Kolarctic CBC – Project KO4178; Conserving our Atlantic salmon as a sustainable resource for people in the North; fisheries and conservation in the context of growing threats and a changing environment.

Report XXI. Genome-wide analysis of the temporal genetic variation of Atlantic salmon populations in Finnmark rivers, northern Norway

Ozerov M.Yu.[1,2], Wennevik V.[3], Niemelä E.[1,4,5], Vasemägi A.[6,7], Vorontsova T.[1], Kalske T.H.[4], Olsen A.A.[4], Høstmark M.S.[4], Smeland A.F.[4]

[1)]Biodiversity Unit, University of Turku, 20014 Turku, Finland
[2)]Department of Biology, University of Turku, 20014 Turku, Finland
[3)]Department of Aquaculture, Institute of Marine Research, Nordnes, PO Box 1870, 5817 Bergen, Norway
[4)]County Governor of Troms and Finnmark, 9800 Vadsø, Norway
[5)]Tmi Olli van der Meer, Hiomonkatu 14, 90850 Haukipudas, Finland
[6)]Department of Aquatic Resources, Institute of Freshwater Research, Swedish University of Agricultural Sciences, Stångholmsvägen 2, 17893 Drottningholm, Sweden
[7)]Chair of Aquaculture, Institute of Veterinary Medicine and Animal Sciences, Estonian University of Life Sciences, 46A Kreutzwaldi St., 51006 Tartu, Estonia

Corresponding author: Mikhail Ozerov, Biodiversity Unit, 20014 University of Turku, Finland, tel. +358 29 450 4241, e-mail: mikoze@utu.fi, mikhail.ozerov@gmail.com.

Statsforvalteren i Troms og Finnmark

PB 700, 9815 VADSØ

www.statsforvalteren/troms-finnmark/

CoASal – KO4178 project

ISBN 978-82-94021-21-5

Date: 31.1.2023

Report XXI. Genome-wide analysis of the temporal genetic variation of Atlantic salmon populations in Finnmark rivers, northern Norway

Ozerov M.Yu.[1,2], Wennevik V.[3], Niemelä E.[1,4,5], Vasemägi A.[6,7], Vorontsova T.[1], Kalske T.H.[4], Olsen A.A.[4], Høstmark M.S.[4], Smeland A.F.[4]

[1]Biodiversity Unit, University of Turku, 20014 Turku, Finland
[2]Department of Biology, University of Turku, 20014 Turku, Finland
[3]Department of Aquaculture, Institute of Marine Research, Nordnes, PO Box 1870, 5817 Bergen, Norway
[4]County Governor of Troms and Finnmark, 9800 Vadsø, Norway
[5]Tmi Olli van der Meer, Hiomonkatu 14, 90850 Haukipudas, Finland
[6]Department of Aquatic Resources, Institute of Freshwater Research, Swedish University of Agricultural Sciences, Stångholmsvägen 2, 17893 Drottningholm, Sweden
[7]Chair of Aquaculture, Institute of Veterinary Medicine and Animal Sciences, Estonian University of Life Sciences, 46A Kreutzwaldi St., 51006 Tartu, Estonia

Abstract

Atlantic salmon populations in Norway face a variety of threats, which in turn is compounded by climate change, affecting the arctic areas most. In northern Norway salmon populations' status remains relatively stable, however, the increasing pressure from farming industry in combination with growing mean annual water temperatures in rivers and at sea puts these populations at risk. Investigating temporally replicated samples may assist in the detection of changes in population genetic structure and diversity in time, which in turn aids in evaluation of the consequences of environmental changes, such as deterioration of habitat, climate change, introgression from hatchery supplementations or farmed escapees.

We screened twelve northern Atlantic salmon populations over 9 to 14 years using a high genome coverage SNP-array (> 60,000 SNPs) to evaluate the level of temporal genetic variation on genome-wide level. We aimed to identify SNPs showing significant allele frequency differences among temporally replicated samples within populations and characterize biological functions of those genomic regions.

The observed genome-wide temporal variation in the northernmost populations of Atlantic salmon indicated relatively stable genetic population structure in the majority of studied stocks. However, certain genomic regions were more variable that the others, particularly in populations of smaller stock census sizes, which might indicate both the effects of gene flow from farmed escapees and climate changes. Given fast growing and spreading of salmon aquaculture industry towards northern areas of Norway and fast increase of mean temperature in the Arctic, the generated genome-wide data will serve as a baseline to study genomic changes of salmon populations in the future.

Keywords:

Salmo salar, SNP, genome-wide, temporal genetic variation

Introduction

Atlantic salmon populations in Norway face a variety of threats, including hydropower regulation, degradation of freshwater habitats from land use practices, pollution, parasites, fish diseases, introduced and invasive species, interbreeding with farmed escapees, overexploitation in fisheries, etc. (Forseth et al. 2017). All of these factors are compounded by climate change, which affects the arctic areas most (Kaufman et al. 2009). In northern Norway salmon populations' status remains relatively stable, however the increasing pressure from farming industry (Glover et al. 2019, Fauske 2021) in combination with growing mean annual water temperatures in rivers and at sea (Ferré et al. 2012, Niemelä et al. 2022a, Niemelä et al. 2022b, c) puts these populations at risk. For example, while the spawning stock target achievement of the majority of populations in northern Norway is good, genetic integrity tends to deteriorate (https://lakseregisteret.fylkesmannen.no/). The spread of salmon aquaculture in the area during the last decade (Fauske 2021) and interbreeding of wild salmon with farmed escapees (Glover et al. 2019) is the main factor affecting the loss of genetic integrity in wild populations. Gene flow due to introgressive hybridization with aquaculture salmon causes changes of wild population genetic structure and variation (Glover et al. 2012), which in turn may lead to phenotypic changes and disrupt local adaptations and lead to loss of fitness (Bourret et al. 2011, Glover et al. 2017).

Investigating temporally replicated samples may assist in the detection of changes in population genetic structure and diversity in time, which in turn aids in evaluation of the consequences of environmental changes, such as deterioration of habitat (Yamamoto et al. 2004), climate change, introgression from hatchery supplementations (Vasemägi et al. 2005, Finnegan and Stevens 2008, Ozerov et al. 2016) or farmed escapees (Skaala et al. 2006, Bourret et al. 2011, Glover et al. 2012). The level of temporal variation as an indicator of population genetic integrity has been studied in salmon using various genetic markers including protein polymorphism (Mcelligott and Cross 1991, Jordan et al. 1992), microsatellites (Tessier and Bernatchez 1999, Skaala et al. 2006, Vähä et al. 2008, Glover et al. 2012, Ozerov et al. 2013) and SNPs (Östergren et al. 2021). However, while analyses of tens or hundreds of markers are able to reveal significant mean temporal changes in salmon population structure and variability, it does not allow drawing firm conclusions on which genes and/or regulatory genomic regions have been affected by contemporary natural or human-driven selection.

Advances in high-throughput sequencing technologies and SNP genotyping methods (Houston et al. 2014, Tsai et al. 2016, Wenne et al. 2016, Yáñez et al. 2016, Gao et al. 2020), as well as the availability of a highly continuous and annotated genome assembly (Lien et al. 2016) enable screening of tens to hundred thousands of markers across the Atlantic salmon genome. This allows to shed light on the patterns of genome-wide changes and assists in revealing genomic regions affected, e.g. by identifying genomic regions in wild fish, which are more or less resistant to genetic introgression from aquaculture escapees or hatchery stocking (Kovach et al. 2016, Ozerov et al. 2016), or identifying

genomic regions showing extensive genetic change linked to environmental changes (Kjærner-Semb et al. 2016, Gabián et al. 2022). For example, assessment of the effects of hatchery releases on wild salmon populations in the Gulf of Finland, Baltic Sea, showed that some genomic regions are more vulnerable to introgressive hybridization than the others, which, in turn, may have functional consequences for indigenous populations (Ozerov et al. 2016).

Recent studies have shown that about two thirds of Norwegian populations have been affected by farmed escapees (Glover et al. 2019) and in some rivers this proportion has reached up to 80% (Fiske et al. 2006, Karlsson et al. 2016). Atlantic salmon populations in the north are considered to be relatively pristine and less affected by genetic introgression of farmed genes. For example, recent reports showed that the proportion of aquaculture salmon can reach up to 2-4% of spawning populations in Finnmark rivers (Karlsson et al. 2016, Glover et al. 2019). However, whether some regions of the wild salmon genome are more susceptible for introgression is not known, there is an urgent need to evaluate the potential functional effects of farmed gene flow on wild populations. Furthermore, given a rapid spread of salmon aquaculture northward during recent years, it is important to establish a reliable genome-wide baseline for future evaluation of its consequences.

To evaluate the level of temporal genetic variation on genome-wide level we screened twelve northern Atlantic salmon populations over 9 to 14 years using a high genome coverage SNP-array (> 60,000 SNPs). We aimed to identify SNPs showing significant allele frequency differences among temporally replicated samples within populations and characterize biological functions of those genomic regions. Simultaneously, this work also generated genome-wide baseline of northernmost populations consisting of thousands of SNPs for future applications.

Material and methods

DNA samples

In total, 12 populations of Atlantic salmon representing previous (2006/2010/2011) and recent samples (2019/2020) were chosen for individual genotyping on SNP array (Fig. 1). Both the previous and recent samples were selected among DNA extracts used for the Atlantic salmon baseline generation consisting of juvenile fish (see details on DNA extraction, quality control and microsatellite genotyping in (Ozerov et al. 2017, Ozerov et al. 2022b). In brief, each sample was surveyed for genetic variation at 31 microsatellite DNA loci to ensure high quality of DNA extracts by excluding: a) brown trout or/and Atlantic salmon – brown trout hybrid individuals; and b) contaminated samples, i.e. the samples, which showed the presence of DNA from multiple individuals. Further, to remove the effect of family structure the sibship-reconstruction method implemented in Colony 2.0.6.6 (Jones and Wang 2010) was applied to test for full- and half-sib relationships in each population and all full-siblings except one pair per family were excluded from further analyses (see Ozerov et al. 2017; 2022). In addition, due to possible effect of long-term storage on DNA integrity, previous DNA extracts (2006/2010/2017) were examined

for degradation by visual inspection on 1% agarose gels. Samples containing low molecular weight DNA (indicative of degradation) were discarded from further analyses. In total, high quality DNA extracts of 1152 individuals, were included for SNP genotyping (Table 1).

SNP genotyping and quality control

The initial concentration of DNA extracts was first measured with the NanoDrop™ 1000 (Thermo Scientific) and subsequently diluted to 23 ng/ul. Diluted DNA extracts were analyzed using an Atlantic salmon Axiom™ Genotyping Array, containing probes for 60,251 SNPs, at the Centre for Integrative Genetics (CIGENE), Norway. Genotypes from samples showing a dish quality control (DQC) < 0.82 or call rate < 0.90 were discarded. Only those data from SNPs classified as "*Poly High Resolution*" were retained in the dataset (51,368 SNPs). SNPs with a minor allele frequency (MAF) < 0.05 were also removed. An exact test for Hardy-Weinberg equilibrium (HWE) based on 1000 Monte Carlo permutations of alleles (Guo and Thompson 1992) was performed for each temporal sample separately using hw.test function of pegas 1.1 package in R 4.1.3 (R Core Team 2021). False discovery rate (FDR) was estimated with (Benjamini and Hochberg 1995) procedure using p.adjust function of stats 3.6.2 package in R. SNPs showing deviation from HWE in > 50% of the temporal samples were discarded. Finally, SNPs with > 5% of missing genotypes were excluded. After applying these filters, data from 50,196 SNPs and 1102 individuals remained available for further analyses.

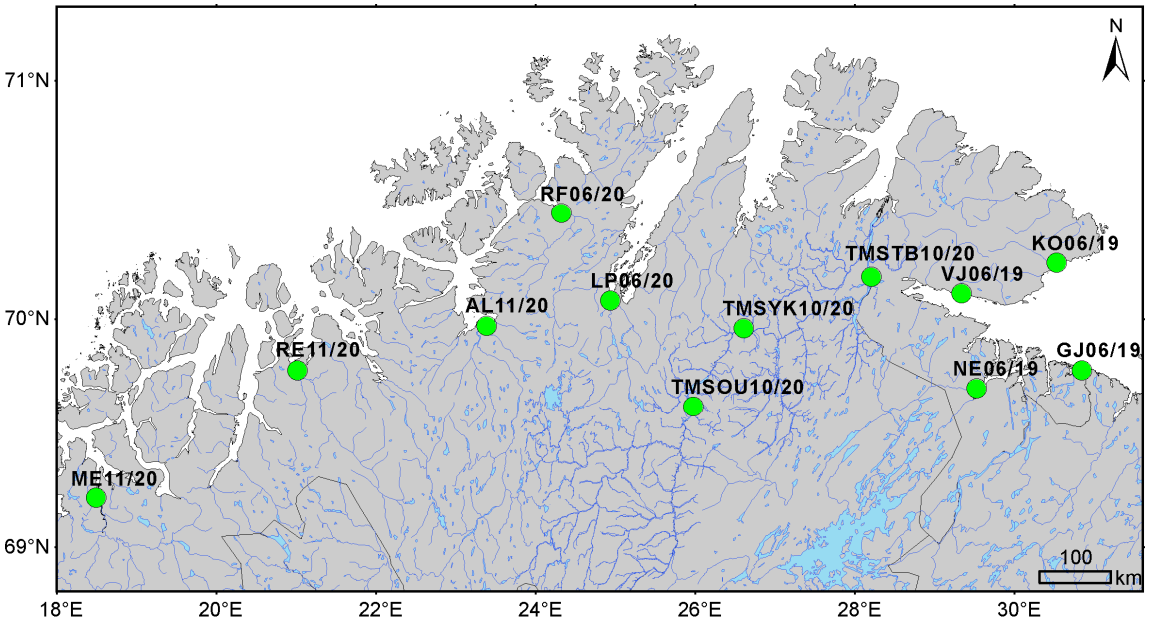

Fig. 1. Map showing geographical locations and sampling years of analysed Atlantic salmon populations.

Table 1. Studied Atlantic salmon populations; year of sampling; population *ID*; salmon population status shown as spawning stock target achievement and spawning potential (*SSTP*) and genetic integrity (*GI*); number (*n*) of analyzed and quality control passed (n_{QC}) samples; observed (H_O) and expected (H_E) heterozygosity; and inbreeding coefficient (*f*).

Population	***Year***	***ID***	***SSTP****	***GI****	***N***	***N_{QC}***	***QC passed, %***	***Ho***	***H_E***	***f***
Grense Jakobselv	2006	GJ06	NA	NA	48	48	100.0	0.397	0.391	-0.027
Grense Jakobselv	2019	GJ19	good	moderate	48	41	85.4	0.404	0.398	-0.005
Neidenelva	2006	NE06	NA	NA	48	48	100.0	0.390	0.389	0.008
Neidenelva	2019	NE19	very good	moderate	48	39	81.3	0.406	0.398	-0.036
Vestre Jakobselva	2006	VJ06	NA	NA	48	48	100.0	0.403	0.403	0.009
Vestre Jakobselva	2019	VJ19	very good	bad	48	48	100.0	0.400	0.404	0.017
Komagelva	2006	KO06	NA	NA	48	48	100.0	0.389	0.385	-0.016
Komagelva	2019	KO19	very good	bad	48	47	97.9	0.397	0.393	-0.002
Tana mainstem Tanabru	2010	TMSTB10	NA	NA	48	48	100.0	0.380	0.383	0.003
Tana mainstem Tanabru	2020	TMSTB20	very bad	good	48	46	95.8	0.389	0.385	-0.010
Tana mainstem Outakoski	2010	TMSOU10	NA	NA	48	48	100.0	0.384	0.387	0.005
Tana mainstem Outakoski	2020	TMSOU20	very bad	good	48	46	95.8	0.394	0.387	-0.011
Tana mainstem Yläköngäs	2010	TMSYK10	NA	NA	48	48	100.0	0.385	0.388	0.004
Tana mainstem Yläköngäs	2020	TMSYK20	very bad	good	48	37	77.1	0.397	0.389	-0.023
Lakselva Porsanger	2006	LP06	NA	NA	48	48	100.0	0.381	0.380	0.008
Lakselva Porsanger	2020	LP20	very good	moderate	48	47	97.9	0.386	0.386	0.010
Repparfjordelva	2006	RF06	NA	NA	48	47	97.9	0.398	0.400	-0.003
Repparfjordelva	2020	RF20	very good	moderate	48	37	77.1	0.410	0.398	-0.035
Altaelva	2010	AL10	NA	NA	48	46	95.8	0.385	0.388	0.009
Altaelva	2020	AL20	very good	moderate	48	48	100.0	0.386	0.386	-0.010
Reisaelva	2011	RE11	NA	NA	48	48	100.0	0.382	0.386	0.021
Reisaelva	2020	RE20	very bad	good	48	47	97.9	0.394	0.387	-0.032
Målselv	2011	ME11	NA	NA	48	48	100.0	0.410	0.410	0.005
Målselv	2020	ME20	very good	very bad	48	46	95.8	0.420	0.412	-0.022
Total					1152	1102	95.7	0.394	0.392	-0.006

*Data obtained from https://lakseregisteret.fylkesmannen.no/

Estimation of basic population genetics statistics

The R package adegenet 2.1.3 (Jombart 2008, Jombart and Ahmed 2011) was used to convert SNP data into a genind object. The basic descriptive statistics for each SNP locus and population (allelic richness, expected and observed heterozygosity) were calculated using basic.stats function of hierfstat 0.5-11 R-package (Goudet 2005). The same package was applied to estimate within-population inbreeding coefficients (FIS) and between-population pairwise fixation indices (F_{ST}; (Weir and Cockerham 1984) among populations using pairwise.WCfst function. Overall population genetic structure was examined by applying principal component analysis (PCA) using the dudi.pca function of the ade4 1.7-16 R-package (Dray and Dufour 2007). A consensus Neighbor-joining tree based on Nei's genetic distances (Nei 1972) and 1000 bootstrap replicates over loci was constructed using aboot function of R-package poppr 2.9.2 (Kamvar et al. 2014) and plotted using R-package ape 5.6-2 (Paradis et al. 2004, Paradis and Schliep 2019). Significance of genetic differentiation per locus between temporal samples within each population was estimated using exact G-test in Genepop 4.7.5 (Raymond and Rousset 1995, Rousset 2008) with the Markov chain parameters set at 100,000 dememorization steps, 100 batches, and 5,000 iterations per batch. P-values of locus-specific genetic differentiation between temporal samples along chromosomes were plotted using R-package CMPlot 3.7.0 (Yin et al. 2021). SNPs were classified as genome-wide significant if P-value was below the Bonferroni threshold for multiple testing (alpha = 0.05) of 0.05/50,196 (total number of SNPs genome-wide) and graded as chromosome-wide significant for P-values below the Bonferroni threshold estimated as 0.05/1,724 (average number of SNPs per chromosome). In our study genome-wide significant threshold was $P \leq 9.961 \times 10^{-7}$, which is equivalent to $-\log10(P) = 6.00$, while chromosome-wide significant threshold was $P \leq 2.900 \times 10^{-5}$, which is equivalent to $-\log10(P) = 4.54$.

SNP annotation and gene ontology (GO) analysis

SNPs were annotated with snpeff 5.0 (Cingolani et al. 2012) using ICSASG v2 Atlantic salmon reference genome sequence and annotation (NCBI: GCA_000233375.4). For GO analysis, salmon genes that are orthologous to human were identified using the rentrez package (Winter 2017) in R. Further candidate genes were classified to GO categories using panther 17.0 (Thomas et al. 2003).

Results

Mean population genetic estimates

The highest level of genetic diversity estimated overall loci was observed in Målselva both in earlier (H_{Oearl} = 0.410, H_{Eprev} = 0.410) and recent (H_{Oearl} = 0.420, H_{Eres} = 0.412) samples, whereas the lowest genetic diversity was observed in Tana Bru (H_{Oearl} = 0.380, H_{Eearl} = 0.383; H_{Ores} = 0.380, H_{Eres} = 0.386) and Lakselva Porsanger samples (H_{Oearl} = 0.381, H_{Eearl} = 0.383; H_{Ores} = 0.386, H_{Eres} = 0.385; Table 2; Fig. 2). Mean observed heterozygosity was higher among recent samples in 11 of 12

comparisons (Welch two sample t-test, $P < 0.05$), whereas mean expected heterozygosity was higher among recent samples in 6 of 12 comparisons (Welch two sample t-test $P < 0.05$; Table 2). Overall level of genetic diversity was slightly higher among recent samples compared to the earlier ones (H_{Oearl} = 0.381 vs. H_{Ores} = 0.386; H_{Eearl} = 0.383 vs. H_{Eres} = 0.385; both Welch two sample t-tests $P < 0.05$).

Table 2. Genetic differentiation ($F_{ST\ temp}$), mean observed ($H_{O\ diff}$) and expected ($H_{E\ diff}$) heterozygosity difference between earlier and recent samples and their respective *P*-values.

Population	$F_{ST\ temp}$	$H_{O\ diff}$	***P*-value** $H_{O\ diff}$	$H_{E\ diff}$	***P*-value** $H_{E\ diff}$
Grense Jakobselv	0.00703	0.007	0.000	0.007	0.000
Neidenelva	0.00485	0.016	0.000	0.009	0.000
Vestre Jakobselva	0.00609	-0.003	0.000	0.002	0.021
Komagelva	0.00382	0.008	0.000	0.008	0.000
Tana mainstem Tanabru	0.00077	0.009	0.000	0.001	0.081
Tana mainstem Outakoski	0.00268	0.011	0.000	0.002	0.051
Tana mainstem Yläköngäs	0.00073	0.010	0.000	0.000	0.935
Lakselva Porsanger	0.00457	0.011	0.000	-0.001	0.076
Repparfjordelva	0.00269	0.005	0.000	0.006	0.000
Altaelva	0.00002	0.001	0.093	-0.002	0.054
Reisaelva	0.00112	0.011	0.000	0.001	0.149
Målselv	0.00070	0.010	0.000	0.002	0.018

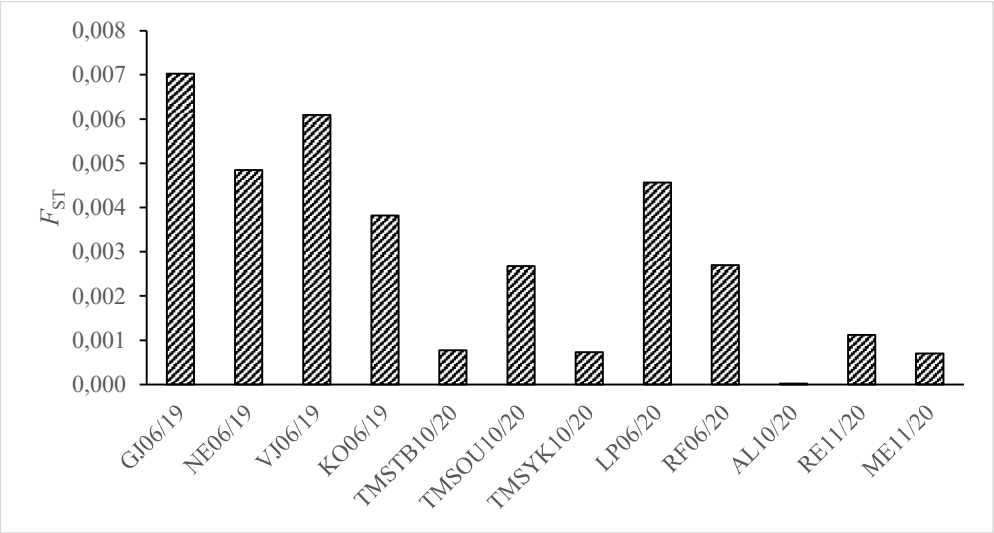

Fig. 2. Mean genetic divergence (F_{ST}) between temporal samples within populations estimated using 50,192 SNPs.

Genetic divergence between temporal samples within populations varied from F_{ST} = 0.00002 in Altaelva 2010 vs. 2020 to F_{ST} = 0.00703 in Grense Jakobselva 2006 vs. 2019 (Fig. 2), which was much lower than mean genetic divergence among earlier (F_{ST} = 0.030) or recent samples (F_{ST} = 0.027). Overall genetic variation due to temporal component within populations (0.29%) was 8.97 times lower than that

due to spatial component among populations (2.60%; Table 3). This indicates that studied populations have been relatively stable over time.

Table 3. Analysis of molecular variance (AMOVA) in the temporal samples of Atlantic salmon populations from northern Norway.

Source of variation	*Sum of squares*	*Percentage variation*	*P*
Among spatial samples	1336870	2.60	0.001
Among temporal samples	299395	0.29	0.001
Among individuals within populations	21087631	-0.54	0.778
Within individuals	21797696	97.65	0.001

Salmon populations subdivision on the neighbor-joining tree (Fig. 2) and PCA plot (Fig. 3) reflected their geographical origin. The temporal samples tended to cluster by the site of origin with a bootstrap support that varied between 99.8% and 100% (Fig. 2).

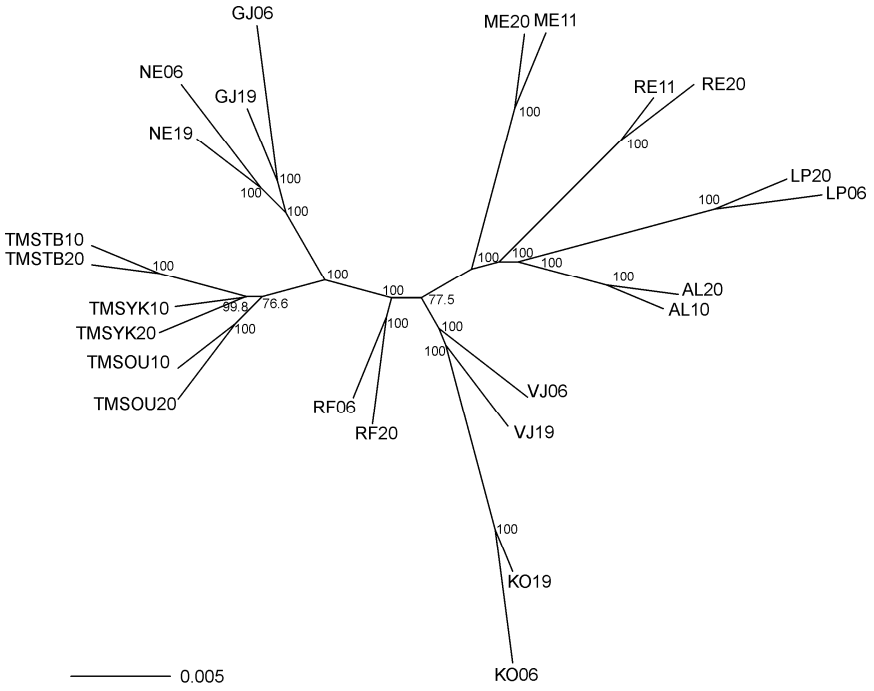

Fig. 2. Unrooted neighbor-joining tree based on Nei's genetic distances showing the genetic relationships between the temporally replicated samples of Atlantic salmon populations in northern Norway. The number on the nodes indicates the bootstrap values (percentage) obtained after 1000 replicates.

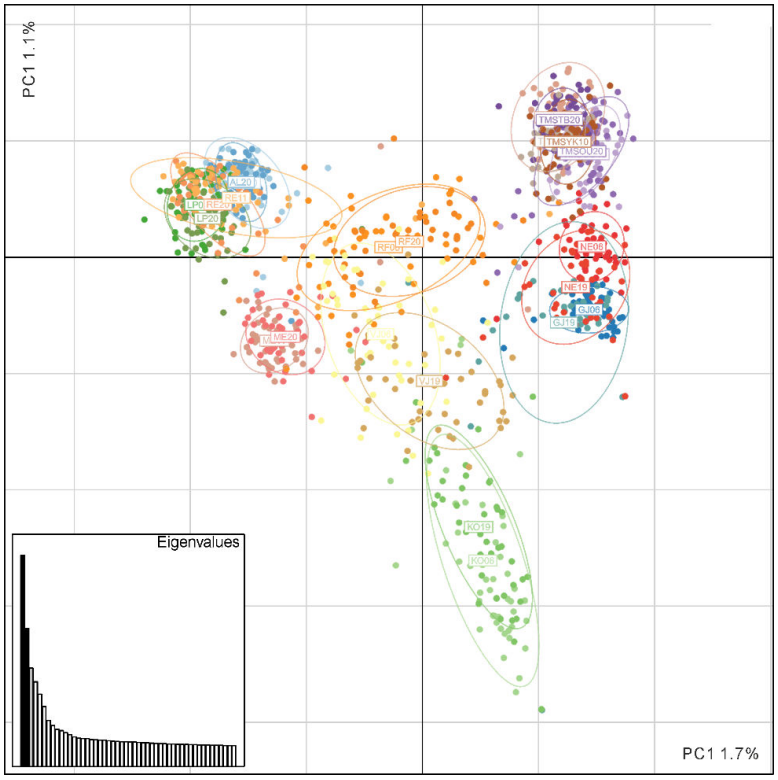

Fig. 3. PCA results of the temporally replicated samples of Atlantic salmon populations from northern Norway.

Candidate SNPs showing large temporal changes

In total, 54 candidate SNPs located in 22 out of 29 chromosomes showed significant genetic differentiation between temporal replicates after applying genome-wide significance threshold (Fig. 4). The highest number of SNPs showing significant genetic differentiation between temporal samples was detected in Grense Jakobselva (n = 13), Vestre Jakobselva (n = 14) and Lakselva Porsanger (n = 10) populations, whereas the lowest was observed in Alta (n = 0), Målselva (n = 1), Tana Yläköngäs (n = 1) and Tana Bru (n = 1) populations. In addition, the correlation between the number of candidate SNPs and genetic divergence between temporal samples within populations (F_{STtemp}) was significant (Pearson's $r = 0.876$, $P < 0.01$). None of the candidate SNPs were overlapped among different geographical locations most probably indicating that temporal changes of genetic structure were shaped by various factors, including environmental changes, different level of introgression by farmed escapees, selection and genetic drift. The majority of candidate SNPs were located in introns (27.9%), followed by intergenic (25.4%), 5K downstream (19.7%) and 5K upstream variants (14.8%; Fig. 5). In total, candidate SNPs were found within or nearby 30 Atlantic salmon genes. The majority of candidate genes were involved in cellular and metabolic processes, and biological regulation (GO biological process; Fig. 6), and in catalytic activity and binding (GO molecular function; Fig. 7).

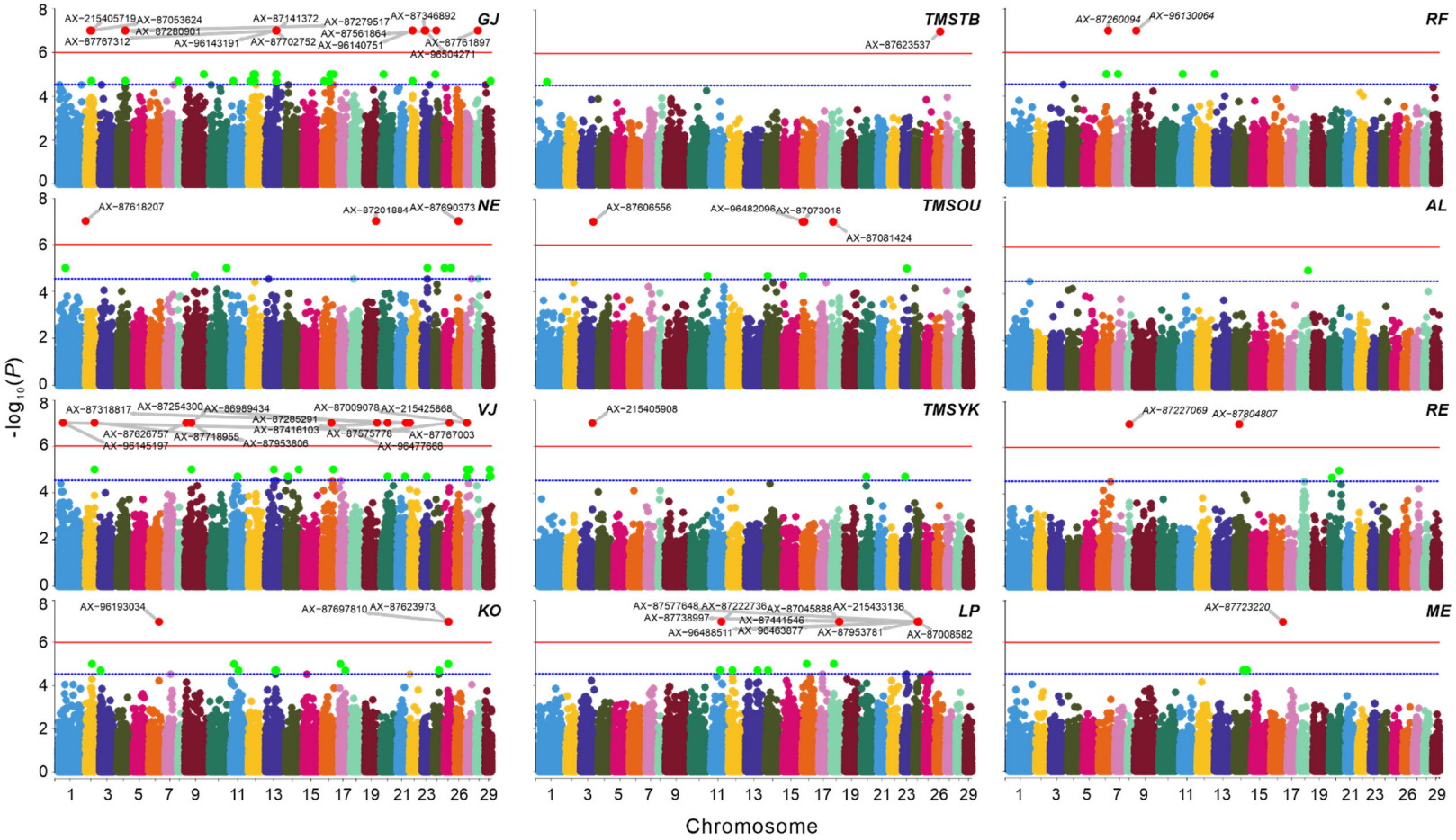

Fig. 4. The Manhattan plot showing highly divergent SNPs between earlier and recent samples in each Atlantic salmon population. The plot is based on -$\log_{10}$ (P-value) of G-test and imputation analysis against chromosome position, each colour represents different chromosome. Solid red and dashed blue lines indicate genome-wide ($P \leq 9.961 \times 10^{-7}$) and chromosome-wide ($P \leq 2.900 \times 10^{-5}$)significant threshold, genetically divergent SNPs passing these thresholds are shown as green and red dots, respectively.

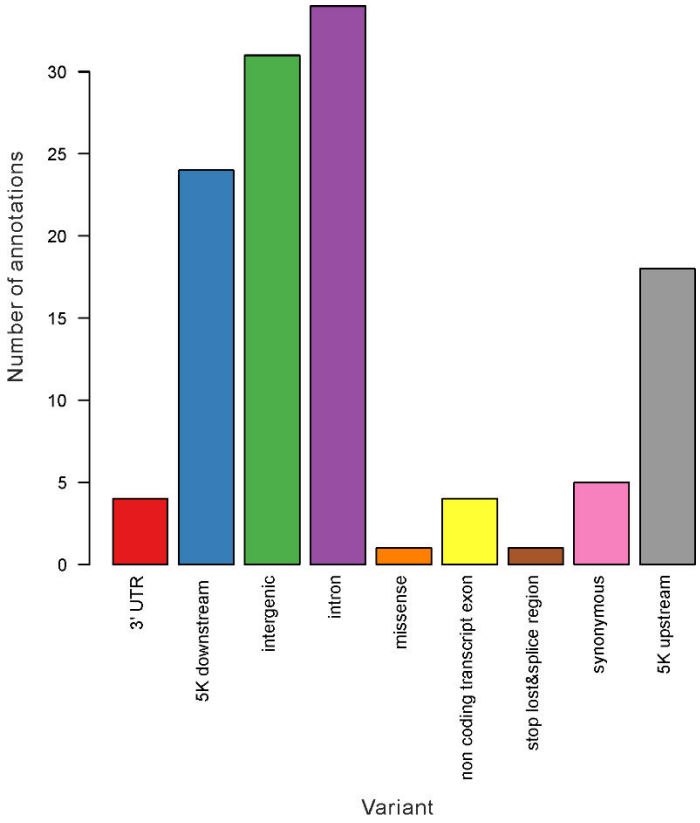

Fig. 5. The number of annotated candidate SNPs in each annotation category. The sum of SNPs per annotation category does not correspond to the total number of SNPs due to multiple annotations of some SNPs located in nearby (<5 K) or overlapped genes.

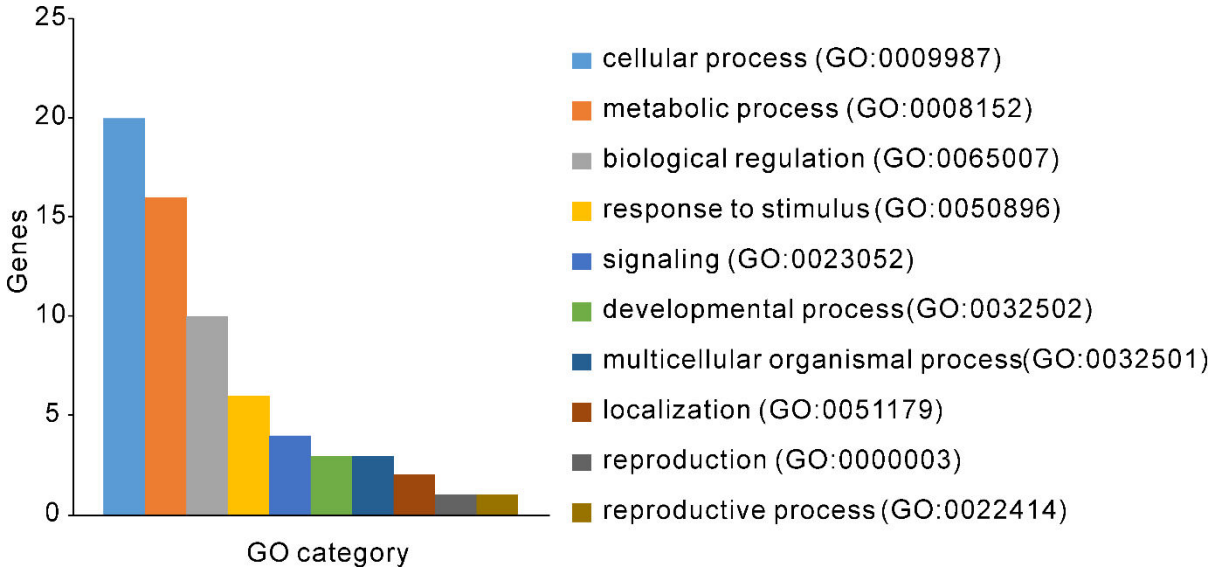

Fig. 6. Candidate genes classified using GO biological process.

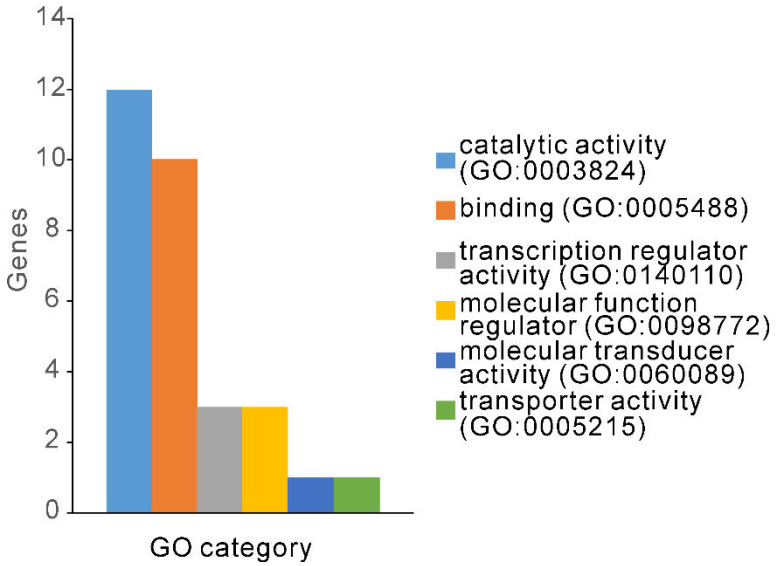

Fig. 7. Candidate genes classified using GO molecular function.

For example, substantial temporal changes of allele frequencies of SNPs located in a genomic region coding two genes (vestigial-like family member 3 gene, *vgll3* and A-kinase anchor protein 11, *akap11*) shown earlier to explain a large proportion in age at maturity variation (Ayllon et al. 2015, Barson et al. 2015), were observed in three populations (Komagelva, Tana mainstem Outakoski and Målselva; Fig. 6). The most drastic changes were detected in Komagelva, where fluctuation of allele frequencies of allele B reached up to 28% (Fig. 8). Further studies, however, required to evaluate the functional effect of these observations.

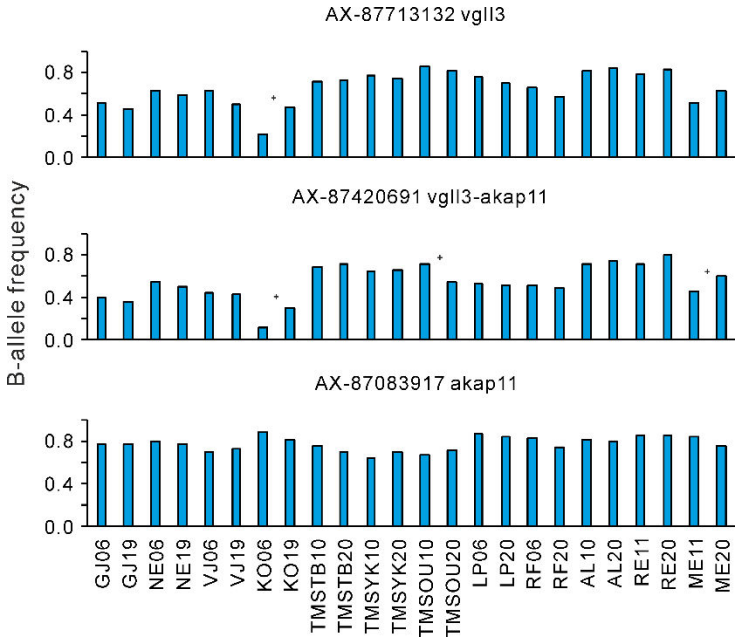

Fig. 8. Temporal changes of B-allele frequency at *vgll3-akap11* genes region in 12 Atlantic salmon populations. Substantial allele frequency changes are indicated with "+".

Discussion

In this report we assessed for fast contemporary changes at genome-wide level in 12 northernmost Atlantic salmon populations. We generated the genetic baseline of > 1,100 individuals genotyped for > 50,000 SNPs. Although the overall level of temporal variation within each population was low, we found a number of SNPs showing significant allele frequency differences among temporally replicated samples within populations and found those candidate SNPs associated with a number of genes.

Overall temporal variation and genetic structure

The overall temporal genetic variation among all studied populations was relatively stable, and supported by low level of genetic divergence between temporal samples within populations and by AMOVA results, which showed that genetic variation explained by the temporal component within populations was nearly nine times lower than that explained by the spatial component among populations. Similarly, grouping of temporal samples according to their river of origin as well as reflection of geographical pattern on both the neighbor-joining tree (Fig. 2) and PCA plot (Fig. 3) support the stability of the overall genetic structure among populations.

On the other hand, we observed a reduction of the overall level of genetic divergence among populations, from F_{ST} = 0.030 in 2006-2011 to F_{ST} = 0.027 in 2019-2020. In addition, we observed a modest, but significant, increase of the overall level of genetic diversity among the recent samples compared to the earlier ones. Moreover, mean H_E was significantly higher in six of 12 populations sampled in 2019-2020, and in three of those estimated genetic integrity status was classified as “bad” or “very bad”, whereas in the other three – as “moderate”. It should be noted, that no significant temporal changes of H_E were observed in four populations with “good” and two with “moderate” estimated genetic integrity status.

The observed pattern of overall decrease of genetic divergence and increase of genetic diversity among the recent samples might be shaped by possible gene flow from farmed salmon escapees and/or increased straying due to climate change. For example, decrease of genetic divergence among Norwegian salmon populations as a result of introgression from farmed stocks was shown in several studies (Skaala et al. 2006, Bourret et al. 2011, Glover et al. 2012, Glover et al. 2013, Glover et al. 2017). On the other hand, genetic diversity increase as a consequence of gene flow from farmed escapees contradicts the common knowledge that genetic diversity of aquaculture strains is often lower compared to that in wild populations (Ryman and Laikre 1991, Blanchet et al. 2008, Araki and Schmid 2010) and it is expected that increased introgression of farmed fish will reduce genetic diversity of wild populations (Eldridge et al. 2009, Bourret et al. 2011, Jasper et al. 2013). However, farmed escapees may carry new alleles not observed in wild population and thus increase genetic variability of wild salmon stocks (Verspoor 1998, Skaala et al. 2006), particularly those of small census size. Between-river migration of adults, or straying, also contributes to increased gene flow among wild populations

and may show similar patterns of decreased genetic divergence and increased genetic diversity with time. For example, increase of water temperature due to global climate change caused enhanced straying in southern European populations (Valiente et al. 2010, Horreo et al. 2011). Thus, we cannot exclude a synergistic effect of gene flow from farmed escapees and climate warming shaping the observed temporal patterns of genetic variation in northern populations of Atlantic salmon. However, further studies required to disentangle factors causing the observed changes of genetic structure and variability.

Candidate SNPs showing large temporal changes

While the observed changes of genetic variation overall SNPs were modest, we found 54 candidate SNPs, which showed significantly high genetic divergence between temporal replicates. However, the detected SNPs did not overlap among populations, indicating that various factors, such as genetic drift, environmental changes, degree of introgression from famed escapees, selection might have influenced temporal variation within geographical locations. On the other hand, the density of SNPs along the genome was not high, approximately one SNP per 54,916 bp, and the performed analyses might failed capturing temporal variation of non-covered genomic regions. Thus, more thorough analyses of the data is required to draw firm conclusions, including outlier tests, detection of runs of homozygosity and genome scan using sliding-window approaches. Nevertheless, a large proportion of candidate SNPs was found in regulatory regions (5K upstream, 5K downstream and 3'UTR), and only a few SNPs were located in protein coding sequences. This observation is corroborated by earlier studies indicated that selection predominantly influenced regulatory regions rather than protein coding sequences (Fraser 2013, Verta and Jones 2019, Fagny and Austerlitz 2021, Ozerov et al. 2022a).

Interestingly, we observed substantial changes of allele frequencies in *vgll3-akap11* gene region in three populations. This region is important in controlling age at maturity both in wild and domesticated salmon, with non-synonymous mutations in *vgll3* gene explained 33–36% of phenotypic variation (Ayllon et al. 2015, Barson et al. 2015). For example, Besnier et al. (2022) observed the decrease of age at maturity among salmon males in the river Etneelva population, which has been strongly affected by farmed escapees for many years. Therefore, we assume that temporal changes of allele frequencies at *vgll3-akap11* genomic region in the northernmost populations might be also due to increased gene flow from escaped farm salmon.

Conclusions

The observed genome-wide temporal variation in the northernmost populations of Atlantic salmon indicated relatively stable genetic population structure in the majority of studied stocks. However, certain genomic regions were more variable that the others, particularly in populations of smaller stock census sizes, which might indicate both the effects of gene flow from farmed escapees and climate changes. Given fast growing and spreading of salmon aquaculture industry towards northern areas of Norway and fast increase of mean temperature in Arctic area, the generated genome-wide data will

serve as a backbone to study genomic changes of salmon populations in the future. Due to delays caused by Covid-19 pandemics, more thorough analyses of the data were not performed in the frameworks of the Kolarctic ENPI CBC – CoASal project (KO4178), however, we plan to continue the exploration of the generated data in the future.

Acknowledgements

We thank Eevaliisa Kivilahti for assistance during fieldwork sampling and Abdullah Sami Madhun, Synne Winterthun and Johanne Øyro at IMR and Tutku Aykanat (University of Helsinki) for laboratory assistance. We acknowledge CSC-IT Center for Science, Finland for providing generous computational resources. This study was funded by the European Union, Kolarctic ENPI CBC project KO4178 "Conserving our Atlantic salmon as a sustainable resource for people in the North; fisheries and conservation in the context of growing threats and a changing environment" (CoASal). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the report. This publication has been produced with the assistance of the European Union, but the contents can in no way be taken to reflect the views of the European Union.

Literature

Araki, H., and C. Schmid. 2010. Is hatchery stocking a help or harm? Evidence, limitations and future directions in ecological and genetic surveys. Aquaculture **308**:S2-S11.

Ayllon, F., E. Kjærner-Semb, T. Furmanek, V. Wennevik, M. F. Solberg, G. Dahle, G. L. Taranger, K. A. Glover, M. S. Almén, C. J. Rubin, R. B. Edvardsen, and A. Wargelius. 2015. The vgll3 Locus Controls Age at Maturity in Wild and Domesticated Atlantic Salmon (Salmo salar L.) Males. Plos Genetics **11**:e1005628.

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

Benjamini, Y., and Y. Hochberg. 1995. Controlling the False Discovery Rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological) **57**:289-300.

Besnier, F., F. Ayllon, Ø. Skaala, M. F. Solberg, P. T. Fjeldheim, K. Anderson, S. Knutar, and K. A. Glover. 2022. Introgression of domesticated salmon changes life history and phenology of a wild salmon population. Evolutionary Applications **15**:853-864.

Blanchet, S., D. J. Paez, L. Bernatchez, and J. J. Dodson. 2008. An integrated comparison of captive-bred and wild Atlantic salmon (Salmo salar): Implications for supportive breeding programs. Biological Conservation **141**:1989-1999.

Bourret, V., P. T. O'Reilly, J. W. Carr, P. R. Berg, and L. Bernatchez. 2011. Temporal change in genetic integrity suggests loss of local adaptation in a wild Atlantic salmon (Salmo salar) population following introgression by farmed escapees. Heredity (Edinb) **106**:500-510.

Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen, L. Wang, S. J. Land, X. Lu, and D. M. Ruden. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. Fly **6**:80-92.

Dray, S., and A.-B. Dufour. 2007. The ade4 package: Implementing the duality diagram for ecologists. 2007 **22**:20.

Eldridge, W. H., J. M. Myers, and K. A. Naish. 2009. Long-term changes in the fine-scale population structure of coho salmon populations (Oncorhynchus kisutch) subject to extensive supportive breeding. Heredity (Edinb) **103**:299-309.

Fagny, M., and F. Austerlitz. 2021. Polygenic adaptation: Integrating population genetics and gene regulatory networks. Trends in Genetics **37**:631-638.

Fauske, M. 2021. Key figures from Norwegian aquaculture industry 2020. Fiskeridirektoratet.

Ferré, B., J. Mienert, and T. Feseker. 2012. Ocean temperature variability for the past 60 years on the Norwegian-Svalbard margin influences gas hydrate stability on human time scales. Journal of Geophysical Research: Oceans **117**.

Finnegan, A. K., and J. R. Stevens. 2008. Assessing the long-term genetic impact of historical stocking events on contemporary populations of Atlantic salmon, Salmo salar. Fisheries Management and Ecology **15**:315-326.

Fiske, P., R. A. Lund, and L. P. Hansen. 2006. Relationships between the frequency of farmed Atlantic salmon, Salmo salar L., in wild salmon populations and fish farming activity in Norway, 1989–2004. ICES Journal of Marine Science **63**:1182-1189.

Forseth, T., B. T. Barlaup, B. Finstad, P. Fiske, H. Gjøsæter, M. Falkegård, A. Hindar, T. A. Mo, A. H. Rikardsen, E. B. Thorstad, L. A. Vøllestad, and V. Wennevik. 2017. The major threats to Atlantic salmon in Norway. ICES Journal of Marine Science **74**:1496-1513.

Fraser, H. B. 2013. Gene expression drives local adaptation in humans. Genome Research **23**:1089-1096.

Gabián, M., P. Morán, M. Saura, and A. Carvajal-Rodríguez. 2022. Detecting Local Adaptation between North and South European Atlantic Salmon Populations. Biology **11**:933.

Gao, G., M. R. Pietrak, G. S. Burr, C. E. Rexroad, 3rd, B. C. Peterson, and Y. Palti. 2020. A New Single Nucleotide Polymorphism Database for North American Atlantic Salmon Generated Through Whole Genome Resequencing. Front Genet **11**:85.

Glover, K. A., C. Pertoldi, F. Besnier, V. Wennevik, M. Kent, and O. Skaala. 2013. Atlantic salmon populations invaded by farmed escapees: quantifying genetic introgression with a Bayesian approach and SNPs. BMC Genetics **14**:74.

Glover, K. A., M. Quintela, V. Wennevik, F. Besnier, A. G. Sorvik, and O. Skaala. 2012. Three decades of farmed escapees in the wild: a spatio-temporal analysis of Atlantic salmon population genetic structure throughout Norway. PLoS One **7**:e43129.

Glover, K. A., M. F. Solberg, P. McGinnity, K. Hindar, E. Verspoor, M. W. Coulson, M. M. Hansen, H. Araki, Ø. Skaala, and T. Svåsand. 2017. Half a century of genetic interaction between farmed and wild Atlantic salmon: Status of knowledge and unanswered questions. Fish and Fisheries **18**:890-927.

Glover, K. A., K. Urdal, T. Næsje, H. Skoglund, B. Florø-Larsen, H. Otterå, P. Fiske, M. Heino, T. Aronsen, H. Sægrov, O. Diserud, B. T. Barlaup, K. Hindar, G. Bakke, I. Solberg, H. Lo, M. F. Solberg, S. Karlsson, Ø. Skaala, A. Lamberg, Ø. Kanstad-Hanssen, R. Muladal, O. T. Skilbrei, and V. Wennevik. 2019. Domesticated escapees on the run: the second-generation monitoring programme reports the numbers and proportions of farmed Atlantic salmon in > 200 Norwegian rivers annually. ICES Journal of Marine Science **76**:1151-1161.

Goudet, J. 2005. hierfstat, a package for r to compute and test hierarchical F-statistics. Molecular Ecology Notes **5**:184-186.

Guo, S. W., and E. A. Thompson. 1992. Performing the exact test of Hardy-Weinberg proportion for multiple alleles. Biometrics **48**:361-372.

Horreo, J. L., G. Machado-Schiaffino, F. Ayllon, A. M. Griffiths, D. Bright, J. R. Stevens, and E. Garcia-Vazquez. 2011. Impact of climate change and human-mediated introgression on southern European Atlantic salmon populations. Global Change Biology **17**:1778-1787.

Houston, R. D., J. B. Taggart, T. Cézard, M. Bekaert, N. R. Lowe, A. Downing, R. Talbot, S. C. Bishop, A. L. Archibald, J. E. Bron, D. J. Penman, A. Davassi, F. Brew, A. E. Tinch, K. Gharbi, and A. Hamilton. 2014. Development and validation of a high density SNP genotyping array for Atlantic salmon (Salmo salar). BMC Genomics **15**:90.

Jasper, J. R., C. Habicht, S. Moffitt, R. Brenner, J. Marsh, B. Lewis, E. Creelman Fox, Z. Grauvogel, S. D. Rogers Olive, and W. S. Grant. 2013. Source-sink estimates of genetic introgression show influence of hatchery strays on wild chum salmon populations in Prince William Sound, Alaska. PLoS One **8**:e81916.

Jombart, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. Bioinformatics **24**:1403-1405.

Jombart, T., and I. Ahmed. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. Bioinformatics **27**:3070-3071.

Jones, O. R., and J. Wang. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. Molecular Ecology Resources **10**:551-555.

Jordan, W. C., A. F. Youngson, D. W. Hay, and A. Ferguson. 1992. Genetic Protein Variation in Natural Populations of Atlantic Salmon (Salmo salar) in Scotland: Temporal and Spatial Variation. Canadian Journal of Fisheries and Aquatic Sciences **49**:1863-1872.

Kamvar, Z. N., J. F. Tabima, and N. J. Grünwald. 2014. Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. PeerJ **2**:e281.

Karlsson, S., O. H. Diserud, P. Fiske, and K. Hindar. 2016. Widespread genetic introgression of escaped farmed Atlantic salmon in wild salmon populations. ICES Journal of Marine Science **73**:2488-2498.

Kaufman, D. S., D. P. Schneider, N. P. McKay, C. M. Ammann, R. S. Bradley, K. R. Briffa, G. H. Miller, B. L. Otto-Bliesner, J. T. Overpeck, and B. M. Vinther. 2009. Recent warming reverses long-term arctic cooling. Science **325**:1236-1239.

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

Kovach, R. P., B. K. Hand, P. A. Hohenlohe, T. F. Cosart, M. C. Boyer, H. H. Neville, C. C. Muhlfeld, S. J. Amish, K. Carim, S. R. Narum, W. H. Lowe, F. W. Allendorf, and G. Luikart.

2016. Vive la résistance: genome-wide selection against introduced alleles in invasive hybrid zones. Proc Biol Sci **283**.

Lien, S., B. F. Koop, S. R. Sandve, J. R. Miller, M. P. Kent, T. Nome, T. R. Hvidsten, J. S. Leong, D. R. Minkley, A. Zimin, F. Grammes, H. Grove, A. Gjuvsland, B. Walenz, R. A. Hermansen, K. von Schalburg, E. B. Rondeau, A. Di Genova, J. K. A. Samy, J. Olav Vik, M. D. Vigeland, L. Caler, U. Grimholt, S. Jentoft, D. Inge Våge, P. de Jong, T. Moen, M. Baranski, Y. Palti, D. R. Smith, J. A. Yorke, A. J. Nederbragt, A. Tooming-Klunderud, K. S. Jakobsen, X. Jiang, D. Fan, Y. Hu, D. A. Liberles, R. Vidal, P. Iturra, S. J. M. Jones, I. Jonassen, A. Maass, S. W. Omholt, and W. S. Davidson. 2016. The Atlantic salmon genome provides insights into rediploidization. Nature **533**:200-205.

Mcelligott, E. A., and T. F. Cross. 1991. Protein variation in wild Atlantic salmon, with particular reference to southern Ireland. Journal of Fish Biology **39**:35-42.

Nei, M. 1972. Genetic Distance between Populations. Am Nat **106**:283-292.

Niemelä, E., E. Hassinen, T. Kalske, and M. Høstmark. 2022a. REPORT II. Graphs indicating long-term sea temperatures in the coastal areas in Northern Norway and in Northern Kola Peninsula (Russia); monthly and annual fluctuations with clear warming trends, simultaneous changes between monitoring areas and affects to ice break- up dates and timing of salmon catches in the River Tana watershed, Kolarctic ENPI CBC – CoASal project (KO4178) report, ISBN 978-82-94021-01-7, 38 pp.

Niemelä, E., E. Hassinen, J. Kuusela, T. Kalske, and M. Høstmark. 2022b. REPORT I. Changing weather conditions, air temperature and precipitation, are driving annual and long-term changes in subarctic salmon environment and in the timing of salmon migrations in the River Tana watershed through air and water temperature synchrony, Kolarctic ENPI CBC – CoASal project (KO4178) report, ISBN 978-82-94021-00-0, 53 pp.

Niemelä, E., E. Hassinen, J. Kuusela, T. Kalske, and M. Høstmark. 2022c. REPORT III. Summary from the long-term mean monthly river temperatures and monthly cumulative temperatures as an index from the changes in juvenile salmon freshwater environment; graphical illustrations from the temperature variations within and between the rivers in Northern Norway and Finland, Kolarctic ENPI CBC – CoASal project (KO4178) report, ISBN 978-82-94021-02-4, 23 pp.

Östergren, J., S. Palm, J. Gilbey, G. Spong, J. Dannewitz, H. Königsson, J. Persson, and A. Vasemägi. 2021. A century of genetic homogenization in Baltic salmon—evidence from archival DNA. Proceedings of the Royal Society B: Biological Sciences **288**:20203147.

Ozerov, M., K. Noreikiene, S. Kahar, M. Huss, A. Huusko, T. Kõiv, M. Sepp, M.-E. López, A. Gårdmark, R. Gross, and A. Vasemägi. 2022a. Whole-genome sequencing illuminates multifaceted targets of selection to humic substances in Eurasian perch. Molecular Ecology **31**:2367-2383.

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

Ozerov, M., V. Wennevik, E. Niemelä, A. Vasemägi, S. Prusov, T. Vorontsova, T. Kalske, A. Olsen, M. Høstmark, and A. Smeland. 2022b. Application of the updated genetic baseline for genetic stock identification of Atlantic salmon in commercial fisheries in northern Norway, Kolarctic ENPI CBC – CoASal project (KO4178) report, 43 pp.

Ozerov, M. Y., R. Gross, M. Bruneaux, J. P. Vähä, O. Burimski, L. Pukk, and A. Vasemägi. 2016. Genomewide introgressive hybridization patterns in wild Atlantic salmon influenced by inadvertent gene flow from hatchery releases. Molecular Ecology **25**:1275-1293.

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

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of Phylogenetics and Evolution in R language. Bioinformatics **20**:289-290.

Paradis, E., and K. Schliep. 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics **35**:526-528.
R Core Team. 2021. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
Raymond, M., and F. Rousset. 1995. Genepop (Version-1.2) - Population-genetics software for exact tests and ecumenicism. Journal of Heredity **86**:248-249.
Rousset, F. 2008. genepop'007: a complete re-implementation of the genepop software for Windows and Linux. Mol Ecol Resour **8**:103-106.
Ryman, N., and L. Laikre. 1991. Effects of supportive breeding on the genetically effective populaiton size. Conserv Biol **5**:325-329.
Skaala, O., V. Wennevik, and K. Glover. 2006. Evidence of temporal genetic change in wild Atlantic salmon, Salmo salar L., populations affected by farm escapees. ICES Journal of Marine Science **63**:1224-1233.
Tessier, N., and L. Bernatchez. 1999. Stability of population structure and genetic diversity across generations assessed by microsatellites among sympatric populations of landlocked Atlantic salmon (Salmo salar L.). Molecular Ecology **8**:169-179.
Thomas, P. D., A. Kejariwal, M. J. Campbell, H. Mi, K. Diemer, N. Guo, I. Ladunga, B. Ulitsky-Lazareva, A. Muruganujan, S. Rabkin, J. A. Vandergriff, and O. Doremieux. 2003. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. Nucleic acids research **31**:334-341.
Tsai, H. Y., D. Robledo, N. R. Lowe, M. Bekaert, J. B. Taggart, J. E. Bron, and R. D. Houston. 2016. Construction and Annotation of a High Density SNP Linkage Map of the Atlantic Salmon (Salmo salar) Genome. G3 (Bethesda) **6**:2173-2179.
Vähä, J.-P., J. Erkinaro, E. Niemelä, and C. R. Primmer. 2008. Temporally stable genetic structure and low migration in an Atlantic salmon population complex: implications for conservation and management. Evolutionary Applications **1**:137-154.
Valiente, A. G., E. Beall, and E. V. A. Garcia-Vazquez. 2010. Population genetics of south European Atlantic salmon under global change. Global Change Biology **16**:36-47.
Vasemägi, A., R. Gross, T. Paaver, M. L. Koljonen, and J. Nilsson. 2005. Extensive immigration from compensatory hatchery releases into wild Atlantic salmon population in the Baltic sea: spatio-temporal analysis over 18 years. Heredity (Edinb) **95**:76-83.
Verspoor, E. 1998. Genetic impacts on wild Atlantic salmon (Salmo salar L .) stocks from escaped farm conspecifics : An assessment of risk. Ottawa.
Verta, J.-P., and F. C. Jones. 2019. Predominance of cis-regulatory changes in parallel expression divergence of sticklebacks. eLife **8**:e43785.
Weir, B. S., and C. C. Cockerham. 1984. Estimating F-statistics for the analysis of population structure. Evolution **38**:1358-1370.
Wenne, R., A. Drywa, M. Kent, K. K. Sundsaasen, and S. Lien. 2016. SNP Arrays for Species Identification in Salmonids. Methods in Molecular Biology **1452**:97-111.
Winter, D. J. 2017. rentrez: An R package for the NCBI eUtils API. The R Journal **9**:520-526.
Yamamoto, S., K. Morita, I. Koizumi, and K. Maekawa. 2004. Genetic Differentiation of White-Spotted Charr (Salvelinus leucomaenis) Populations After Habitat Fragmentation: Spatial–Temporal Changes in Gene Frequencies. Conservation Genetics **5**:529-538.
Yáñez, J. M., S. Naswa, M. E. López, L. Bassini, K. Correa, J. Gilbey, L. Bernatchez, A. Norris, R. Neira, J. P. Lhorente, P. S. Schnable, S. Newman, A. Mileham, N. Deeb, A. Di Genova, and A. Maass. 2016. Genomewide single nucleotide polymorphism discovery in Atlantic salmon (Salmo salar): validation in wild and farmed American and European populations. Mol Ecol Resour **16**:1002-1011.
Yin, L., H. Zhang, Z. Tang, J. Xu, D. Yin, Z. Zhang, X. Yuan, M. Zhu, S. Zhao, X. Li, and X. Liu. 2021. rMVP: A memory-efficient, visualization-enhanced, and parallel-accelerated tool for genome-wide association study. Genomics, Proteomics & Bioinformatics **19**:619-628.