included in this study using a dataset of 883 individually genotyped fish from the Teno River (combined data from Barson *et al.*, 2015; Pritchard *et al.*, 2016, 2018). Parameters for haploblock estimation followed Pritchard *et al.*, 2018, where they were relaxed from Plink defaults because these often returned multiple small haploblocks breaking up a single selective sweep (`--no-small-max-span --blocks-inform-frac 0.8 --blocks-max-kb 5000 --blocks-strong-lowci 0.55 --blocks-strong-highci 0.85 --blocks-recomb-highci 0.8`). We defined haploblock boundaries as the positions halfway between the outermost haploblock SNPs and their closest non-haploblock SNP. We combined any neighbouring haploblocks containing retained SNPs and within 10kB of each other into a single block. We annotated each haploblock that contained an Outlier or GEA SNP with the overlapping NCBI coding genes using the *intersect* function of BEDTools.

RESULTS

Genetic variation & population genetic structure: Mean observed heterozygosity ranged from 0.259 to 0.358, and was generally lowest in the SW White Sea populations and highest in the Barents Sea populations (Table 1). We retained 7 PCs from the PCA of the LD-pruned 30-population dataset. The first two PCs (Fig. 2) arranged the populations geographically, with PC1 separating the Barents Sea samples from all other populations and PC2 separating the SW White Sea populations from all others. The remaining PCs (Fig. S1) separated individual rivers in the E and SW White Sea (Nilma, Pongoma, Pulonga, Suma, SD_Emtsa & Megra) from the rest of the populations. Thus,

overall, we observed two clusters of genetically more similar populations, hereafter ‘Barents’, (comprising the 8 rivers opening to the Barents Sea), and ‘Kola’ (comprising the 16 rivers on the Kola Peninsula opening into the White Sea), plus six outlying populations in the E and SW White Sea that were strongly differentiated from the two main regional populations clusters and from one another.

When the six outlying populations were removed, a single PC separating the Barents and Kola population clusters captured most of the variation in the dataset. PCAdapt identified 132 SNPs as contributing disproportionately ($q < 0.05$) to differentiation along this PC. These SNPs were distributed across 24 of the 29 chromosomes and associated with 80 overlapping or downstream protein coding genes and 7 documented structural variants (Fig. S2, Table S2).

TreeMix without migration identified two strongly supported clades in the extended 52-population dataset, one comprising the Baltic, Lake Onega and Lake Ladoga populations and the second including all Barents and White Sea populations and three additional landlocked populations (Luzhma, Pisto and Kamennaya, Fig. 3). When migration events were allowed, TreeMix inferred multiple alternative topographies across replicate runs for most values of m , making it difficult to select an unequivocally best model. However, all models inferred two long-distance migration events: one from the Baltic clade into the Karelian landlocked populations, and one from the Baltic into the Teno (Fig. 3). No model inferred migration from the Baltic into the White Sea populations.

Genomic signatures of local selection within regional population groups: Three hundred and five SNPs were retained as ‘Outlier SNPs’ in the Barents analysis, 318 in the Kola analysis and 201 in the Teno re-analysis (Table S3). These SNPs occurred within 263 independent haploblocks in Barents, 184 haploblocks in Kola, and 80 haploblocks in Teno (Table S3). Only four of these outlying haploblocks were shared between all three regional groups, while a further 13 were shared between two groups (Table 2, Fig. S3). In 10 of these 13 haploblocks, one or more Outlier SNPs were also shared between regional groups (Table 2); the other three shared haploblocks were labelled by different Outlier SNPs in the different groups. When examining the association between genomic variation and upstream catchment area far fewer SNPs were retained as ‘GEA SNPs’: 20 SNPs in Barents, within 17 haploblocks; 14 SNPs in Kola, within 13 haploblocks, and 123 SNPs in Teno, within 48 haploblocks (Table S3). Only a single catchment-associated haploblock was shared between regional groups; this haploblock, on chromosome Ssa11, was also a shared outlier (Table 2, Fig. S3).

DISCUSSION

Here, we used an allelotyping approach in combination with a 220K SNP array to characterize the population structure of Atlantic salmon in northeastern Europe, and ask whether the same loci emerge as locally adaptive candidates in independent analyses of different geographic regions. Our results largely recapitulated phylogeographic patterns that were previously inferred from much smaller marker sets. While most candidate loci were unique to a single geographic region, several genomic segments repeatedly showed strong evidence for differential local selection across all three regions. The results from this study validate allelotyping as a reliable and cost-effective alternative to individual genotyping when SNP genotyping arrays are available and the underlying population genetic structure is known.

Genome-wide SNPs recapitulate population structure inferred from other markers.

Analysis of genome-wide SNP allele frequencies reiterated the population genetic structure found in earlier studies of the Atlantic salmon of northeastern Europe using microsatellites (Tonteri *et al.* 2005, 2009) and mtDNA (Asplund *et al.* 2004). We observed a clear genetic transition between the Barents and White Seas, with White Sea populations (Kola, E White Sea & SW White Sea) being distinct from Barents populations in the PCA and forming a well-supported clade in the TreeMix analysis. This White Sea cluster was distant from the similarly well-supported Baltic/Onega/Ladoga clade, with no inference of gene flow between the two: thus, our results do not support previous hypotheses of a close phylogeographic connection between the White Sea and the Baltic Sea (Kazakov & Titov, 1991; Makhrov *et al.*, 2005). We observed further genetic differentiation between Kola Peninsula White Sea populations and E & SW White Sea populations. The SW White Sea populations in particular were strongly differentiated from one other and exhibited relatively low heterozygosity, suggesting that the observed structure in this region has been driven by population isolation and strong drift.

Our results are consistent with ancestral allopatry of Atlantic salmon in the Barents and White Seas, with subsequent gene flow between the two population clusters. In such a situation of secondary contact, genomic segments with elevated differentiation between the clusters may contain loci that are resistant to gene flow because they confer locally adaptive traits or otherwise reduce hybrid fitness (Wu, 2001). This pattern, however, can also emerge via other population genetic processes including linked purifying selection in areas of reduced recombination such as inversions (Cruickshank & Hahn, 2014; Lotterhos, 2019). Our analysis found segments of elevated differentiation between the Barents and Kola clusters to be distributed across 26 of the 29 chromosomes, with only a small number co-occurring with known inversions (Table S2). Further investigation of possible genomic barriers to gene flow between the Barents Sea and White Sea

Atlantic salmon lineages could improve our understanding of their recent evolutionary history and inform their management.

Strong candidates for differential local selection repeatedly emerge in replicated analyses.

When analysing the Barents, Kola and Teno datasets independently, we identified numerous haploblocks that were unusually differentiated among populations within each region and so potentially harboured locally adaptative loci. Only four of these outlying haploblocks, however, were shared between all three geographical regions. For three of these four haploblocks (on chromosomes Ssa09, Ssa12 and Ssa25) there is powerful additional evidence for the presence of a selective target. This includes observations of locally extended haplotype homozygosity in the previous Teno River analysis (Pritchard *et al.* 2018), and well-documented genotype-phenotype associations. The Ssa25 haploblock encompasses *vgll3-akap11*, a large-effect locus that strongly influences age at sexual maturity throughout the European range of Atlantic salmon and probably further afield (Ayllon *et al.*, 2015; Barson *et al.*, 2015; Kusche *et al.*, 2017). Age, and therefore size, at sexual maturity is expected to have varying effects on fitness depending upon the physical and biological environment a salmon encounters during return spawning run; thus, differential selection among rivers is expected. Correspondingly, the age-at-maturity distribution of males and females varies substantially among the rivers included in this study (Primmer *et al.*, 2006; Vähä *et al.*, 2007). The haploblock encompassing *six6*, on Ssa09, has emerged repeatedly as a candidate for differential local selection throughout the range of Atlantic salmon (Pritchard *et al.*, 2018). Variation in this genomic segment correlates with both age at maturity and seasonal timing of the return spawning migration (Barson *et al.*, 2015; Cauwelier *et al.*, 2017; Pritchard *et al.*, 2018). *Major histocompatibility complex II*, on the Ssa12 haploblock, is well known for its role in pathogen response. For example, in Atlantic salmon, genetic variation at this *MHCII*-containing segment is implicated in resistance to *piscine myocarditis virus* (Hillestad *et al.*, 2020). Although much previous work on *MHCII* has focused on maintenance of genetic variability by balancing selection, there is strong evidence of differential local selection on *MHCII* genotype in other salmonid species (McClelland *et al.*, 2013; Larson *et al.*, 2014, 2019).

In the Teno River Pritchard *et al.* (2018) found a correlation between *vgll3* and *six6* allele frequencies and upstream catchment area, a surrogate for river flow volume. This association was not observed in Barents or Kola, and over all we found rather few environmentally-associated haploblocks when analysing the allelotyped data. This may be due to the unbalanced nature of the Barents and Kola datasets, which contained many rivers with small catchment areas and a few with very large ones (Table 1). Only one genomic tract, on chromosome Ssa11, was associated with upstream catchment

area in more than one regional group; this haplotype, containing the genes *numa1* and *zfhx3*, was also a shared outlier. It has been observed to co-vary with seasonal migration timing in the Teno (Pritchard *et al.*, 2018), and is unusually differentiated between populations of Atlantic salmon in northern and southern Norway (Kjærner-Semb *et al.*, 2016).

Our results are likely to underestimate the true overlap between locally selected targets among the three different regional groups of populations that we examined. The identification of outliers by combining evidence from several different statistical approaches is expected to be biased towards loci that show the strongest signals of selection within the Barents, Kola or Teno datasets (Forester *et al.*, 2018). We also expect variation among datasets in the power we have to identify outliers due to the different number of populations examined, and the presence of many SNP markers linked to a single candidate locus which reduces the number of markers available to label other candidate loci when a fixed-rank cut-off for selection is applied. Additionally, in contrast to whole-genome sequence data, the average density of SNPs on the array (~1 SNP per 15kb) is unlikely to provide sufficient coverage of all possible selected variants. Further, in some cases, strong selection may have driven a variant to fixation in one of the geographic areas. For example, in Barents and Kola, we observe particularly strong evidence for differential local selection at a segment 18.8Mb along chromosome Ssa09 (Table 2). This locus, which closely flanks a genomic segment recently implicated in adaptation to a landlocked life history (Kjærner-Semb *et al.*, 2020) is barely variable within the Teno populations analysed. We note, however, that the vast majority of identified ‘candidate loci’, including those previously inferred in the Teno River study of Pritchard *et al.* (2018), are unique to one of the three independent analyses. We suggest that this reflects a combination of false positives and truly different genomic architecture of local adaptation among regions.

Our results add to the growing evidence that adaptive diversification does not only occur by polygenic changes but can also involve large phenotypic shifts driven by single genes (or tightly linked genetic clusters). Such large-effect loci, previously considered unlikely to contribute to complex locally adaptive traits, now been identified in a wide range of taxa (reviewed in Oomen *et al.*, 2020). Here, we have identified several genomic segments with evidence for strong differential local selection among neighbouring populations of salmon across a wide area of northern Europe, including two phylogeographically distinct lineages. Several of these segments have previously been identified as selective targets across the broader Atlantic salmon distribution (Pritchard *et al.* 2018) and even in other salmonid taxa (Pritchard *et al.* 2018, Veale & Russello, 2017, Larson *et al.* 2019). Although the expectation remains that most local adaptation is polygenic, an influence of large effect loci can greatly alter evolutionary dynamics (Oomen *et al.*, 2020). These strongly selected loci in Atlantic salmon are ideal targets for experimental validation of their role in local adaptation by

characterizing their phenotypic consequences and examining their influence on fitness in different environments.

ACKNOWLEDGMENTS

We would like to thank Jan Nilsson, Jaakko Erkinaro, Anti Vasemägi, and others who participated in samples collection. We are deeply grateful to Anni Tonteri and Mikhail Ozerov for extracting DNA for some of the samples more than 10 years ago. Special thanks to the reviewers for useful comments that improved the manuscript. This research has been funded by the Academy of Finland (grants 302873, 284941, 314254).

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: CRP, VLP, KJZ. Performed the experiments: KJZ.

Analysed the data: VLP, KJZ. Contributed reagents/materials/analysis tools: AEV, JL, CRP. Wrote the paper: KJZ, VLP, CRP. Commented on the manuscript: JL.

The authors have no conflict of interest to declare.

DATA ARCHIVING STATEMENT

Raw data and code used in the analyses are archived in Dryad Digital Repository:

<https://doi.org/10.5061/dryad.sn02v6x2t>

SUPPORTING INFORMATION

Supplementary Tables and Supplementary Figures can be found online at:

Three supplementary tables are included as separate .csv files:

Table S1: Details of the Atlantic salmon populations included in additional datasets: Teno river, Baltic Sea, Lake Ladoga, Lake Onega, and landlocked lakes. Population numbers (Key) match those on Figure 1.

Table S2: Details of the SNPs contributing to the differentiation along the PC separating Barents and Kola populations, as identified by PCAdapt.

Table S3: Haploblocks (HB) containing candidate locally selected SNPs. SNPs were retained based on outlier analysis or association with upstream catchment area.

REFERENCES

Asplund, T., Veselov, A., Primmer, C.R., Bakhmet, I., Potutkin, A., Titov, S., *et al.* 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. *Ann. Zool. Fenn.* **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. *Mol. Ecol.* **24**: 5158–5174.
- Ayllon, F., Kjærner-Semb, E., Furmanek, T., Wennevik, V., Solberg, M.F., Dahle, G., *et al.* 2015. The *vgl3* locus controls age at maturity in wild and domesticated Atlantic salmon (*Salmo salar* L.) males. *PLoS Genet.* **11**: e1005628.
- Barson, N.J., Aykanat, T., Hindar, K., Baranski, M., Bolstad, G.H., Fiske, P., *et al.* 2015. Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature* **528**: 405-408.
- Bernatchez, L., Renaut, S., Whiteley, A.R., Derome, N., Jeukens, J., Landry, L., *et al.* 2010. On the origin of species: Insights from the ecological genomics of lake whitefish. *Phil. Trans. R. Soc. B: Biological Sciences* **365**: 1783-1800.
- Bertolotti, A.C., Layer, R.M., Gundappa, M.K., Gallagher, M.D., Pehlivanoglu, E., Nome, T., *et al.* 2020. The structural variation landscape in 492 Atlantic salmon genomes. *bioRxiv* 2020.05.16.099614.
- Bourret, V., Kent, M.P., Primmer, C.R., Vasemägi, A., Karlsson, S., Hindar, K., *et al.* 2013. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Mol. Ecol.* **22**: 532–51.
- Bradbury, I.R., Hamilton, L.C., Dempson, B., Robertson, M.J., Bourret, V., Bernatchez, L., *et al.* 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. *Mol. Ecol.* **24**: 5130–5144.
- Buehler, D., Holderegger R., Brodbeck S., Schnyder E., Gugerli F. 2014. Validation of outlier loci through replication in independent data sets: a test on *Arabis alpina*. *Ecol. Evol.* **4**: 4296– 4306.
- Cauwelier, E., Gilbey, J., Sampayo, J., Stradmeyer, L. & Middlemas, S.J. 2017. Identification of a single genomic region associated with seasonal river return timing in adult Scottish Atlantic salmon (*Salmo salar* L.) identified using a genome-wide association study. *Can. J. Fish. Aquat. Sci.* **75**: 1427-1435.
- 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**, 1–16.

- Coop, G., Witonsky, D., Di Rienzo, A. & Pritchard, J.K. 2010. Using environmental correlations to identify loci underlying local adaptation. *Genetics* **185**: 1411–1423.
- Cruickshank, T.E. & Hahn, M.W. 2014. Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Mol. Ecol.* **23**: 3133–3157.
- de Villemereuil, P. & Gaggiotti, O.E. 2015. A new F_{ST} -based method to uncover local adaptation using environmental variables. *Methods Ecol. Evol.* **6**: 1248–1258.
- Dodson, J.J., Aubin-Horth, N., Thériault, V. & Páez, D.J. 2013. The evolutionary ecology of alternative migratory tactics in salmonid fishes. *Biol. Rev.* **88**: 602–625.
- Felsenstein, J. 1981. Evolutionary trees from DNA sequences: A maximum likelihood approach. *J. Mol. Evol.* **17**: 368–376.
- 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.
- Forester, B.R., Lasky, J.R., Wagner, H.H. & Urban, D.L. 2018. Comparing methods for detecting multilocus adaptation with multivariate genotype-environment associations. *Mol. Ecol.* **27**: 2215–2233.
- François, O., Martins, H., Caye, K. & Schoville, S.D. 2016. Controlling false discoveries in genome scans for selection. *Mol. Ecol.* **25**: 454–469.
- 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. *Mol. Biol. Evol.* **30**: 1687–1699.
- Gautier, M. 2015. Genome-wide scan for adaptive divergence and association with population-specific covariates. *Genetics* **201**: 1555–1579.
- Glover, K.J., Solberg, M.F., McGinnity, P., Hindar, K., Verspoor, E., Coulson, M.W., Hansen, M.M., Araki, H., Skaala, Ø. & Svåsand, T. 2017. Half a century of genetic interaction between farmed and wild Atlantic salmon: Status of knowledge and unanswered questions. *Fish Fish.* **18**: 890–927.
- Günther, T. & Coop, G. 2013. Robust identification of local adaptation from allele frequencies. *Genetics* **195**: 205–220.
- Haas, R.J. & Payseur, B.A. 2015. Fifteen years of genome-wide scans for selection: trends, lessons, and unaddressed genetic sources of complication. *Mol. Ecol.* **25**: 5–23.
- Hess, J.E., Zendt, J.S., Matala, A.R. & Narum, S.R. 2016. Genetic basis of adult migration timing in

anadromous steelhead discovered through multivariate association testing. *Proc. Biol. Sci.* **283**: 20153064.

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**: 608.

Hoban, S., Kelley, J.L., Lotterhos, K.E., Antolin, M.F., Bradburd, G., Lowry, D.B., *et al.* 2016. Finding the genomic basis of local adaptation: Pitfalls, practical solutions, and future directions. *Am. Nat.* **188**: 379–397.

Jensen, A. 1999. Cessation of the Norwegian drift net fishery: changes observed in Norwegian and Russian populations of Atlantic salmon. *ICES J. Mar. Sci.* **56**: 84–95.

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. *J. Fish Biol.* **39**: 1–6.

Kijas, J., Elliot, N., Kube, P., Evans, B., Botwright, N., King, H., *et al.* 2017. Diversity and linkage disequilibrium in farmed Tasmanian Atlantic salmon. *Anim. Genet.* **48**: 237–241.

King, T.L., Kalinowski, S.T., Schill, W.B., Spidle, A.P. & Lubinski, B.A. 2001. Population structure of Atlantic salmon (*Salmo salar* L.): A range-wide perspective from microsatellite DNA variation. *Mol. Ecol.* **10**: 807–821.

Kjærner-Semb, E., Ayllon, F., Furmanek, T., Wennevik, V., Dahle, G., Niemelä, E., *et al.* 2016. Atlantic salmon populations reveal adaptive divergence of immune related genes - a duplicated genome under selection. *BMC Genomics* **17**: 610.

Kjærner-Semb, E., Edvardsen, R.B., Ayllon, F., Vogelsang, P., Furmanek, T., Rubin, C.J., *et al.* 2020. Comparison of anadromous and landlocked Atlantic salmon genomes reveals signatures of parallel and relaxed selection across the Northern Hemisphere. *Evol. Appl.* eia.13129.

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. *Ecol. Evol.* **7**: 5797–5807.

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 major histocompatibility complex in ecotypes of sockeye salmon across Alaska. *Mol. Ecol.* **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. *Mol. Ecol.* **23**:

5448–5461.

- Lehnert, S.J., Bentzen, P., Kess, T., Lien, S., Horne, J.B., Clément, M., *et al.* 2019. Chromosome polymorphisms track trans-Atlantic divergence and secondary contact in Atlantic salmon. *Mol. Ecol.* **28**: 2074–2087.
- Lehnert, S.J., Kess, T., Bentzen, P., Clément, M. & Bradbury, I.R. 2020. Divergent and linked selection shape patterns of genomic differentiation between European and North American Atlantic salmon (*Salmo salar*). *Mol. Ecol.* **me.15480**.
- Liu, J., Shikano, T., Leinonen, T., Cano, J.M., Li, M.-H. & Merila, J. 2014. Identification of major and minor QTL for ecologically important morphological traits in three-spined sticklebacks (*Gasterosteus aculeatus*). *G3* **4**: 595–604.
- Lotterhos, K.E.K. 2019. The effect of neutral recombination variation on genome scans for selection. *G3* **9**:1851-1867.
- Lotterhos, K.E.K. & Whitlock, M.M.C. 2015. The relative power of genome scans to detect local adaptation depends on sampling design and statistical method. *Mol. Ecol.* **24**: 1031-1046.
- Luu, K., Bazin, E. & Blum, M.G.B. 2017. *pcadapt*: an R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Resour.* **17**: 67–77.
- Makhrov, A.A., Artamonova, V.S., Christoforov, O.L., Murza, I.G. and Altukhov, Y.P. 2004. Hybridization between Atlantic salmon *Salmo salar* L. and brown trout *S. trutta* L. upon artificial propagation. *Russ. J. Genet.* **40**: 1258–1263
- Makhrov, A.A., Verspoor, E., Artamonova, V.S. & O’Sullivan, M. 2005. Atlantic salmon colonization of the Russian Arctic coast: pioneers from North America. *J. Fish Biol.* **67**: 68–79.
- McClelland, E.K., Ming, T.J., Tabata, A., Kaukinen, K.H., Beacham, T.D., Withler, R.E., *et al.* 2013. Patterns of selection and allele diversity of class I and class II major histocompatibility loci across the species range of sockeye salmon (*Oncorhynchus nerka*). *Mol. Ecol.* **22**: 4783-4800.
- Oomen, R.A., Kuparinen, A. & Hutchings, J.A. 2020. Consequences of single-locus and tightly linked genomic architectures for evolutionary responses to environmental change. *bioRxiv* 2020.01.31.928770.
- Ozerov, M., Vähä, J.-P., Wennevik, V., Niemelä, E., Svenning, M.-A., Prusov, S., *et al.* 2017. Comprehensive microsatellite baseline for genetic stock identification of Atlantic salmon (*Salmo salar* L.) in northernmost Europe. *ICES J. Mar. Sci.* **74**: 2159–2169.
- Ozerov, M., Vasemägi, A., Wennevik, V., Niemelä, E., Prusov, S., Kent, M., *et al.* 2013. Cost-effective genome-wide estimation of allele frequencies from pooled DNA in Atlantic salmon

(*Salmo salar* L.). *BMC Genomics* **14**: 12.

- 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. *Can. J. Fish. Aquat. Sci.* **69**: 1947–1958.
- Ozerov, M.Y., 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 Genet.* **14**: 88.
- Pickrell, J.K., Pritchard, J.K., Cavalli-Sforza, L., Edwards, A., Felsenstein, J., Cann, R., *et al.* 2012. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet.* **8**: e1002967.
- Ponomareva, E.V. 2007. *Population Structure of Atlantic Salmon (Salmo salar L) from European North of Russia*. Doctoral Thesis Dissertation, Lomonosov Moscow State University, Moscow, 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. *Proc. KarRC RAS* **11**: 100-108. (In Russian; available <http://resources.krc.karelia.ru/krc/doc/publ2007/Trudy-KRC-Ussue-11.pdf>, last accessed 25th July 2020).
- Primmer, C.R., Veselov, A.J., Zubchenko, A., Poututkin, A., Bakhmet, I. & Koskinen, M.T. 2006. Isolation by distance within a river system: genetic population structuring of Atlantic salmon, *Salmo salar*, in tributaries of the Varzuga River in northwest Russia. *Mol. Ecol.* **15**: 653–666.
- Prince, D.J., O’Rourke, S.M., Thompson, T.Q., Ali, O.A., Lyman, H.S., Saglam, I.K., *et al.* 2017. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Sci. Adv.* **3**: e1603198.
- Pritchard, V.L., Erkinaro, J., Kent, M.P., Niemelä, E., Orell, P., Lien, S., *et al.* 2016. Single nucleotide polymorphisms to discriminate different classes of hybrid between wild Atlantic salmon and aquaculture escapees. *Evol. Appl.* **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. *Mol. Ecol.* **27**: 2560-2575.
- Quinlan, A.R. & Hall, I.M. 2010. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841-842.
- Quinn, T.P., McGinnity, P., Reed, T.E. & Bradford, M. 2016. The paradox of “premature

migration” by adult anadromous salmonid fishes: patterns and hypotheses. *Can. J. Fish. Aquat. Sci.* **73**: 1015-1030.

- R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL <https://www.R-project.org/>.
- Ravinet, M., Westram, A., Johannesson, K., Butlin, R., André, C. & Panova, M. 2015. Shared and non-shared genomic divergence in parallel ecotypes of *Littorina saxatilis* at a local scale. *Mol. Ecol.* **25**: 287-305
- Reelstab, C., Gugerli, F., Eckert, A.J., Hancock, A.M. and Holderegger, R. 2015. A practical guide to environmental association analysis in landscape genomics. *Mol. Ecol.* **24**: 4348-4370.
- Roesti, M., Hendry, A.P., Salzburger, W. & Berner, D. 2012. Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Mol. Ecol.* **21**: 2852–2862.
- Rougemont, Q., & Bernatchez, L. 2018. The demographic history of Atlantic Salmon (*Salmo salar*) across its distribution range reconstructed from Approximate Bayesian Computations. *Evolution* **72**, 1261-1277.
- Storey, J.D., Bass, A.J., Dabney, A. & Robinson, D. 2020. qvalue: Q-value estimation for false discovery rate control. R package version 2.20.0, <http://github.com/jdstorey/qvalue>.
- Sylvester, E.V.A., Beiko, R.G., Bentzen, P., Paterson, I., Horne, J.B., Watson, B., *et al.* 2018. Environmental extremes drive population structure at the northern range limit of Atlantic salmon in North America. *Mol. Ecol.* **27**: 4026–4040. Blackwell Publishing Ltd.
- Tonteri, A., Titov, S., Veselov, A., Zubchenko, A., Koskinen, M.T., Lesbarrères, D., *et al.* 2005. Phylogeography of anadromous and non-anadromous Atlantic salmon (*Salmo salar*) from northern Europe. *Ann. Zool. Fenn.* **42**: 1-22
- 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. *Can. J. Fish. Aquat. Sci.* **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. *Evol. Lett.* **2**: 355-367.
- Vähä, J.-P.P., Erkinaro, J., Niemelä, E. & Primmer, C.R. 2007. Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Mol. Ecol.* **16**: 2638-2654.
- Veale, A.J. & Russello, M.A. 2017. An ancient selective sweep linked to reproductive life history evolution in sockeye salmon. *Sci. Rep.* **7**: 1747.

- Verspoor, E., McCarthy, E.M., Knox, D., Bourke, E.A. & Cross, T.F. 1999. The phylogeography of European Atlantic salmon (*Salmo salar* L.) based on RFLP analysis of the ND1/16sRNA region of the mtDNA. *Biol. J. Linn. Soc.* **68**: 129-146.
- Weigland, H. & Leese, F. 2018. Detecting signatures of positive selection in non-model species using genomic data. *Zool. J. Linn. Soc.* 184: 528-583.
- 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. *Ecol. Evol.* **9**: 6901-6921.
- Wickham, H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York.
- Wu, C.I. 2001. The genic view of the process of speciation. *J. Evol. Biol.* **14**: 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 Genet.* **14**: e1007717.
- Zueva, K.J., Lumme, J., Veselov, A.E., Kent, M.P. & Primmer, C.R. 2018. Genomic signatures of parasite-driven natural selection in north European Atlantic salmon (*Salmo salar*). *Mar. Genomics* **39**: 26–38.
- Zueva, K.J., Lumme, J., Veselov, A.E., Kent, M.P., Lien, S. & Primmer, C.R. 2014. Footprints of directional selection in wild Atlantic salmon populations: evidence for parasite-driven evolution? *PLoS One* **9**: e91672.

Figure 1. Map of the sampling locations. Colours indicate populations analysed in detail in this study, while grey indicates populations added for the TreeMix analysis. Different shapes represent major geographic regions (circles: Barents & White Seas, squares: Baltic Sea, triangles: landlocked freshwater locations), while colours indicate sub-regions within the Barents & White Seas (green: Barents Sea; blue: Kola Peninsula; purple: South-West White Sea; yellow: East White Sea). See Pritchard *et al.* (2018) for the sampling locations for the reanalysed Teno River dataset.

Figure 2. Results of Principal Component Analysis. Position of each population is plotted along the first two PCs. Colours and population indices correspond to Figure 1 and Table 1 (green: Barents Sea; blue: Kola Peninsula; purple: SW White Sea; yellow: E White Sea).

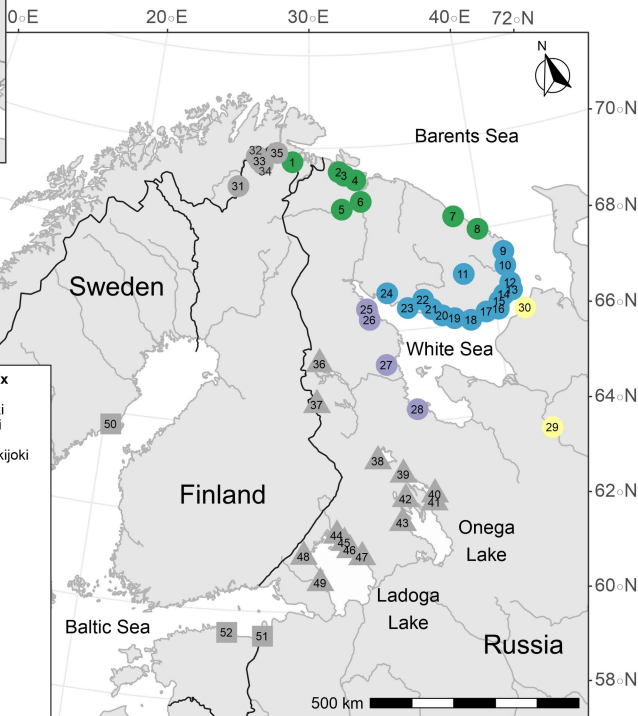
Figure 3: Results of the TreeMix analysis. Block colours show geographic regional groupings as in Figures 1 & 2. Numbers indicate nodal support based on 100 bootstrap trees without migration. Red arrows show long-distance gene flow events inferred by TreeMix, with numbers indicating migration weight. The tree is unrooted.

Table 1. Details of the studied Atlantic salmon populations: regional grouping, sample year and location, catchment area of the river, number of individuals pooled, and observed heterozygosity (H_0). Population numbers match those on Figure 1.

Key	River	Region	Lat	Long	Catchment area (km ²)	Sample year	DNA extracted	Pool size	H ₀
1	Näätämo	Barents Sea	69.7	28.98	2962	2006-08	2012	4 x 60	0.316
2	Titovka	Barents Sea	69.48	31.82	1226	2000	2006	38	0.343
3	Zapadnaya Litsa	Barents Sea	69.4	32.15	1688	2000	2015	43	0.348
4	Ura	Barents Sea	69.29	32.82	1029	2000	2015	44	0.358
5	Tuloma	Barents Sea	68.67	31.94	18231	1998	2015	40	0.352
6	Kola	Barents Sea	68.82	33.08	3846	2000	2015	40	0.351
7	Drozdivka	Barents Sea	68.33	38.42	468	2001	2006	48	0.314
8	Yokanga	Barents Sea	67.99	39.71	5944	2001	2015	39	0.353
9	Kachkovka	Kola Peninsula	67.43	40.95	843	2008	2008	66	0.349
10	Ponoi	Kola Peninsula	67.12	40.92	15467	2008	2008	83	0.314
11	Ponoi Lebyazia	Kola Peninsula	67.07	38.57	714	2001	2015	48	0.342
12	Danilovka	Kola Peninsula	66.74	41.02	262	2008	2015	48	0.327
13	Snezitsa	Kola Peninsula	66.58	40.69	235	2008	2008	25	0.333
14	Sosnovka	Kola Peninsula	66.5	40.58	582	2008	2008	47	0.314
15	Babya	Kola Peninsula	66.38	40.29	348	2008	2008	25	0.325
16	Lihodeevka	Kola Peninsula	66.33	40.17	308	2008	2008	53	0.337
17	Ust' Pyalka	Kola Peninsula	66.2	39.5	261	2008	2008	45	0.333
18	Strelna	Kola Peninsula	66.07	38.63	2774	2008	2008	64	0.311
19	Chavanga	Kola Peninsula	66.15	37.77	1212	2008	2008	42	0.334
20	Indera	Kola Peninsula	66.24	37.14	285	2008	2008	60	0.295
21	Varzuga	Kola Peninsula	66.4	36.62	9836	2008	2008	48	0.298
22	Yapoma	Kola Peninsula	66.62	36.20	180	2000	2015	34	0.328
23	Olenitsa	Kola Peninsula	66.47	35.33	403	2000	2015	46	0.309
24	Umba	Kola Peninsula	66.82	34.28	6248	2001	2001	44	0.331
25	Nilma	SW White Sea	66.5	33.13	164	2005	2005	39	0.265
26	Pulonga	SW White Sea	66.27	33.25	733	2005	2005	57	0.259
27	Pongoma	SW White Sea	65.28	34.00	1200	2005	2005	41	0.293
28	Suma	SW White Sea	64.28	35.40	2041	1999	2000	36	0.263
29	S. Dvina Emtsa	E White Sea	63.51	41.83	14100	2001	2015	42	0.323
30	Megra	E White Sea	66.15	41.57	2180	2001	2015	36	0.335

Table 2. Haploblocks containing candidate locally selected SNPs that were identified in more than one regional population group. Candidate SNPs identified from combined outlier analyses (“Outlier SNPs”) or associations with upstream catchment area (“GEA SNPs”). **Chr:** Chromosome; **HB_Start, HB_end:** start and end position of haploblock in bp; **# shared candidate SNPs:** how many SNPs in the haploblock were classed as candidate SNPs in two or more regional groups, compared to total number of shared SNPs in the haploblock; **Annotation:** annotated protein coding genes overlapping the haploblock.

Outlier SNPs	GEA SNPs	Chr	HB_Start	HB_End	# shared candidate SNPs	Annotation
Barents, Kola	na	Ssa03	8923406	9051536	1/19	<i>rahyl, gimap8</i>
Barents, Kola	na	Ssa09	18808558	18882390	0/19	<i>cd276, kiaa1522, rbbp4, zbtb80s</i>
Barents, Kola, Teno	Teno	Ssa09	24572313	24956797	28/55	<i>fgfr1l, tacc3, kif15, rd3, tdrd9, rtn1, lrrc9, pcnxc4, dhrs7, ppm1a, six6, pgbd4, six1</i>
Barents, Kola	na	Ssa09	70676323	70691232	1/1	<i>ppargc1b</i>
Kola, Teno	na	Ssa10	19245658	19414999	0/25	<i>polrmt, fgf10, mcpt, cfd</i>
Kola, Teno	Kola, Teno	Ssa11	19198258	19305775	2/9	<i>numa1, zfx3,</i>
Barents, Kola	na	Ssa12	25150145	25570229	1/23	<i>apob, prkca, cacng5,</i>
Barents, Kola	na	Ssa12	60985380	61035597	1/6	<i>slc41a1, etnk1, sox13</i>
Barents, Kola, Teno	na	Ssa12	61472809	61699297	6/45	<i>mdfi, tmem183a, ppfia1, tfeb, mhcll-dab,</i>
Kola, Teno	na	Ssa12	61735215	61901187	0/8	<i>perk8, foxp4</i>
Barents, Kola	na	Ssa13	71035682	72586328	0/52	<i>dlc1, sgcg, mrpl22, gemin5, cnot8, fam114a2, mfabp3, glnt10, cdhrs11, ine, flot2a, tbx6l, eral, fam222b</i>
Barents, Kola	na	Ssa18	58936002	58952501	1/2	<i>mfa3l</i>
Barents, Kola	na	Ssa20	3321666	3367040	1/12	<i>suds3</i>
Barents, Kola, Teno	na	Ssa20	45834673	46328021	1/69	<i>abnak, hipk4, pld3, fkb, irf2bp1, rpa34, nova1, gpr4, eml1, gcgr, cyp2m1, clip3, LOC106580757, akap12, tmem87b</i>
Barents, Kola	na	Ssa21	320945	642361	0/17	<i>mycbp2, fbxl3, cln5</i>
Barents, Kola	na	Ssa21	12187008	12190232	0/21	<i>LOC106581842</i>
Barents, Kola, Teno	na	Ssa25	28639461	28832410	0/20	<i>vglb3, akap11, tnfrsf11, epsti1, dnajc15</i>



Core Populations

- 1 - Näätämö
- 2 - Titovka
- 3 - Zapadnaya Litsa
- 4 - Ura
- 5 - Tuloma
- 6 - Kola
- 7 - Drozdovka
- 8 - Yokanga
- 9 - Kachkovka
- 10 - Ponoï
- 11 - Ponoï Lebyazia
- 12 - Danilovka
- 13 - Sneznitsa
- 14 - Sosnovka
- 15 - Babya
- 16 - Lihodeevka
- 17 - Ust' Pyalka
- 18 - Strelna
- 19 - Chavanga
- 20 - Indera
- 21 - Varzuga
- 22 - Yapoma
- 23 - Olenitsa
- 24 - Umba
- 25 - Nilma
- 26 - Pulonga
- 27 - Pongoma
- 28 - Suma
- 29 - S. Dvina Emtsa
- 30 - Megra

Added for TreeMix

- 31 - Teno Inarijoki
- 32 - Teno Tsarsjoki
- 33 - Teno Kevojoki
- 34 - Teno Utsjoki
- 35 - Teno Pulmankijoki
- 36 - Pisto
- 37 - Kamennaya
- 38 - Luzhma
- 39 - Kumsa
- 40 - Pyalma
- 41 - Tuba
- 42 - Lizma
- 43 - Shuya
- 44 - Sysky
- 45 - Uuksa
- 46 - Tulema
- 47 - Vidlitsa
- 48 - Hiitola
- 49 - Taipale
- 50 - Vindel
- 51 - Narva
- 52 - Kunda

Geographic Region

○ Barents & White Seas

- Barents Sea
- Kola Peninsula
- SW White Sea
- E White Sea

△ Landlocked lakes

□ Baltic Sea

