RESEARCH PROJECT 2022 - 2023

Identification of selection sweeps following recent domestication : parallel evolution in farmed Atlantic salmon strains from two geographical origins

Pauline Buso[1] | Célian Diblasi[2] | Domniki Manousi[2] | Jun Soung Kwak[2] | Arturo Vera Ponce de Leon[2] | Kristina Stenløkk[2] | Nicola Barson[2]* | Marie Saitou[2]*

[1]University Paul Sabatier and UPVD, Master degree Functional Biology & Ecology (FBE), EUR-TULIP

[2]Centre for Integrative Genetics (CIGENE) and Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, As, Norway

***Corresponding authors:**
Marie Saitou: marie.saitou@nmbu.no ;
Nicola Barson: nicola.barson@nmbu.no

ABSTRACT

Directional selection through domestication has left detectable footprints in the genome of farmed animals, while producing important phenotypic changes. Aquaculture, in particular, has witnessed recent and independent domestication events worldwide, making it a relatively new development in the evolutionary history of domesticated species. This provides a unique opportunity to study the parallel effects of domestication on individuals undergoing similar selection pressures. In this context, Atlantic salmon (*Salmo salar*) is a promising model to study the genetic consequences of domestication, as the reared strains have undergone intense artificial selection over the last half century in multiple geographical regions. However, parallel evolution of different lineages has not been widely proven. Here we show evidence of parallel and population-specific selection signatures in two strains of Atlantic salmon from Western Norway and Canada. Using a genome-wide scan and three outlier methods, several loci were detected as being putatively under selection and were shared between the methods, revealing 45 and 34 genes in the Norwegian and Canadian breeding populations respectively. Our results show that domestication occurs on genes encoding immune response, reproduction and sensory processes and support previous findings that these are traits of interest for breeding programmes. We believe that such traits may be polygenic and will require future analysis to understand them further. Nevertheless, our study provides valuable information on the genetic architecture of traits of practical importance to aquaculture and contributes to our understanding of the early patterns and effects of domestication in Atlantic salmon.

KEYWORDS: S*almo salar*, aquaculture, whole-genome sequencing, selection signatures, genome scan, F_{ST}, Tajima's D, haplotype patterns.

INTRODUCTION

Animal domestication has shaped the genetic landscape of various species throughout history worldwide. This process, whereby a population adapts and mates in a human-controlled environment, takes various forms, ranging from extensive and intensive breeding, and has occurred for economic, social, or symbolic purposes. Spatial and temporal patterns of domestication, as well as archaeological data, show that domestication began with dogs during the Pleistocene and then spread to livestock species such as cattle, pigs, sheep, and goats in relation to nutrition at the end of the most recent ice age (~12,000 B.P.) (Larson et al., 2014). Domestication involves several degrees of intensification, which, in various ways and over varying periods of time, can lead to the emergence of domesticated lineages shaped by humans. The initial stages of domestication include transport, captivity, and acclimatization, followed by selective breeding that leads to significant changes in phenotypes of domestic animals. These changes are reflected in the genome as detectable selection signatures (Vigne, 2011). Genomic evolution of domesticated species has been mainly driven by two elements; one is the small effective population size, due to the fact that the founder population is initiated by a small subset of the wild population and that the subsequent gene flow between domestic and wild populations is limited. Another key element is artificial selection, through which humans have selected beneficial traits, including high yield, appearance, faster reproduction and disease resistance (Mignon-Grasteau et al., 2005). Meanwhile, it is also known that extensive breeding programs and inbreeding brought unwanted by-product phenotypes, for instance, breed-specific diseases such as hepatopathies in pedigree dogs (Watson, 2017), leukocyte adhesion deficiency in cattle (Gholap P.N., Kale D.S. and Sirothia A.R, 2014), inferior brachygnathia in sheep (Greber et al., 2013).

A fundamental question in evolutionary biology is to understand the molecular mechanisms that have shaped domesticated species and the impact of artificial selection on genome architecture. By comparing genomic signatures of domesticated species with their wild counterparts, it is possible to distinguish traits that have been exposed to natural and artificial selection (J.R Stinchcombe & H.E Hoekstra, 2008; Zhao et al., 2015; Brito et al.,

2017) and to identify functional information useful for the design and/or updating of breeding programmes for practice improvement (Qanbari & Simianer, 2014). Technological breakthroughs in genome sequencing and the establishment of high-quality reference genome sequences have provided strong opportunities to investigate the effect of breeding programs on domesticated species genomes (Druet et al., 2013). Genome-wide association analysis (GWAS) and quantitative trait locus (QTL) analyses have been proven to be relatively effective methods to investigate the statistical association between genomic variation and phenotypic variation at the genome-wide scale, and at the single nucleotide or single gene resolution (Lee et al., 2015; Barson et al., 2015; Misztal et al., 2021). Moreover, new genome scan approaches have been developed to efficiently and accurately identify selection footprints using SNP arrays or whole genome sequencing (Jensen, Foll, & Bernatchez, 2016). In livestock species, numerous studies have been conducted to understand genetic factors that are artificially selected for their favorable traits in economic, sustainability and nutritional contexts. For example, studies have identified genes encoding milk or meat production in cattle (Bomba et al., 2015), development, milk, meat and wool production in sheep (Fariello et al., 2014), growth, appetite and metabolite regulation in chicken (Carl-Johan Rubi et al., 2010), muscle development in pigs (Amaral et al., 2011).

In aquaculture species, domestication is more recent compared to any other livestock species and important knowledge gaps still remain (Teletchea & Fontaine, 2012). As for mammals, several breeding programs have been implemented independently, in crustaceans, mollusks and fishes (Houston et al., 2020; Gjedrem, 2012) such as in Rainbow trout (*Oncorhynchus mykiss*) (Gjedrem, 1992), Coho salmon (*Oncorhynchus kisutch*) (Neira et al., 2014), Nile Tilapia (*Oreochromis niloticus*) and Atlantic salmon (*Salmo salar*) (Gjedrem, 2010). This also makes aquaculture species highly tractable for genomic research on domestication, as they offer an opportunity to study the genomic effects of intense artificial selection over a short period of time in multiple geographic regions.

Among others, Atlantic salmon is an especially interesting species in studying evolutionary aspects of domestication. Firstly, Atlantic salmon underwent a whole genome duplication event ~80M years ago inducing variation and redundancy, which can increase tolerance to new, potentially adaptive mutations (Lien et al., 2016). Secondly, Atlantic salmon is a widely distributed anadromous species characterized by a strong tendency to reproduce in the rivers of their birth, which encourages genetic divergence among populations (Klemetsen et al., 2003; Lehnert et al., 2018). Pleistocene glaciations resulted in repeated periods of isolation and range expansion of species (Lehnert et al., 2018). In response to this global climatic change, wild populations of Atlantic salmon have diverged into four main lineages including three clades (Atlantic, Barents/White Sea, Baltic) in Europe and one in North America (Bourret et al., 2013). These lineages show differences in demographic history (Bourret et al., 2013; Rougemont et Bernatchez, 2018) and in karyotype, as the chromosome number is polymorphic in North American lineage (Brenna-Hansen et al., 2012). As a result, several independent breeding programmes have emerged around the world. Atlantic salmon populations have undergone intense artificial selection in parallel over the last half century, ranging from 5 to 12 generations (Glover, 2017). Even in different geographic areas, and using different source populations, artificial selection in breeding programs was directed at similar key traits. GWAS, QTL (Houston et al., 2009; Gutierrez et al., 2012; Gutierrez et al., 2014) and recent genome-wide scans using allelic frequency differentiation (F_{ST}) and haplotype extended patterns (EHH) (Liu et al., 2017; López et al., 2019*b*; Naval-Sanchez et al., 2020) have shown that the major traits targeted by selection in fish are linked to flesh quality, growth, immune resistance and advanced sexual maturity (Gjedrem, 2012; Glover et al., 2017; Houston & Macqueen, 2019). Such replicated directional selection suggests we may observe parallel evolution between domestic strains from different geographical regions. In Atlantic salmon, a previous investigation of Scottish and Canadian breeding populations using 220K SNP array revealed a few overlapping domestication footprints between lineages (López et al., 2019*a*). In the present study, we aim to investigate signatures of parallel domestication by comparing whole genome sequencing data, rather than array data to search for non-neutral signatures comprehensively and in an unbiased way, using two pairs of wild and farmed Atlantic salmon populations from two geographical areas, Western Norway and Canada.

MATERIAL & METHODS

Sampling and genotyping

The present study was conducted using a total of 370 individuals from two pairs of wild and farmed Atlantic salmon populations in two geographic areas. The purpose of this design was to perform two independent comparisons between domestic strains and their wild counterpart populations in order to investigate genomic regions subject to putative artificial selection and their potential functions. Paired-end, short-read whole-genome sequencing data was retrieved from multiple publicly available datasets. We used n=80 farmed samples derived from 3 different source rivers in Canada (NCBI Project No. PRJEB34225), n = 112 farmed Norwegian samples from the MOWI strain that is derived from the River Bolstad, in the Vosso drainage, River Årøy and from marine captures near Osterfjord and Sotra, in western Norway (Glover et al., 2009) (NCBI Project No. PRJEB47441). As the closest available wild counterparts, we used n = 80 wild Canadian individuals from 8 Quebec rivers and n = 98 wild Norwegian individuals from 17 rivers in Western Norway (NCBI Project No. PRJEB38061, Bertolotti et al., 2020) (see Figure 1.).

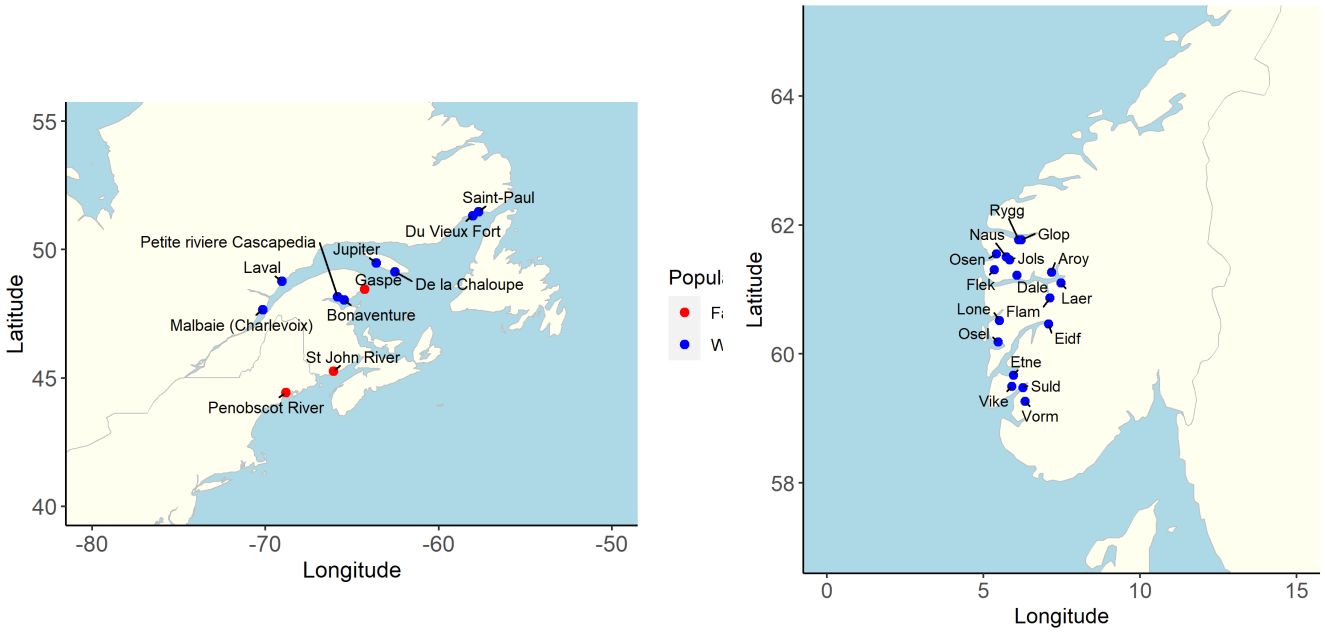


FIGURE 1. Sampling map showing the geographical coordinates of the Canadian and West Norwegian rivers of origin of the farmed populations in red and the wild populations in blue.

Data preparation

Mapping, variant calling and quality control were performed independently for each of the four datasets (Farmed Norwegian, Wild Norwegian, Farmed Canadian, Wild Canadian) using the same pipeline and parameters. Illumina reads were mapped to the Atlantic salmon genome (Ssal_v3.1; GCA_905237065.2) using SAMtools (Li et al., 2009). Duplicate reads were marked with gatk4-spark:4.3.0.0 MarkDuplicates. Genetic variation was identified using gatk4-spark:4.3.0.0 HaplotypeCaller with default parameters and the individual genotypes were merged with gatk4-spark:4.3.0.0 GenotypeGVCFs. To avoid the detection of false positives, variants were filtered by the following parameters with gatk4-spark:4.3.0.0 (Mckenna et al., 2010) : "QD < 2.0" which filters out variants with a

low quality score by depth of coverage; "QUAL < 50.0" to sort variants with a low-quality score; SOR > 4.0 that removes variants where the number of reads supporting the reference allele is significantly skewed towards one strand (forward or reverse) compared to the other; FS > 60.0: to filter out variants whose reads supporting the reference allele versus the alternative allele are significantly different between the two strands; and finally ReadPosRankSum < -8.0: to remove variants with a low score for the difference between the mean position of the reference allele and the alternative allele among the reads supporting the variant (https://gatk.broadinstitute.org/hc/en-us). From this large pruning, we kept only biallelic markers (SNPs) with depth greater than 4 and less than 30, confident call rate greater than 90% (GQ) and missingness lower than 10% using VCFtools (Danecek et al., 2011). These and later filtering were applied on a merged file containing the 4 populations in order to retain good quality, common variants and avoid quality bias.

Population structure analysis

We first used two clustering methods to investigate genetic structure among populations and have a broad representation of the genetic differentiation between Canadian and West Norwegian lineages. Further filtering removed loci deviating from Hardy–Weinberg equilibrium, loci with minor allele frequency (MAF) lower than 5% and linked SNPs (i.e. 50 SNP windows, 10 SNP sliding windows and an R^2 threshold of 0.5) using PLINK v1.9 (Purcell et al., 2007). We ran ADMIXTURE (Alexander, Novembre, Lange, 2015) with a number of ancestral populations set from 1 to 10 (K). The optimal K was selected based on the lowest cross-validation error. We then conducted a Principal Component Analysis (PCA) using PLINK v1.9 on the same filtered dataset. The results of both analyses (i.e eigenvectors and genetic variation from PCA; genetic components from ADMIXTURE) were visualized in R 4.2.2 using the *ggplot2* (Wickham, 2011) and *pophelper* (Francis, 2016) R packages.

Detection of selection signatures

For the detection of putative genomic regions under artificial selection in domestic strains, we applied a genome scan using three statistical methods based on (1) population differentiation (F_{ST}) (Weir & Cockerham, 1984; Beaumont & Balding, 2004), (2) site frequency (Tajima's D) (Tajima, 1989), and (3) cross population test of extended haplotype homozygosity (XP-EHH) (Sabeti et al., 2002). These tests were first carried out to compare farmed and wild populations from the same geographical region independently in order to detect regions specific to each area. Then, we investigated whether some regions are shared by populations of different origins, indicative of parallel evolution. In the context of domestication, we expect two main scenarios in terms of selective sweeps. A soft sweep happens when the selected allele is found in multiple haplotypes due to standing variation or recurrent mutations. In contrast, a hard sweep occurs when a new mutation emerges and rapidly spreads through the population, resulting in complete linkage between the selected allele and the surrounding haplotype. A hard sweep will occur rapidly as an allele that segregates for a long time before selection starts will be found in multiple haplotypes because of recombination and mutation.
First, F_{ST} identifies selection using differences in allele frequencies between populations, here between domesticated and wild populations from each geographic region independently. These statistics were calculated according to the pairwise Weir and Cockerham's (1984) estimator using VCFtools (Danecek et al., 2011) in 5,000 bp non-overlapping sliding windows using wild individuals as the reference population. We extracted the top 1% values as outliers. As F_{ST} allows the detection of large changes in allele frequency, the test is particularly sensitive to hard sweeps, which involve strong selection. However, when using the sliding window approach, F_{ST} values are primarily influenced by the surrounding haplotype. Here the window size is typically small enough that the background haplotype is not a critical factor and indicates that F_{ST} values are unlikely to distinguish between hard and soft sweeps. The strength of selection will matter and it may be problematic to differentiate it from genetic drift following a founder bottleneck and small effective population size of captivity using this test.

Secondly, Tajima's D (T_D) assumes that neutral genetic variation is maintained by a balance between mutation, genetic drift, and gene flow. The test is computed as the ratio between the number of segregating sites and the average pairwise nucleotide differences. Deviations from this neutral evolution can indicate putative selection or demographic events. Markers under selection can be expressed as positive and negative values, while zero values indicate a putative absence of selection. A significant positive T_D value may indicate an excess of intermediate-frequency alleles, which could be caused by balancing selection, population subdivision, or recent population expansion. A significant negative T_D value may indicate an excess of low-frequency alleles, which could be caused by positive selection, population bottlenecks, or purifying selection. In our study, T_D was conducted on the two farmed populations independently by non-overlapping sliding windows of 20,000 bp. As the domestication involved bottlenecks and focus on directional selection, we are more interested in negative values. The lowest values at the 1% level were taken as the significance level.

Finally, we investigated patterns of extended haplotypes between pairs of populations using the *rehh* R package (Gautier, Klassmann, Vitalis, 2017). Haplotype analyses are based on extended haplotype homozygosity (EHH) that represents the model of a selective sweep in which an adaptive de novo mutation appears on a haplotype and rapidly evolves towards fixation, thus reducing diversity around the locus (Sabeti et al., 2002). A hard sweep induces a signal of high haplotype homozygosity to be observed extending from an adaptive locus as strong rapid selection causes strong linkage that is not broken by recurrent mutation or recombination. In other words, under conditions of positive selection, a beneficial new allele can rapidly increase in frequency and create long stretches of homozygosity, or identical copies of the haplotype, around the selected site. This process results in longer and linked haplotypes detectable by EHH methods. In our case, we applied a cross-population (XP-EHH) test that compares the distribution of haplotype homozygosity between farmed and wild populations. When there is divergent selection between the two environments (wild and aquaculture), the allele selected in the focal population (farm) will be on a longer haplotype. This signal will not be present when selection is concordant between the two environments. Negative scores of XP-EHH suggest that selection occurred in the reference population, whereas positive scores show selection in the domestic population. Unlike Tajima and F_{ST}, these tests cannot directly handle SNP genotypes because they require haplotypes. Added to the first filtering (see *Data preparation* section), variants with Minor Allele Frequency < 0.05 were removed. The remaining variants were phased to infer the haplotypes of each population using BEAGLE v4.1 (Browning S.R & Browning B.L, 2007) with a window size of 10,000 bp and 600 bp overlap. One-sided p-values were obtained as $-\log10(1-\Phi(\text{XP-EHH}))$, where Φ(XP-EHH) represents the Gaussian distribution function for the statistic. False Discovery Rate (FDR) test was performed to estimate the number of false positives and determine the significant threshold. We used $-\log10(\text{p-value}) = 4$ (alpha = 0,0001) as a threshold to define significant values of XP-EHH. As XP-EHH searches for the extended haplotype, which is typical of a hard sweep, this method will be more sensitive in detecting hard sweeps than soft selective sweeps. In soft sweeps, the selected allele may have arisen and segregated as neutral variation in the wild, resulting in it being found in multiple haplotypes and the weaker signals extended haplotypes may be difficult to detect using XP-EHH, depending on the allele frequency at the time of domestication (Innan & Kim, 2004). Nevertheless, the combination of the three methods is particularly promising to increase the probability of detecting true directional selection signals.

SNP annotation

Significant windows and/or SNPs from the three comparative analyses were intersected using BEDtools (Quinlan, 2014) to examine markers detected by multiple methods and thus increase the repeatability of the results. For each geographical region, three intersections were tested: (i) F_{ST}/T_D, (ii) F_{ST}/XP-EHH, (iii) T_D/XP-EHH. Candidate variants detected by two or more methods were then intersected with annotated genes in the Atlantic salmon reference genome (Ssal_v3.1; GCA_905237065.2) using BEDtools window to identify variants under selection that are contained in the same windows as the genes or that are directly mapped to one or more genes. The overlapping

genes were input into ShinyGO 0.77 (Ge, Jung, Yao, 2020; http://bioinformatics.sdstate.edu/go/) to extract biological functions and enrichment.
We further investigated whether any of the candidate variants could have an impact on the overlapping genes. For this, variants from BEDtools intersect were entered into the Variant Effect Predictor tool on Ensembl v.109 to determine whether they have a regulatory function in the target gene. The coding sequences (mRNA) were used to study the potential changes in sequence and protein caused by the variant using Salmobase (Samy et al., 2017; http://www.salmobase.org/). The protein domain and structure were observed using HMMER 3.3.2 (https://www.ebi.ac.uk/Tools/hmmer/).

Gene expression

As a final step, we measured the expression of all detected genes with or without a coding variant within their overlapping window. We used RNA sequencing data from the NCBI Project No. SRP011583 including data on the following tissues: brain, eye, gill, gut, head kidney, heart, kidney, liver, muscle, nose, ovary, caecum, skin, spleen and testis.

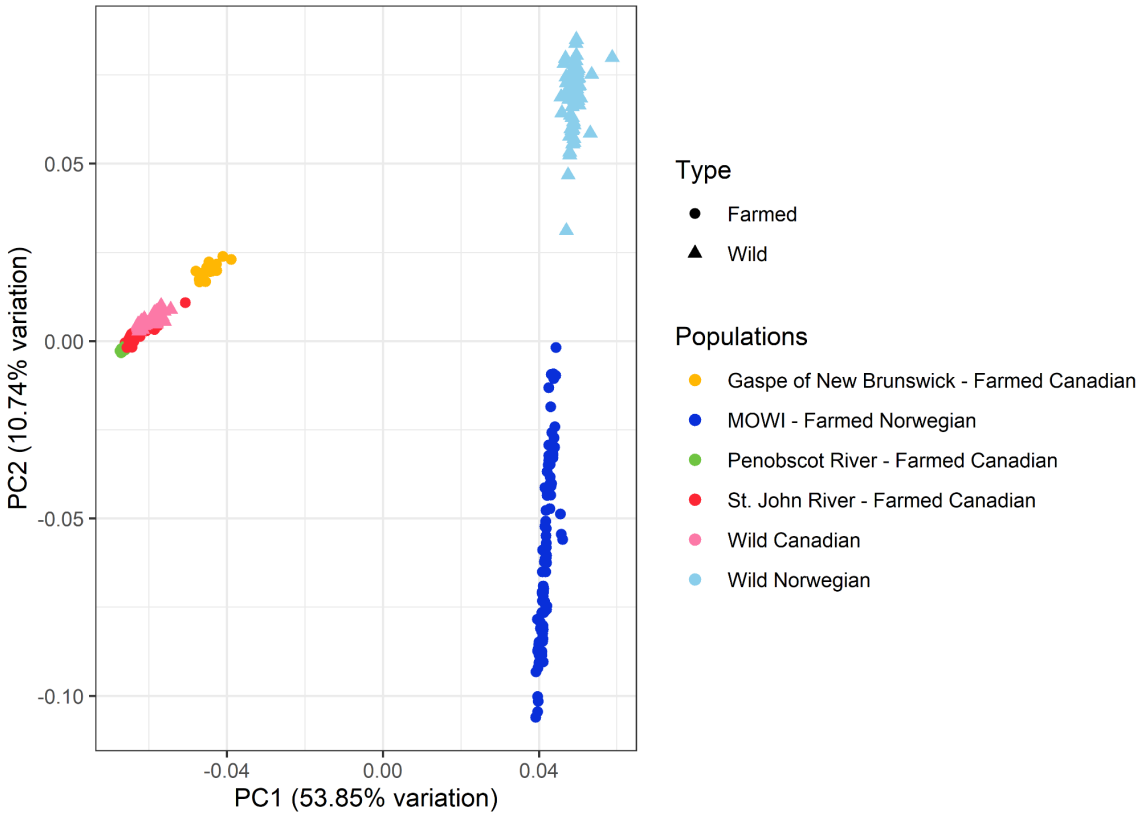
RESULTS

Genotyping and SNPs filtering

After initial mapping to the Atlantic salmon genome (v3.1) and variant calling there were 30,785,285 SNPs. Depth, low quality and missingness filtering left 1,139,312 variants for each of the four populations. Added to this first filtering, 1,053,324 variants were remaining after MAF cut-off of 0.05 for phasing and XP-EHH analysis. The genetic structure analyses were conducted on a total of 148,433 neutral markers after LD pruning, Hardy-Weinberg filtering and MAF cut-off from the first dataset containing 1,139,312 variants.

Genetic Structure analysis

(a)



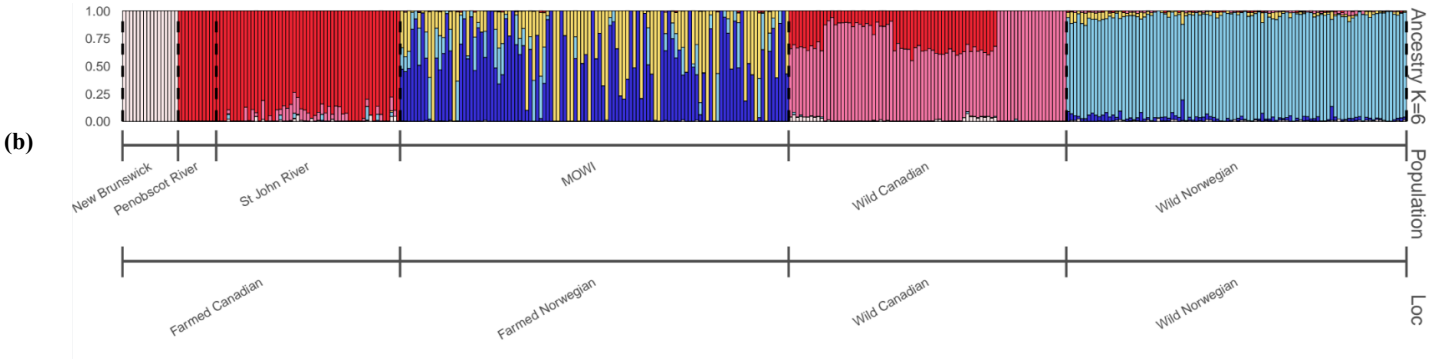


FIGURE 2. Genetic structure among populations. (a) Principal Component Analysis of Atlantic salmon populations using 148, 433 neutral SNPs highlighting trans-Atlantic differences. Samples are colored by type (triangle for wild ; spot for farmed) and population. The distribution of the explained genetic variance along the 10 PCs is presented in the **Supplementary material Fig 1**. (b) ADMIXTURE results based on the same populations and neutral markers as the PCA. See methods for details regarding the analysis. The plot shows the degree of ancestry and segregation of populations using 6 genetic components (K=6) according to the lowest cross-validation error (see **Supplementary material Fig.2**).

The PCA analysis (**Fig 2. (a)**) showed four distinct clusters: (1) the farmed Norwegian strain, (2) their wild counterpart, (3) farmed Canadian from St John and Penobscot rivers and wild Canadian, (4) the farmed Canadian strain from the Gaspe of New Brunswick. The first two principal components (PCs) jointly accounted for 64.59% of the total variance, with PC1 (53.85% of variance) separating Canadian and West Norwegian lineages, and PC2 (10.74%) distinguishing between wild and domestic populations. In agreement with the PCA, the ancestry analysis (**Fig 2. (b)**) also showed clear separation between the two geographical regions, with each lineage having its own main genetic component shared to varying degrees between wild and farmed individuals from the same geographical region. The breeding population on the Norwegian side was rather heterogeneous genetically, unlike the wild population. Wild individuals in Canadian populations seem genetically closer to farmed animals, especially those from St. John, but samples from the Penobscot River only shared their main genetic component with the wild. The farmed population from the Gaspe of New Brunswick showed an independent genetic pattern, and this strain-specific component was also visible in the wild at a low level. A small portion of the wild European genome persisted in both wild and farmed Canadian individuals, indicating introgression on a broad scale.

Genomic markers putatively under selection distributed across the genome

The F_{ST} calculation identified 3177 outlier windows putatively under divergent selection in Norwegian populations according to the threshold of the top 1% (**Fig 3. (a)**) . The SNPs included in these windows were distributed among almost all the chromosomes (**Supplementary material Fig 3.**). In Canadian strains, 3088 windows were detected (**Fig. 3 (b)**) along the majority of chromosomes (**Supplementary material Fig 4.**), similar to what we observed for the Norwegian populations.

Defining the bottom 1% as an outlier threshold in T_D, the analysis revealed 1053 windows under selection in the farmed Norwegian population (**Fig 3. (a)**) and 1032 in the Canadian population (**Fig 3. (b)**). In Norwegians, T_D follows the trends of the F_{ST} analysis. The significant windows are spread over all chromosomes whereas in the Canadian breed it is very homogeneous with some chromosomes having only one or two significant windows (**Supplementary material Fig 3 & 4.**).

Haplotype differentiation analysis between domestic and wild populations identified 1914 significant positions (**Fig 3. (a)**) in the farmed Norwegian population and 1982 SNPs in Canadian farmed strain (**Fig 3. (b)**).

Significant genomic regions shared among methods and overlapping genes

BEDtools intersect allowed us to see if some regions under selection were detected by multiple methods and, therefore, increase the resolution and the precision of the final candidates markers. Among the three tests in the Norwegian comparisons, 54 significant shared windows were detected by F_{ST}/T_D intersect, 144 windows by F_{ST}/XP-EHH intersect and no shared windows between T_D and XP-EHH (**Fig 3. (a)**). In Canadian populations, the

first intersection between F_{ST} and T_D identified 9 shared windows followed by 363 shared windows in the second intersection between F_{ST} and XP-EHH. The last intersect T_D/XP-EHH follows the same trend as the Norwegian populations and did not identify any shared position under selection between the two methods (**Fig 3. (b)**). These candidate regions were mapped to genes using Atlantic salmon genome annotation of Ensembl v.109. It is usual that BEDtools do not distinguish between genes that are detected several times. Since F_{ST} and T_D were run using non-overlapping windows, several windows, especially side-by-side windows, may map to the same gene. Similarly in the F_{ST} and XP-EHH intersect: XP-EHH use direct positions of SNPs instead of windows. This results in several XP-EHH SNPs that can intersect the same F_{ST} window, and causes multiple detection of the same gene overlapping with the F_{ST} window. A last possibility is that the same window is mapped with different genes. These three situations cause redundancy in the genes detected. Removing these redundancies, we obtained a total of 45 genes (i.e 26 detected by F_{ST}/T_D intersect; 19 detected by F_{ST}/XP-EHH intersect) in Norwegian farmed strain and 34 genes (i.e 8 detected by F_{ST}/T_D intersect; 26 detected by F_{ST}/XP-EHH intersect) in Canadian. The position, name and description of the corresponding genes are available in the **Supplementary material** (**Table S1 and S2**). In Norwegian samples, genes were distributed on 17 chromosomes, mainly on chromosomes 1, 14 and 5 with four genes identified on each, chromosomes 9 and 21 with five genes and chromosome 27 with six genes. In the Canadian population, genes were distributed among 11 chromosomes, four of which had a larger number of genes identified. Nine genes were found in chromosome 4, 13 in chromosome 7, ten in chromosome 14 and four in chromosome 22. Only five genes in each of the two populations were uncharacterised loci.

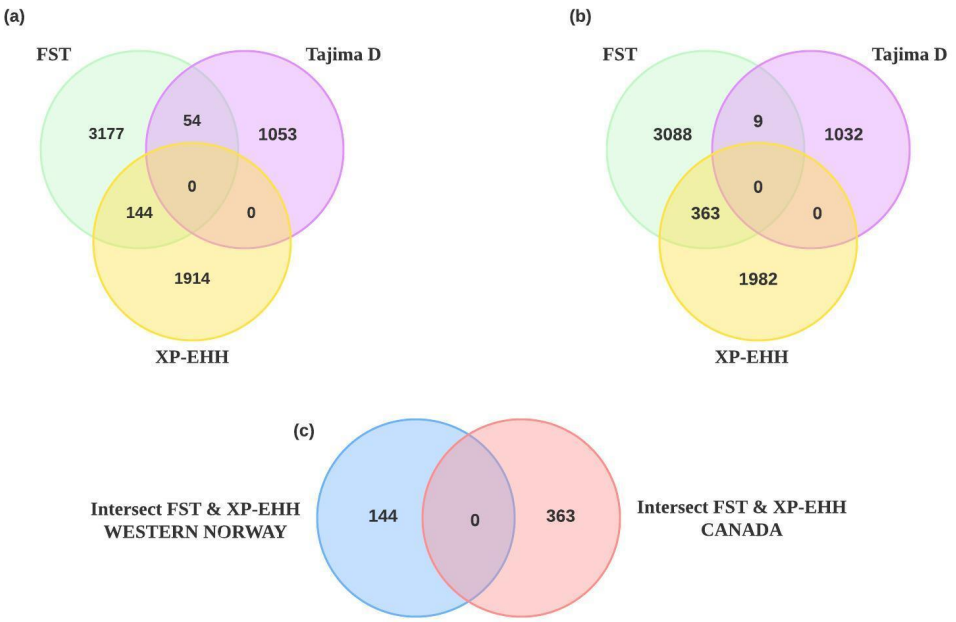


FIGURE 3. Venn diagrams showing shared SNPs with evidence of selection among three independent tests (F_{ST}, Tajima's D, XP-EHH) in (a) West Norwegian populations (b) Canadian populations and the overlap of loci identified from F_{ST} and XP-EHH intersect between the two geographical regions.

A gene ontology (GO) enrichment analysis on the total of outlier sequences found in both populations revealed that five and two categories were significantly over-represented in farmed Norwegian and Canadian strains respectively, representing metabolic (phosphate by-products) and cellular processes (e.g cell activity (adhesion, differentiation, proliferation), morphogenesis, reorganization of the actin cytoskeleton for cell movement and gene expression) (**Table 1**).

Apart from the enrichment analysis, a lot a genes are also involved in metabolic, cellular and signaling pathways but a small number of genes also code for processes associated with immune response (e.g *il21*, the locus

ENSSSAG00000078726 orthologous to *lgals17* in Norwegian; *gemin5*, *tapasin*, *tap2a* in Canadian populations), early development (e.g *eed* and *thsd7a* in Canadians*),* and neurogenesis (e.g *ngfb*, *grn* in Canadians).

Enrichment FDR	Fold Enrichment	Pathway	Genes	Population
4.8E-02	78.72	Phosphonate and phosphinate metabolism	pcyt1aa	West Norway
4.8E-02	15.14	Inositol phosphate metabolism	PI4KB, plcd4a	West Norway
4.8E-02	12.05	Phosphatidylinositol signaling system	PI4KB, plcd4a	West Norway
4.8E-02	7.96	Focal adhesion	col4a2 , ENSSSAG00000058972, CRK	West Norway
4.8E-02	15.14	ECM-receptor interaction	col4a2, ENSSSAG00000058972	West Norway
3.1E-02	79.52	Threonine-type peptidase activity	PSB6, ggt1b	Canada
3.1E-02	14.37	Intracellular protein-containing complex	PSB6, 2a5e, cnot8, TRPC4AP	Canada

TABLE 1. Enrichment analysis on genes detected in both farmed populations

Prediction of SNP function and/or impact on the overlapping genes

The next step was to look at whether overlapping variants in the gene windows had an impact on gene function. For this we input the candidate variants of the F_{ST}/XP-EHH intersect into the Variant Effect Predictor (VEP) on Ensembl (version 109). We chose to test only this intersect because it is the one with the largest number of shared windows. We found 14 SNPs impacting five genes in the Canadian farmed strain and two genes in Norwegian farmed strain. Three types of high impact SNPs were identified: (1) missense variant which means that a single base pair substitution alters the genetic code to produce a different amino acid than the reference amino acid at that position. Some missense variants change the function of the protein; (2) stop-lost variant when the mutation happens within translational termination codons, which could result in the continued translation of the messenger RNA and elongate the protein; and (3) splice acceptor variant that changes the 2 base region at the 3' end of an intron.
Among the overlapping genes in the Canadian breeding strain, one SNP is supposed to have an impact on the uncharacterised locus *ENSSSAG00000118501*. Additionally, we found ten SNPs impacting *plekhd1*, which plays a role in intracellular signaling or the cytoskeleton structure; *tapasin*, *thsd7a* and *eed* described above. In the Norwegian farmed population, three SNPs impact the locus *ENSSSAG00000041320* corresponding to the orthologous gene *rarb* in zebrafish, predicted to play a role in cellular processes during embryonic morphogenesis; and the *ENSSSAG00000078726* locus that corresponds to the orthologous gene *lgals17* described above as well. RNA sequencing data revealed that *plekhd1* and *rarb* are highly expressed in the eyes and *thsd7a* in the brain. *eed*, *lgals17* and *tapasin* are globally expressed in all tissues but *eed* seems to be particularly expressed in ovaries and testis. The position of the SNPs impacting genes, and gene expression are shown in Figures 4 and 5.

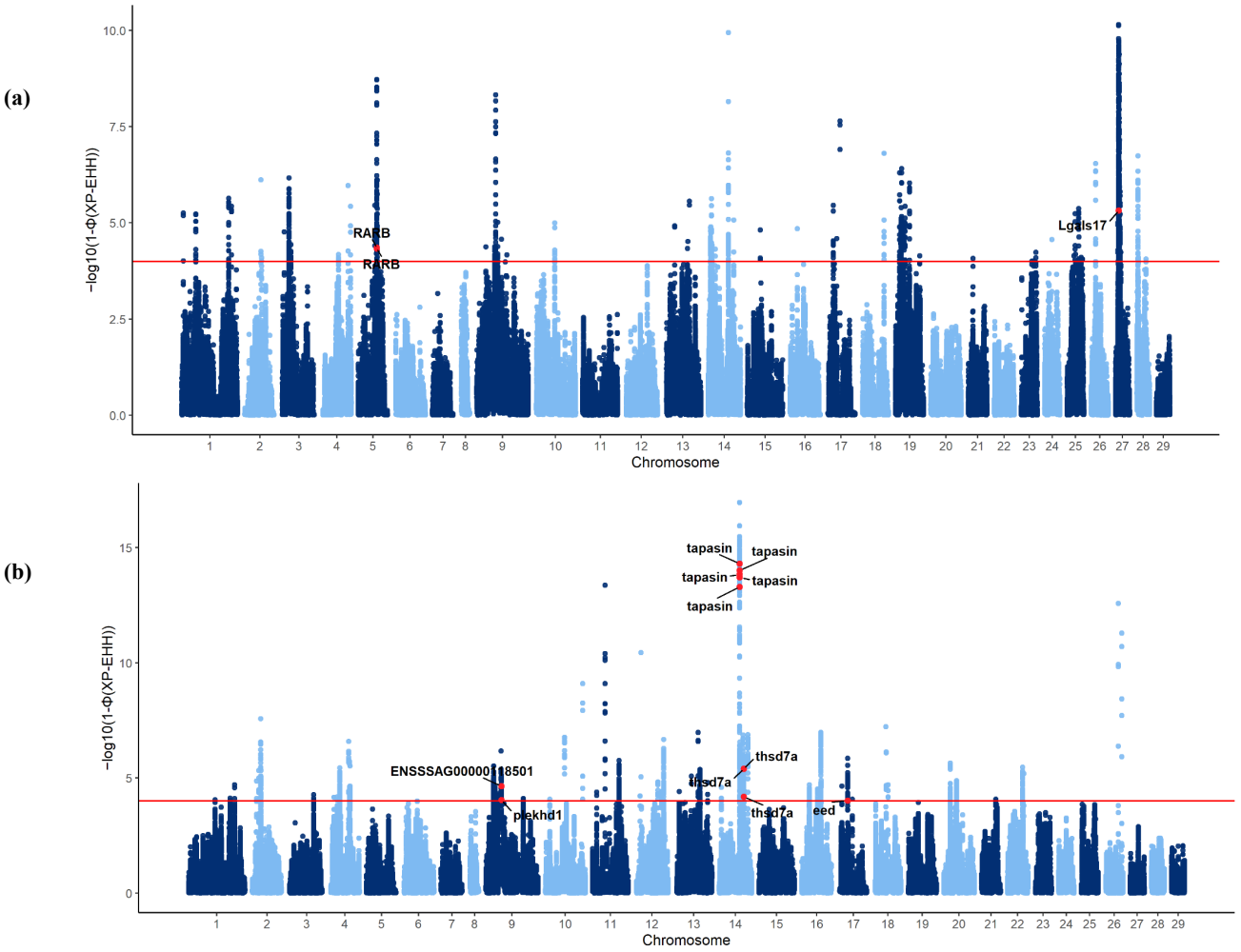


Figure 4. Distribution of the 14 SNPs under selection impacting genes (missense, stop-lost, slice-acceptor variants). The Manhattan plot represents the distribution of all XP-EHH results in the **(a)** Norwegian and **(b)** Canadian farmed populations. SNPs from the intersect between F_{ST} and XP-EHH are highlighted in red with the label of associated gene.

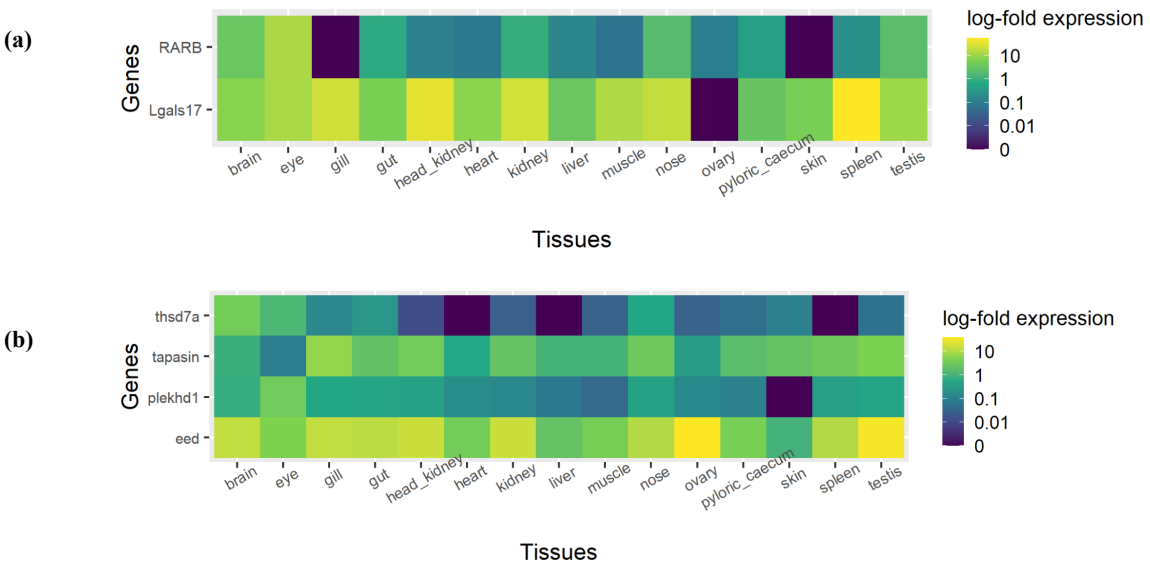


FIGURE 5. Expression level of genes with missense, stop-lost and splice-acceptor variants coding in (a) West Norwegian domestic strains, and in (b) Canadian domestic strains.

The gene sequences were then observed on Salmobase in order to recognise mutations in codon and posterior amino acid changes. Details of the gene overlapping selected variants, allelic frequencies, associated genes and changes in amino acids and codons are summarized in **Table 2**. Most of the mutations do not take effect in the protein domain and do not affect the protein structure except for variants on *tapasin* gene in the Canadian strain.

To verify the selection signal detected at the 14 loci, additional haplotype analyses were conducted. It was assumed that the allele under selection, associated with the extended haplotype, has a higher frequency in the domestic population. **Figures 7 to 10** in the **Supplementary material** show examples of the EHH distribution and the furcation patterns of the reference and alternative allele for two markers. This represents which allele has the extended haplotype and allows us to verify the selection signal identified in the XP-EHH and F_{ST} overlap by comparing it with the allelic frequency for the given marker. Overall, our results are consistent between the selection signal detected by the XP-EHH method and the allelic frequencies.

Parallel signal of selection among the Norwegian and Canadian lineages

Investigating parallel evolution, we used the F_{ST}/XP-EHH intersect with the hypothesis to find positions under selection, shared between the Canadian and Norwegian breeding strains. The BEDtools results (**Fig 3. (c)**) revealed no shared genomic positions between the domestic strains from different geographical origins. However, the overlap of genes with variants specific to each geographical region suggests similar target traits, especially regarding the immune-related genes.

DISCUSSION

The present study investigated how domestication shapes the genomic landscape of animals to understand how human-driven selection can induce genetic changes. In the last 50 years, Atlantic salmon populations have undergone intense artificial selection resulting in the emergence of strains that exhibit specialized phenotypic traits (Milla et al., 2021). This process has caused significant phenotypic and genotypic differences between wild and domestic populations, leading to the opportunity to compare farmed individuals with their wild counterparts to search for selection signatures from domestication (Glover et al., 2009; 2017). Here, we performed genome-wide scans for footprints of artificial selection, using whole genome sequencing data from two geographically isolated Atlantic salmon strains, in Canada and Norway, by integrating three statistical methods. Whole genome sequencing can produce much larger datasets than conventionally used SNP arrays, and allows an increase in the resolution and precision in detecting sweeps. Genome sequencing data also opens the possibility of detecting putative functional variants that directly underpin artificial selection. We used the latest chromosome anchored long read reference genome for European Atlantic salmon (GCA_905237065.2) (Stenløkk et al., 2022), offering the possibility to find new variants, in addition to possibly confirming some found in the literature, and to search for evidence of parallel evolution.

Genetic differentiation between European and Canadian lineages

Clustering analysis allowed us to get an overview of the genetic structure and differentiation between the two lineages, then between farmed and wild from the same geographical region. Both PCA and ADMIXTURE showed a clear genetic separation between the two geographical regions. The differences in the demographic history, the anadromous way of life can explain this separation and reinforce the genetic differentiation between populations from these two geographical regions. Only a few small traces of Wild European persist among farmed Canadian individuals and suggest introgression events (**Fig. 2 (b)**). Previous studies have focused on potential introgression that may occur between farmed and wild fishes sharing the same geographic origin or between two different

TABLE 2. Gene overlapping selected variants, allelic frequencies, associated genes and changes in amino acids and codons.

				Allelic frequencies									
				Domestic strain		Wild strain							
Chr	SNP position	Reference allele	Alternative allele	Reference	Alternative	Reference	Alternative	Gene	Variant type	Population	Amino acids changes	Codons changes	Consequences
ssa09	48294282	G	C	0,019	0,981	0,262	0,738	NA	missense	Canadian	P/A	Ccc/Gcc	-
ssa09	48349333	G	A	0,206	0,794	0,575	0,425	plekhd1	missense	Canadian	A/T	Gcc/Acc	Mutation occurs in PH domain
ssa14	64331704	C	A	0,944	0,056	0,694	0,306	tapasin	stop lost	Canadian	*/Y	taG/taT	12 additional amino acids. Out-of-Immunoglobulin V-set domain
ssa14	64332437	T	A	0,862	0,138	0,894	0,106	tapasin	missense	Canadian	H/L	cAc/cTc	Out-of-Immunoglobulin V-set domain, destabilized protein structure
ssa14	64332602	C	G	0,99375	0,00625	0,781	0,219	tapasin	missense	Canadian	W/S	tGg/tCg	Out-of-Immunoglobulin V-set domain, destabilized protein structure
ssa14	64333061	C	G	0,031	0,969	0,162	0,838	tapasin	missense	Canadian	E/Q	Gaa/Caa	Mutation in Immunoglobulin V-set domain, destabilized protein structure
ssa14	64333641	G	A	0,969	0,031	0,844	0,156	tapasin	missense	Canadian	T/M	aCg/aTg	-
ssa14	76489218	A	C	0,975	0,025	0,687	0,313	thsd7a	missense	Canadian	D/E	gaT/gaG	Several protein domains, mutation out of domains
ssa14	76547818	T	C	0,981	0,019	0,969	0,031	thsd7a	splice acceptor	Canadian	-	-	-
ssa14	76547838	T	C	0,081	0,919	0,406	0,594	thsd7a	missense	Canadian	K/E	Aaa/Gaa	Mutation out of domains
ssa17	26765566	C	T	0,937	0,063	0,769	0,231	eed	missense	Canadian	V/I	Gtt/Att	Mutation close to the start codon,out of WD domain.
ssa05	57560371	A	G	0,263	0,737	0,816	0,184	rarb	missense	Norwegian	M/T	aTg/aCg	-
ssa05	57560633	A	T	0,737	0,263	0,219	0,781	rarb	missense	Norwegian	V/E	gTg/gAg	-
ssa27	10802925	C	T	0,062	0,938	0,189	0,811	lgals17	missense	Norwegian	E/K	Gag/Aag	-

lineages. For instance, Bradbury et al., 2022 demonstrate that some Canadian farmed salmon and escaped fish showed significantly higher levels of European ancestry compared to Canadian wild individuals. Although introgression between diverging lineages has been used to increase production in captive stocks, it can also induce the maladaptation of wild individuals to a given natural environment if the farmed fish escape and reproduce with the wild strains (Baskett et al., 2013; Fleming, 1995), affecting the genetic diversity and long-term life history traits of wild populations.

Because domestication is recent in salmonids, it was expected to find a higher level of shared genetic components between wild and farmed individuals from the same geographical region compared to other species with longer domestication history. The population-specific genetic signature of farmed fish can be explained by several elements. Firstly, although the farmed population is derived from a wild subpopulation, the significant bottleneck associated with artificial selection and selective breeding has definitely shaped the farmed populations, resulting in unique genetic components in the farmed. This phenomenon is more easily observed in farmed Norwegians, where the main wild genetic component fades away in favor of other genetic components that are much less shared with the wild. Other arguments for this variation revolve around the mixed origin of the breeding population and/or the maintenance of distinct breeding lineages in different age classes. It is important to mention that we did not sample the exact progenitor populations. Thus, for the Norwegian wild samples we do not have samples from the Vosso watershed. Some of the unshared components in the MOWI samples may reflect this unsampled progenitor population.

The wild Canadian samples are also not directly the source populations for the aquaculture strains. In this lineage, there are also very few wild genetic components, following the same direction of evolution of genetic structure as the farmed Norwegians. Individuals from Gaspe of New Brunswick have their own component, shared with some wild samples at a low level, but it is difficult here to conclude on a scenario explaining this difference.

It would be interesting to calculate the inbreeding coefficient between populations as it is a factor that can influence diversity and genetic structure due to selective breeding.

Parallel and non-parallel selection signatures

Overall, our results revealed population-specific footprints of domestication that are detected by multiple statistical methods, between F_{ST} and XP-EHH and between F_{ST} and Tajima's D, while we did not observe shared genomic regions under selection between the two lineages. This is consistent with the latest findings on parallel evolution in Atlantic salmon from López et al., 2019*a* and Naval-Sanchèz et al., 2020. that highlighted only a few genomic regions of overlap between populations with North American and European origins. This can be explained by genetic differentiation that took place during demographic history as well as by differences in nutrition, management or environmental conditions. Moreover, several population genomic studies on Atlantic salmon have suggested that geographically isolated populations may exhibit different genetic pathways leading to similar phenotypes (Bourret et al., 2013; Elmer et al., 2014). This indicates that parallel evolution of complex traits can occur via various evolutionary routes, particularly when genetic drift and inbreeding strongly influence allele frequencies.

Although we did not observe parallel evolution at the same genomic regions, we observed overlapped functions of genes under selection in the two lineages. Indeed, a few genomic regions harboring genes related to the immune response were identified in both domestic strains. Some missense and stop-lost variants overlapped with an orthologous locus of *lgals17* gene in Norwegian farmed population and with *tapasin* in Canadian farmed population. Among several functions, *lgals17* is predicted to be involved in T cell activation in Zebrafish. Another study on Nile tilapia (*Oreochromis niloticus*), a species that is widely farmed and commercially important, found that *lgals17* was upregulated in response to bacterial infection (Barría et al., 2021). This gene was expressed in several different tissues, including the liver, spleen, and kidney, which are important organs for immune function in fish (Barría et al., 2021). In our study we found the locus highly expressed in spleen, kidney and gut tissues, suggesting that *lgals17* may play an important role in the immune response in Atlantic salmon as well. However,

more research is needed to fully understand the function of the present gene, the orthologs loci and its potential implications for fish health and productivity. Regarding *tapasin,* studies have shown that this gene is highly conserved across various domesticated animals (Subramaniam et al., 2010). *tapasin* is generally located in the major histocompatibility complex (MHC) class II region and helps stabilize the peptide loading complex (PLC) of MHC class I molecules for immune recognition with T cells (Ortmann et al., 1997; Lehner et al., 1998; Garbi et al., 2003). In salmonids, several studies already investigated the similar role of *tapasin* in Rainbow trout and Atlantic salmon (Landis et al., 2006; Jørgensen, Grimholt, Gjøen, 2007; Sever et al., 2014). The detection of missense and stop-lost variants impacting these genes may provide additional information on their regulation in Atlantic salmon. No such effects were predicted for the locus orthologous to *Lgals17* but a few variants within *tapasin* were predicted to destabilize the protein structure. One of them in particular is located in the immunoglobulin V-set domain. Further studies would be interesting to deeper investigate the consequences of these mutations on the gene.

Apart from them, other overlapping genes with no predicted functional SNPs have been identified and could play a role in immune response in Atlantic salmon. First, *il21* (interleukin 21) in the Norwegian farmed population and *gemin5* and *tap2a* in the Canadian farmed population. The function(s) of interleukin 21 may vary between species, but it is generally involved in regulating immune responses by the activation and proliferation of various immune cells (Spolski & Leonard, 2008). *Gemin5* is predicted to act upstream of or within T cell differentiation in Zebrafish (Iwanami et al., 2016). Furthermore, *tap2a* is considered as a gene associated with *tapasin* as it encodes a protein that transports peptides from the cytoplasm of cells into the endoplasmic reticulum (ER), where they can be loaded onto major histocompatibility complex (MHC) class I molecules for presentation to T cells (Hansen et al., 2007). These findings support the view that immune resistance is a highly selected trait in domestication in Atlantic salmon (Kjærner-Semb et al., 2016).

Immune response is not the only main trait targeted by domestication. Some interesting genes with high-impact SNPs were detected such as *eed* and *thsd7a*, both involved in early development. *thsd7a* is known to code for the thrombospondin I and the formation of blood vessels during embryonic development in Zebrafish (Wang et al., 2011) and *eed* corresponds to Embryonic Ectoderm Development and is a highly conserved gene in vertebrates (Faust et al., 1995). We also know from the literature that *eed* could be involved in kisspeptin (*KISS*) neuropeptide regulation which globally regulates reproductive functions in teleost fish. *KISS* has been shown to play an essential role in the activation of the hypothalamic-pituitary-gonadal (HPG) axis, which is necessary for the regulation of puberty, gonadal development and gametogenesis by the release of various hormones (Ohga, Selvaraj, & Matsuyama, 2018). *eed* gene can regulate the expression of *KISS* by modifying the histone code at the promoter region of the *KISS* gene. This modification can either enhance or repress the expression of *KISS*, depending on the context and other regulatory factors and has been shown in Rainbow trout (Allen, 2015). Although the role of this gene in Atlantic salmon reproduction has not been proven, our results show that *eed* is highly expressed in salmon reproductive organs. This suggests a role similar to that in Rainbow trout which would be interesting to investigate in a future study.

Apart from the highlighted parallel traits, we found in Canadian farmed strain the genes *ngfb* and *grn,* involved respectively in nerve growth or neuro differentiation and sensory system development. Previous studies on Atlantic salmon population from North America and whole genome sequencing also found 3 genes involved in neurogenesis (*arvcf, sema6a, erbby*) (Naval-Sanchez et al. 2020; Bertolotti et al., 2015). This suggests a relationship with behavioral and/or environmental sensitivity traits that could be interesting for breeding programs in the context of captivity, environmental conditions and perhaps stress. Related to growth, we know from Lopez et al., 2019 that *ube2f* gene is coding for ubiquitin-conjugating enzyme E2-F. This gene is believed to play a role in the growth and development of skeletal muscle, as it undergoes protein breakdown and replacement during different stages. In teleost fish, this process involves E2-ubiquitin-conjugating enzymes and E3-ubiquitin ligases. Here, we found *ube2e*

gene in Norwegian farmed strain, coding for ubiquitin-conjugating enzyme E2 as well, but it is too poorly documented in fish to suggest that it plays a similar role.

Inconsistencies among methods and limits of genome scans

The intersection between Tajima's D and XP-EHH failed to detect shared position under putative selection. This lack of congruence between the methods is probably due to the statistics underlying each approach: Tajima's D is based on site frequencies, F_{ST} on allelic frequency differentiation and XP-EHH is a haplotype comparison method. As the three different tests can detect different patterns, even if SNPs are only detected by one test, they can still be real selection targets. The variants from F_{ST}/Tajima's D intersect were not tested in the Variant Effect Predictor tool in this study but it would be interesting to look at this in a future analysis and complete the present results.
The use of three methods effectively serves to increase detection power but also increases the probability of finding false positives. For this purpose, an FDR was applied, in particular the XP-EHH method, which allowed us to define a threshold of significance authorizing 5% of false positives on all significant SNPs. Additional analyses of the distribution of extended haplotypes and bifurcation analyses on each of the 14 high-impact SNPs allowed us to verify the power and reliability of detection using the XP-EHH method, and then in the overlap with F_{ST}. Most of the results showed logical signals between the allele with the extended haplotype and the allelic frequencies, under the assumption that the selected allele was the one with the highest frequency in farmed individuals. However, the detection signal is not always obvious and may be related to genetic drift and bottleneck events. Indeed, it is frequent as the domestic population undergoes a reduction in size, such as during the early stages of domestication when a small number of individuals are selected for breeding. This is the case for a variant on chromosome 14 around 64 Mbp which shows that both alleles have extended haplotypes. Our methods likely detect hard sweeps that cannot explain all selection events occurring in domestic strains. But using XP-EHH, we find that it is also possible to observe soft sweeps as some bifurcation plots showed that the focal allele is present in multiple haplotypes.

Moreover, such well-known traits of interest for breeding programs, such as immune resistance, growth, and advanced sexual maturity may be controlled by polygenic genes (i.e controlled by many genes with high variance of effect size) (Gagnaire & Gaggiotti, 2016; Barson et al., 2015; Sinclair-Waters et al., 2020). Polygenic selection may imply a subtle change of allele frequencies at several loci that would be hard to detect using genome scan methods that mainly focus on moderate to high shifts in allele frequency at few independent loci. To overcome this limitation, alternative models, such as a quantitative genetic framework, can be employed to investigate polygenic selection. Studies on Atlantic salmon have already reported a large number of different quantitative trait loci (QTL) associated with growth late sexual maturity (Houston et al., 2009; Gutierrez et al., 2012; Gutierrez et al., 2014). Future studies combining population genomic and quantitative genetic approaches would provide a more comprehensive method for uncovering the molecular footprints of domestication in Atlantic salmon.

Complexity of salmon genome and potential bias

The salmon genome is complex due to whole genome duplication and chromosome fusions in the North American lineage in particular (Lien et al., 2016). Duplicate regions are certainly a source of variation, but their sequencing results in regions of low coverage and unreliable reads, as they may correspond to several locations in the genome. In the present study, we can cite some positions on chromosomes 1, 2, 4, 5 11, 16 and 26 in Norwegian strains that are located in regions with medium or high sequences similarity and a few positions in chromosomes 2, 11 and 17 in Canadian populations. Note that the positions of chromosome 5 and chromosome 17 are those corresponding to the *rarb* and *eed* genes respectively. Therefore, results in these regions should be interpreted with caution.
Chromosomal rearrangements between European and North American Atlantic salmon have been well identified. For instance the chromosome 23 of the North American lineage was created from the translocation and then fusion of the p arm of European chromosomes 1 and 23. North American chromosome 8 is made from the fusion of

acrocentric European chromosomes 8 and 29 and chromosome 26 is a fusion of European chromosomes 26 and 28. Although these major structural differences are important, the order of markers in these chromosomal regions is relatively conserved but the chromosomal rearrangements may bias the detection of genomic regions under selection in the Canadian strain when using European Atlantic salmon as a reference genome.

CONCLUSION

Selective breeding, artificial selection and bottleneck related to domestication have left detectable selection signatures and resulted in significant phenotypic changes in farmed Atlantic salmon populations. Using genome-wide scan, our study identified both parallel and non-parallel genomic regions putatively under selection in farmed strain compared to their wild counterparts. Some of them belong to genes that may be associated with domestication-related traits such as immune response, sexual maturity and behavior. However, the candidate regions were not shared between different methods or lineages, suggesting that domestication occurs differently in genomes of Canadian and Norwegian populations due to their distinct evolutionary history and genetic variation. Alternative methods that test the genomic signal of polygenic adaptation may be required to reveal the genetic basis of traits of interest. The complexity of the salmon genome makes some variants difficult to interpret, hence the importance of supporting our results with additional analyses. Nevertheless, our study provides valuable insights into the genetic architecture of traits of practical importance to aquaculture and helps to deepen our understanding of the patterns resulting from recent domestication in Atlantic salmon.

ACKNOWLEDGEMENTS

I would like to thank my supervisors Mrs. Marie Saitou and Mrs. Nicola Barson for their supervision, support and guidance in the realization of this project. I also thank all the members of MSLab: Célian Diblasi, Domniki Manousi, Jun Sung Kwak, Erik Sandertund Roed for their support and their participation in this project as well as Arturo Vera-Ponce de Leon for his precious help in my training in bioinformatics. I thank Joris Bertrand from the Laboratoire Génome et Développement des Plantes (LGDP) in Perpignan for his advice on genetic structure analysis. Finally, I would like to thank on a larger scale the whole BIOVIT team and the CIGENE laboratory for having welcomed me so well, it was a pleasure to collaborate together. This work bearing the reference EUR TULIP-GSR N°ANR-18-EUR-0019 has benefited from a grant managed by the French National Research Agency (ANR) in the framework of the program « Investing for the Future ». The study was supported by The Research Council of Norway (grant nos. 275310 and 221734). We acknowledge the use of the Orion computing cluster at the Norwegian University of Life Sciences (NMBU).

REFERENCES

Larson, Greger, Dolores R. Piperno, Robin G. Allaby, Michael D. Purugganan, Leif Andersson, Manuel Arroyo-Kalin, Loukas Barton. 2014. « Current Perspectives and the Future of Domestication Studies ». *Proceedings of the National Academy of Sciences* 111, nº 17 : 6139‑46. https://doi.org/10.1073/pnas.1323964111.

Vigne, Jean-Denis. 2011. « The Origins of Animal Domestication and Husbandry: A Major Change in the History of Humanity and the Biosphere ». *Comptes Rendus Biologies* 334, nº 3 : 171‑81.https://doi.org/10.1016/j.crvi.2010.12.009.

Mignon-Grasteau, Sandrine, Alain Boissy, Jacques Bouix, Jean-Michel Faure, Andrew D. Fisher, Geoffrey N. Hinch, Per Jensen, et al. 2005. « Genetics of Adaptation and Domestication in Livestock ». *Livestock Production Science* 93, nº 1 : 3‑14. https://doi.org/10.1016/j.livprodsci.2004.11.001.

Watson, Penny. 2017. « Canine Breed-Specific Hepatopathies ». *Veterinary Clinics of North America: Small Animal Practice* 47, nº 3 : 665‑82. https://doi.org/10.1016/j.cvsm.2016.11.013.

Gholap, Prakash N., D. S. Kale and A. R. Sirothia. 2014. « Genetic Diseases in Cattle: A Review ». *Research Journal of Animal, Veterinary and Fishery Sciences* 2, n°2 : 24-33.

Greber, Deborah, Marcus Doherr, Cord Drögemüller, et Adrian Steiner. 2013. « Occurrence of Congenital Disorders in Swiss Sheep ». *Acta Veterinaria Scandinavica* 55, nº 1 : 27. https://doi.org/10.1186/1751-0147-55-27.

Stinchcombe, J R, et H E Hoekstra. 2008. « Combining Population Genomics and Quantitative Genetics: Finding the Genes Underlying Ecologically Important Traits ». *Heredity* 100, nº 2 : 158‑70. https://doi.org/10.1038/sj.hdy.6800937.

Zhao, Fuping, Sinead McParland, Francis Kearney, Lixin Du, et Donagh P Berry. 2015. « Detection of Selection Signatures in Dairy and Beef Cattle Using High-Density Genomic Information ». *Genetics Selection Evolution* 47, nº 1 : 49. https://doi.org/10.1186/s12711-015-0127-3.

Brito, Luiz F., James W. Kijas, Ricardo V. Ventura, Mehdi Sargolzaei, Laercio R. Porto-Neto, Angela Cánovas, Zeny Feng, Mohsen Jafarikia, et Flávio S. Schenkel. 2017. « Genetic Diversity and Signatures of Selection in Various Goat Breeds Revealed by Genome-Wide SNP Markers ». *BMC Genomics* 18, nº 1 : 229. https://doi.org/10.1186/s12864-017-3610-0.
Qanbari, Saber, et Henner Simianer. 2014. « Mapping Signatures of Positive Selection in the Genome of Livestock ». *Livestock Science* 166 : 133‑43. https://doi.org/10.1016/j.livsci.2014.05.003.

Druet, T., L. Pérez-Pardal, C. Charlier, et M. Gautier. 2013. « Identification of Large Selective Sweeps Associated with Major Genes in Cattle ». *Animal Genetics* 44, nº 6 : 758‑62. https://doi.org/10.1111/age.12073.

Lee, Young-Sup, Hyeonsoo Jeong, Mengistie Taye, Hyeon Jeong Kim, Sojeong Ka, Youn-Chul Ryu, et Seoae Cho. 2015. « Genome-Wide Association Study (GWAS) and Its Application for Improving the Genomic Estimated Breeding Values (GEBV) of the Berkshire Pork Quality Traits ». *Asian-Australasian Journal of Animal Sciences* 28, nº 11 : 1551‑57. https://doi.org/10.5713/ajas.15.0287.

Barson, Nicola J., Tutku Aykanat, Kjetil Hindar, Matthew Baranski, Geir H. Bolstad, Peder Fiske, Céleste Jacq, et al. 2015. « Sex-Dependent Dominance at a Single Locus Maintains Variation in Age at Maturity in Salmon ». *Nature* 528, nº 7582 : 405‑8. https://doi.org/10.1038/nature16062.

Misztal, Ignacy, Ignacio Aguilar, Daniela Lourenco, Li Ma, Juan Pedro Steibel, et Miguel Toro. 2021. « Emerging Issues in Genomic Selection ». *Journal of Animal Science* 99, nº 6. https://doi.org/10.1093/jas/skab092.

Jensen, Jeffrey D., Matthieu Foll, et Louis Bernatchez. 2016. « The Past, Present and Future of Genomic Scans for Selection ». *Molecular Ecology* 25, nº 1 : 1‑4. https://doi.org/10.1111/mec.13493.

Bomba, Lorenzo, Ezequiel L Nicolazzi, Marco Milanesi, Riccardo Negrini, Giordano Mancini, Filippo Biscarini, Alessandra Stella, Alessio Valentini, et Paolo Ajmone-Marsan. 2015. « Relative Extended Haplotype Homozygosity Signals across Breeds Reveal Dairy and Beef Specific Signatures of Selection ». *Genetics Selection Evolution* 47, nº 1 : 25. https://doi.org/10.1186/s12711-015-0113-9.

Fariello, Maria-Ines, Bertrand Servin, Gwenola Tosser-Klopp, Rachel Rupp, Carole Moreno, International Sheep Genomics Consortium, Magali San Cristobal, et Simon Boitard. 2014. « Selection Signatures in Worldwide Sheep Populations ». Édité par David Caramelli. *PLoS ONE* 9, nº 8. https://doi.org/10.1371/journal.pone.0103813.

Amaral, Andreia J., Luca Ferretti, Hendrik-Jan Megens, Richard P. M. A. Crooijmans, Haisheng Nie, Sebastian E. Ramos-Onsins, Miguel Perez-Enciso, Lawrence B. Schook, et Martien A. M. Groenen. 2011. « Genome-Wide Footprints of Pig Domestication and Selection Revealed through Massive Parallel Sequencing of Pooled DNA ». Édité par Hans Ellegren. *PLoS ONE* 6, nº 4. https://doi.org/10.1371/journal.pone.0014782.

Teletchea, Fabrice, et Pascal Fontaine. 2014. « Levels of Domestication in Fish: Implications for the Sustainable Future of Aquaculture ». *Fish and Fisheries* 15, nº 2 : 181‑95. https://doi.org/10.1111/faf.12006.

Houston, Ross D., Tim P. Bean, Daniel J. Macqueen, Manu Kumar Gundappa, Ye Hwa Jin, Tom L. Jenkins, Sarah Louise C. Selly, et al. 2020. « Harnessing Genomics to Fast-Track Genetic Improvement in Aquaculture ». *Nature Reviews Genetics* 21, nº 7 : 389‑409. https://doi.org/10.1038/s41576-020-0227-y.

Gjedrem, Trygve. 2012. « Genetic Improvement for the Development of Efficient Global Aquaculture: A Personal Opinion Review ». *Aquaculture* 344‑349 : 12‑22. https://doi.org/10.1016/j.aquaculture.2012.03.003.
Gjedrem, Trygve. 1992. « Breeding Plans for Rainbow Trout ». *Aquaculture* 100, nº 1‑3 : 73‑83. https://doi.org/10.1016/0044-8486(92)90340-Q.

Neira Roberto, Yanez M.J, Lhorente J.P. 2014. “Evolution of Coho Salmon (*Oncorhynchus kisutch*) Breeding Programs”.
Gjedrem, Trygve. 2010. « The First Family-Based Breeding Program in Aquaculture ». *Reviews in Aquaculture* 2, nº 1 : 2‑15. https://doi.org/10.1111/j.1753-5131.2010.01011.x.

Lien, Sigbjørn, Ben F. Koop, Simen R. Sandve, Jason R. Miller, Matthew P. Kent, Torfinn Nome, Torgeir R. Hvidsten, et al. 2016. « The Atlantic Salmon Genome Provides Insights into Rediploidization ». *Nature* 533, nº 7602 : 200‑205. https://doi.org/10.1038/nature17164.

Klemetsen, A., P.-A. Amundsen, J. B. Dempson, B. Jonsson, N. Jonsson, M. F. O’Connell, et E. Mortensen. 2003. « Atlantic Salmon *Salmo Salar* L., Brown Trout *Salmo Trutta* L. and Arctic Charr *Salvelinus Alpinus* (L.): A Review of Aspects of Their Life Histories: Salmonid Life

Histories ». *Ecology of Freshwater Fish* 12, nº 1 : 1‑59. https://doi.org/10.1034/j.1600-0633.2003.00010.x.

Lehnert, Sarah J., Paul Bentzen, Tony Kess, Sigbjørn Lien, John B. Horne, Marie Clément, et Ian R. Bradbury. 2018. « Chromosome Polymorphisms Track Trans-Atlantic Divergence, Admixture and Adaptive Evolution in Salmon ». Preprint. Genomics. https://doi.org/10.1101/351338.

Bourret, Vincent, Matthew P. Kent, Craig R. Primmer, Anti Vasemägi, Sten Karlsson, Kjetil Hindar, Philip McGinnity, Eric Verspoor, Louis Bernatchez, et Sigbjørn Lien. 2013. « SNP-Array Reveals Genome-Wide Patterns of Geographical and Potential Adaptive Divergence across the Natural Range of Atlantic Salmon (*Salmo Salar*) ». *Molecular Ecology* 22, nº 3 : 532‑51. https://doi.org/10.1111/mec.12003.

Rougemont, Quentin, et Louis Bernatchez.2018. « The Demographic History of Atlantic Salmon (*Salmo Salar*) across Its Distribution Range Reconstructed from Approximate Bayesian Computations*: ATLANTIC SALMON HISTORY AND LINKED SELECTION ». *Evolution* 72, nº 6 : 1261‑77. https://doi.org/10.1111/evo.13486.

Brenna-Hansen, Silje, Jieying Li, Matthew P Kent, Elizabeth G Boulding, Sonja Dominik, William S Davidson, et Sigbjørn Lien. 2012. « Chromosomal Differences between European and North American Atlantic Salmon Discovered by Linkage Mapping and Supported by Fluorescence in Situ Hybridization Analysis ». *BMC Genomics* 13, nº 1 : 432. https://doi.org/10.1186/1471-2164-13-432.

Glover, Kevin A, Monica F Solberg, Phil McGinnity, Kjetil Hindar, Eric Verspoor, Mark W Coulson, Michael M Hansen, Hitoshi Araki, Øystein Skaala, et Terje Svåsand. 2017. « Half a Century of Genetic Interaction between Farmed and Wild Atlantic Salmon: Status of Knowledge and Unanswered Questions ». *Fish and Fisheries* 18, nº 5 : 890‑927. https://doi.org/10.1111/faf.12214.

Liu, Lei, Keng Pee Ang, J. A. K. Elliott, Matthew Peter Kent, Sigbjørn Lien, Danielle MacDonald, et Elizabeth Grace Boulding. 2017. « A Genome Scan for Selection Signatures Comparing Farmed Atlantic Salmon with Two Wild Populations: Testing Colocalization among Outlier Markers, Candidate Genes, and Quantitative Trait Loci for Production Traits ». *Evolutionary Applications* 10, nº 3 : 276‑96. https://doi.org/10.1111/eva.12450.

López, Maria E., Laura Benestan, Jean‑Sebastien Moore, Charles Perrier, John Gilbey, Alex Di Genova, Alejandro Maass, et al. 2019. a. « Comparing Genomic Signatures of Domestication in Two Atlantic Salmon (*Salmo Salar* L.) Populations with Different Geographical Origins ». *Evolutionary Applications* 12, nº 1 : 137‑56. https://doi.org/10.1111/eva.12689.

López, María Eugenia, Tyler Linderoth, Ashie Norris, Jean Paul Lhorente, Roberto Neira, et José Manuel Yáñez.2019. (b) « Multiple Selection Signatures in Farmed Atlantic Salmon Adapted to Different Environments Across Hemispheres ». *Frontiers in Genetics* 10 (1 octobre 2019): 901. https://doi.org/10.3389/fgene.2019.00901.

Naval-Sanchez, Marina, Sean McWilliam, Bradley Evans, José M. Yáñez, Ross D. Houston, et James W. Kijas. 2020. « Changed Patterns of Genomic Variation Following Recent Domestication: Selection Sweeps in Farmed Atlantic Salmon ». *Frontiers in Genetics* 11 : 264. https://doi.org/10.3389/fgene.2020.00264.

Houston, R. D., et D. J. Macqueen. 2019. « Atlantic Salmon (*Salmo Salar* L.) Genetics in the 21st Century: Taking Leaps Forward in Aquaculture and Biological Understanding ». *Animal Genetics* 50, nº 1 : 3‑14. https://doi.org/10.1111/age.12748.

Glover, Kevin A., Håkon Otterå, Rolf E. Olsen, Erik Slinde, Geir L. Taranger, et Øystein Skaala. 2009. « A Comparison of Farmed, Wild and Hybrid Atlantic Salmon (Salmo Salar L.) Reared under Farming Conditions ». *Aquaculture* 286, nº 3‑4 : 203‑10. https://doi.org/10.1016/j.aquaculture.2008.09.023.

Bertolotti, Alicia C., Ryan M. Layer, Manu Kumar Gundappa, Michael D. Gallagher, Ege Pehlivanoglu, Torfinn Nome, Diego Robledo, et al. 2020. « The Structural Variation Landscape in 492 Atlantic Salmon Genomes ». *Nature Communications* 11, nº 1 : 5176. https://doi.org/10.1038/s41467-020-18972-x.
Li, Heng, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, et 1000 Genome Project Data Processing Subgroup. 2009; « The Sequence Alignment/Map Format and SAMtools ». *Bioinformatics* 25, nº 16 : 2078‑79. https://doi.org/10.1093/bioinformatics/btp35

McKenna, Aaron, Matthew Hanna, Eric Banks, Andrey Sivachenko, Kristian Cibulskis, Andrew Kernytsky, Kiran Garimella, et al. 2012. « The Genome Analysis Toolkit: A MapReduce Framework for Analyzing next-Generation DNA Sequencing Data ». *Genome Research* 20, nº 9 : 1297‑1303. https://doi.org/10.1101/gr.107524.110.

Danecek, Petr, Adam Auton, Goncalo Abecasis, Cornelis A. Albers, Eric Banks, Mark A. DePristo, Robert E. Handsaker, et al. 2011. « The Variant Call Format and VCFtools ». *Bioinformatics* 27, nº 15 : 2156‑58. https://doi.org/10.1093/bioinformatics/btr330.

Purcell, Shaun, Benjamin Neale, Kathe Todd-Brown, Lori Thomas, Manuel A.R. Ferreira, David Bender, Julian Maller, et al. 2007. « PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses ». *The American Journal of Human Genetics* 81, nº 3 : 559‑75. https://doi.org/10.1086/519795.

Alexander, David H., John Novembre, et Kenneth Lange. 2009. « Fast Model-Based Estimation of Ancestry in Unrelated Individuals ». *Genome Research* 19, nº 9 : 1655‑64. https://doi.org/10.1101/gr.094052.109.

Wickham, Hadley. 2011 « Ggplot2 ». *Wiley Interdisciplinary Reviews: Computational Statistics* 3, nº 2 : 180‑85. https://doi.org/10.1002/wics.147.

Francis, R. M. 2017. « POPHELPER : An R Package and Web App to Analyse and Visualize Population Structure ». *Molecular Ecology Resources* 17, nº 1 : 27‑32. https://doi.org/10.1111/1755-0998.12509.

Weir, B S, et C Clark Cockerham. 1984. « Estimating F-Statistics for the Analysis of Population Structure ».

Tajima, F. 1989. « Statistical Method for Testing the Neutral Mutation Hypothesis by DNA Polymorphism. » *Genetics* 123, nº 3 : 585‑95. https://doi.org/10.1093/genetics/123.3.585.

Sabeti, Pardis C., David E. Reich, John M. Higgins, Haninah Z. P. Levine, Daniel J. Richter, Stephen F. Schaffner, Stacey B. Gabriel, et al. 2002. « Detecting Recent Positive Selection in the Human Genome from Haplotype Structure ». *Nature* 419, nº 6909 : 832‑37. https://doi.org/10.1038/nature01140.

Gautier, Mathieu, Alexander Klassmann, et Renaud Vitalis. 2017. « REHH 2.0: A Reimplementation of the R Package REHH to Detect Positive Selection from Haplotype Structure ». *Molecular Ecology Resources* 17, nº 1 : 78‑90. https://doi.org/10.1111/1755-0998.12634.

Browning, Sharon R., et Brian L. Browning. 2007. « Rapid and Accurate Haplotype Phasing and Missing-Data Inference for Whole-Genome Association Studies By Use of Localized Haplotype Clustering ». *The American Journal of Human Genetics* 81, nº 5 : 1084‑97. https://doi.org/10.1086/521987.

Innan, Hideki, et Yuseob Kim. 2004. « Pattern of Polymorphism after Strong Artificial Selection in a Domestication Event ». *Proceedings of the National Academy of Sciences* 101, nº 29. https://doi.org/10.1073/pnas.0401720101.

Quinlan, Aaron R. 2014. « BEDTools: The Swiss‑Army Tool for Genome Feature Analysis ». *Current Protocols in Bioinformatics* 47, nº 1. https://doi.org/10.1002/0471250953.bi1112s47.

Ge, Steven Xijin, Dongmin Jung, et Runan Yao.2020. « ShinyGO: A Graphical Gene-Set Enrichment Tool for Animals and Plants ». Édité par Alfonso Valencia. *Bioinformatics* 36, nº 8 : 2628‑29. https://doi.org/10.1093/bioinformatics/btz931.
Samy, Jeevan Karloss Antony, Teshome Dagne Mulugeta, Torfinn Nome, Simen Rød Sandve, Fabian Grammes, Matthew Peter Kent, Sigbjørn Lien, et Dag Inge Våge. 2017. « SalmoBase: An Integrated Molecular Data Resource for Salmonid Species ». *BMC Genomics* 18, nº 1 : 482. https://doi.org/10.1186/s12864-017-3877-1.

Milla, Sylvain, Alain Pasquet, Leila El Mohajer, et Pascal Fontaine. 2021. « How Domestication Alters Fish Phenotypes ». *Reviews in Aquaculture* 13, nº 1 : 388‑405. https://doi.org/10.1111/raq.12480.

Stenløkk, Kristina, Marie Saitou, Live Rud-Johansen, Torfinn Nome, Michel Moser, Mariann Árnyasi, Matthew Kent, Nicola Jane Barson, et Sigbjørn Lien. 2022. « The Emergence of Supergenes from Inversions in Atlantic Salmon ». *Philosophical Transactions of the Royal Society B: Biological Sciences* 377, nº 1856. https://doi.org/10.1098/rstb.2021.0195.

Bradbury, Ian R., Sarah Jean Lehnert, Tony Kess, Mallory Van Wyngaarden, Steven Duffy, Amber M. Messmer, Brendan Wringe, et al. 2022. « Genomic Evidence of Recent European Introgression into North American Farmed and Wild Atlantic Salmon ». *Evolutionary Applications* 15, nº 9 : 1436‑48. https://doi.org/10.1111/eva.13454.

Baskett, Marissa L., Scott C. Burgess, et Robin S. Waples. 2013. « Assessing Strategies to Minimize Unintended Fitness Consequences of Aquaculture on Wild Populations ». *Evolutionary Applications* 6, nº 7 : 1090‑1108. https://doi.org/10.1111/eva.12089.

Fleming, Ian A. « Pattern and Variability in the Breeding System of Atlantic Salmon (Salmo Salar), with Comparisons to Other Salmonids », 1998.

Elmer, Kathryn R., Shaohua Fan, Henrik Kusche, Maria Luise Spreitzer, Andreas F. Kautt, Paolo Franchini, et Axel Meyer. 2014. « Parallel Evolution of Nicaraguan Crater Lake Cichlid Fishes via Non-Parallel Routes ». *Nature Communications* 5, nº 1 : 5168. https://doi.org/10.1038/ncomms6168.

Barría, Agustin, Trọng Quốc Trinh, Mahirah Mahmuddin, Carolina Peñaloza, Athina Papadopoulou, Ophelie Gervais, V. Mohan Chadag, John A. H. Benzie, et Ross D. Houston. 2021. « A Major Quantitative Trait Locus Affecting Resistance to Tilapia Lake Virus in Farmed Nile Tilapia (Oreochromis Niloticus) ». *Heredity* 127, nº 3 : 334‑43. https://doi.org/10.1038/s41437-021-00447-4.

Subramaniam, N. Siva, E.F. Morgan, C.Y. Lee, J.D. Wetherall, et D.M. Groth. 2010. « Polymorphism of Sheep MHC Class IIb Gene TAPASIN ». *Veterinary Immunology and Immunopathology* 137, nº 1‑2 : 176‑80. https://doi.org/10.1016/j.vetimm.2010.04.018.

Ortmann, Bodo, James Copeman, Paul J. Lehner, Bhanu Sadasivan, Jethro A. Herberg, Andeas G. Grandea, Stanley R. Riddell, et al. 1997. « A Critical Role for Tapasin in the Assembly and Function of Multimeric MHC Class I-TAP Complexes ». *Science* 277, nº 5330: 1306‑9. https://doi.org/10.1126/science.277.5330.1306.

Lehner, Paul J, Michael J Surman, et Peter Cresswell. 1998. « Soluble Tapasin Restores MHC Class I Expression and Function in the

Tapasin-Negative Cell Line .220 ». *Immunity* 8, nº 2 : 221‑31. https://doi.org/10.1016/S1074-7613(00)80474-4.

Garbi, Natalio, Neeraj Tiwari, Frank Momburg, et Günter J. Hämmerling. 2003. « A Major Role for Tapasin as a Stabilizer of the TAP Peptide Transporter and Consequences for MHC Class I Expression ». *European Journal of Immunology* 33, nº 1 : 264‑73. https://doi.org/10.1002/immu.200390029.

Landis, Eric D., Yniv Palti, Jenefer Dekoning, Robert Drew, Ruth B. Phillips, et John D. Hansen. 2006. « Identification and Regulatory Analysis of Rainbow Trout Tapasin and Tapasin-Related Genes ». *Immunogenetics* 58, nº 1 : 56‑69. https://doi.org/10.1007/s00251-005-0070-5.

Jørgensen, Sven Martin, Unni Grimholt, et Tor Gjøen. 2007. « Cloning and Expression Analysis of an Atlantic Salmon (Salmo Salar L.) Tapasin Gene ». *Developmental & Comparative Immunology* 31, nº 7 : 708‑19. https://doi.org/10.1016/j.dci.2006.10.004.

Sever, Lital, Nguyen T.K. Vo, Niels C. Bols, et Brian Dixon. 2014. « Expression of Tapasin in Rainbow Trout Tissues and Cell Lines and up Regulation in a Monocyte/Macrophage Cell Line (RTS11) by a Viral Mimic and Viral Infection ». *Developmental & Comparative Immunology* 44, nº 1 : 86‑93. https://doi.org/10.1016/j.dci.2013.11.019.

Spolski, Rosanne, et Warren J. Leonard. 2008. « Interleukin-21: Basic Biology and Implications for Cancer and Autoimmunity ». *Annual Review of Immunology* 26, nº 1 : 57‑79. https://doi.org/10.1146/annurev.immunol.26.021607.090316.

Hansen, Michael M., Øystein Skaala, Lasse Fast Jensen, Dorte Bekkevold, et Karen-Lise D. Mensberg. 2007. « Gene Flow, Effective Population Size and Selection at Major Histocompatibility Complex Genes: Brown Trout in the Hardanger Fjord, Norway: Genetic Structure of Brown Trout ». *Molecular Ecology* 16, nº 7.https://doi.org/10.1111/j.1365-294X.2007.03255.x.

Iwanami, Norimasa, Katarzyna Sikora, Andreas S. Richter, Maren Mönnich, Lucia Guerri, Cristian Soza-Ried, Divine-Fondzenyuy Lawir, et al. « Forward Genetic Screens in Zebrafish Identify Pre-MRNA-Processing Pathways Regulating Early T Cell Development ». *Cell Reports* 17, nº 9 (novembre 2016): 2259‑70. https://doi.org/10.1016/j.celrep.2016.11.003.

Kjærner-Semb, Erik, Fernando Ayllon, Tomasz Furmanek, Vidar Wennevik, Geir Dahle, Eero Niemelä, Mikhail Ozerov, et al. 2016. « Atlantic Salmon Populations Reveal Adaptive Divergence of Immune Related Genes - a Duplicated Genome under Selection ». *BMC Genomics* 17, nº 1 : 610. https://doi.org/10.1186/s12864-016-2867-z.

Wang, Chieh-Huei, I-Hui Chen, Meng-Wei Kuo, Pei-Tsu Su, Zih-Yin Lai, Chian-Huei Wang, Wei-Chang Huang, et al.2011. « Zebrafish *Thsd7a* Is a Neural Protein Required for Angiogenic Patterning during Development ». *Developmental Dynamics* 240, nº 6 : 1412‑21. https://doi.org/10.1002/dvdy.22641.

Ohga, Hirofumi, Sethu Selvaraj, et Michiya Matsuyama. 2018. « The Roles of Kisspeptin System in the Reproductive Physiology of Fish With Special Reference to Chub Mackerel Studies as Main Axis ». *Frontiers in Endocrinology* 9 : 147. https://doi.org/10.3389/fendo.2018.00147.

Allen, Melissa S. 2015. « The Genetic Architecture of Spawning Date and the Associations Among Life History Traits and Growth in Rainbow Trout (Oncorhynchus Mykiss) ».

Sinclair-Waters, Marion, Jørgen Ødegård, Sven Arild Korsvoll, Thomas Moen, Sigbjørn Lien, Craig R. Primmer, et Nicola J. Barson. 2020. « Beyond Large-Effect Loci: Large-Scale GWAS Reveals a Mixed Large-Effect and Polygenic Architecture for Age at Maturity of Atlantic Salmon ». *Genetics Selection Evolution* 52, nº 1 : 9. https://doi.org/10.1186/s12711-020-0529-8.

Gagnaire, Pierre-Alexandre, et Oscar E. Gaggiotti. 2016. « Detecting Polygenic Selection in Marine Populations by Combining Population Genomics and Quantitative Genetics Approaches ». *Current Zoology* 62, nº 6 : 603‑16. https://doi.org/10.1093/cz/zow088.

Houston, R. D., S. C. Bishop, A. Hamilton, D. R. Guy, A. E. Tinch, J. B. Taggart, A. Derayat, B. J. McAndrew, et C. S. Haley 2009. « Detection of QTL Affecting Harvest Traits in a Commercial Atlantic Salmon Population ». *Animal Genetics* 40, nº 5 : 753‑55. https://doi.org/10.1111/j.1365-2052.2009.01883.x.

Gutierrez, Alejandro P., Krzysztof P. Lubieniecki, Evelyn A. Davidson, Sigbjørn Lien, Matthew P. Kent, Steve Fukui, Ruth E. Withler, Bruce Swift, et William S. Davidson. 2012.« Genetic Mapping of Quantitative Trait Loci (QTL) for Body-Weight in Atlantic Salmon (Salmo Salar) Using a 6.5 K SNP Array ». *Aquaculture* 358‑359 : 61‑70. https://doi.org/10.1016/j.aquaculture.2012.06.017.

Gutierrez, Alejandro P., Krzysztof P. Lubieniecki, Steve Fukui, Ruth E. Withler, Bruce Swift, et William S. Davidson.2014. « Detection of Quantitative Trait Loci (QTL) Related to Grilsing and Late Sexual Maturation in Atlantic Salmon (Salmo Salar) ». *Marine Biotechnology* 16, nº 1 : 103‑10. https://doi.org/10.1007/s10126-013-9530-3.

SUPPLEMENTARY MATERIAL

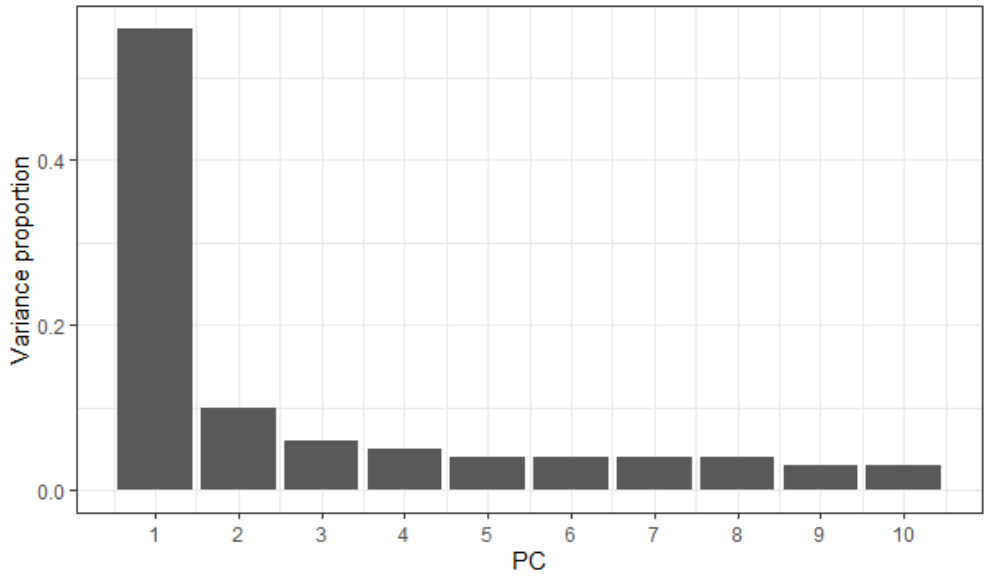


Figure 1. Genetic variance proportion distributed alongside the 10 PCs of the Principal Component Analysis (PCA).

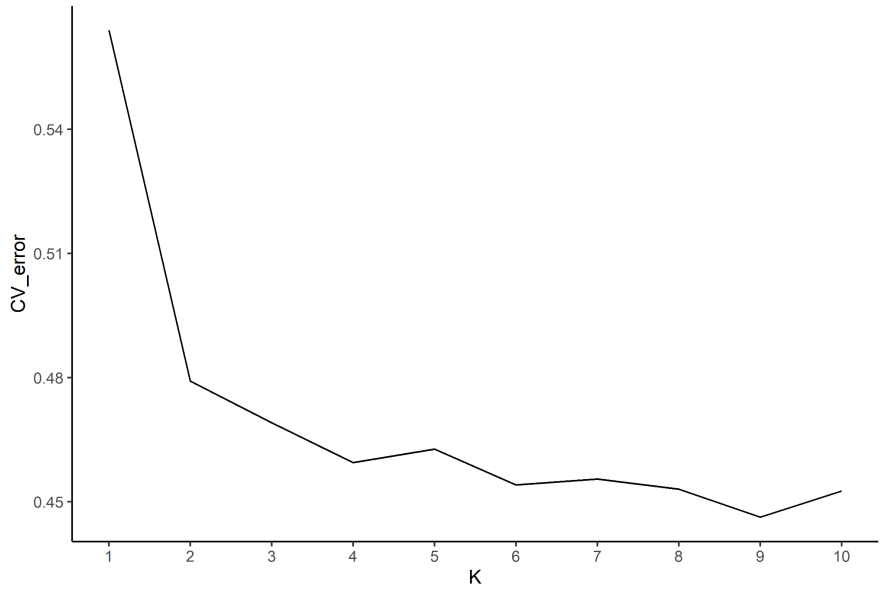


Figure 2. Cross-validation errors of the 10 ancestral components (K) tested with ADMIXTURE.

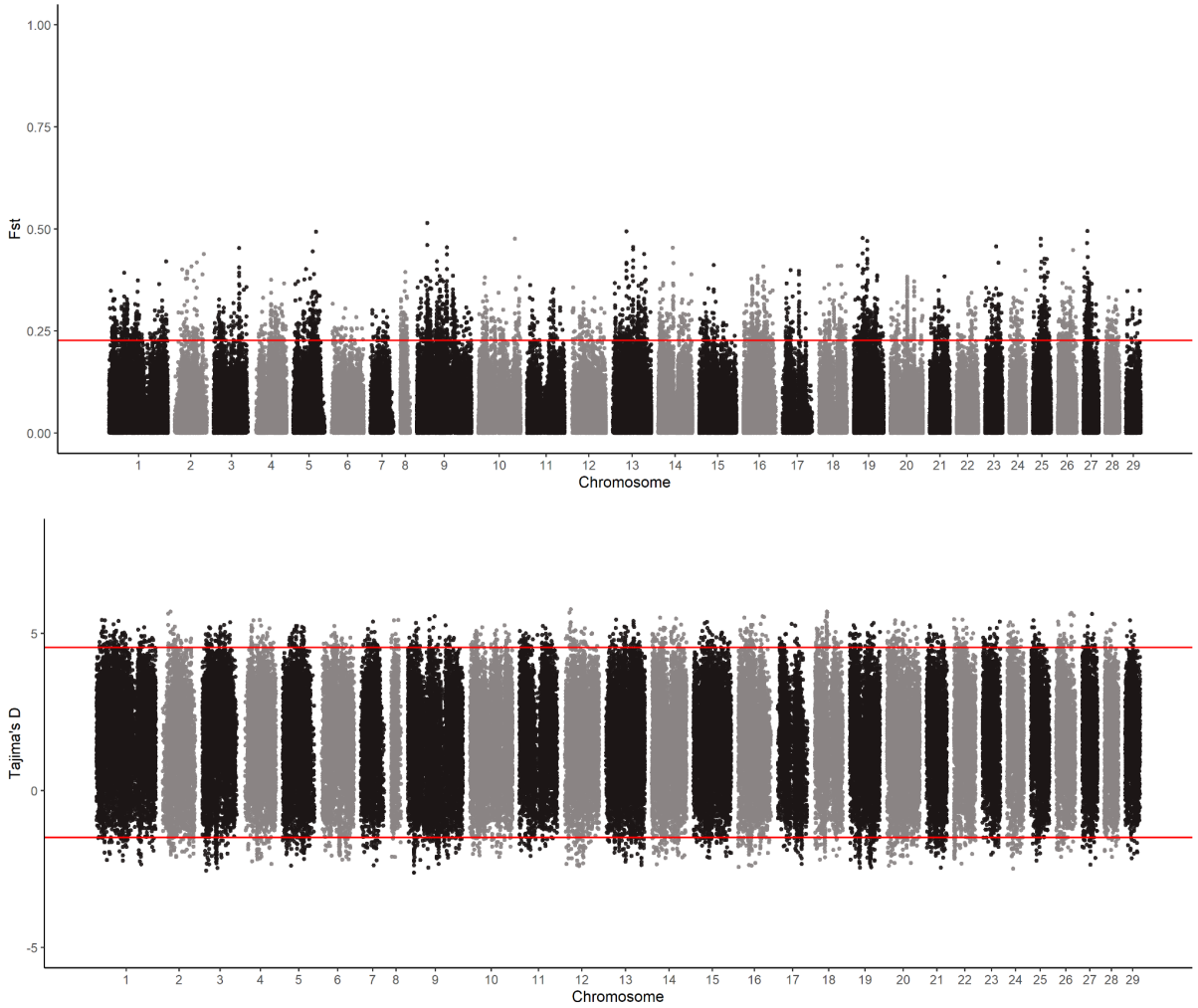


Figure 3. Manhattan plots showing the F_{ST} distribution and Tajima's D distribution alongside the chromosomes of the Norwegian populations.

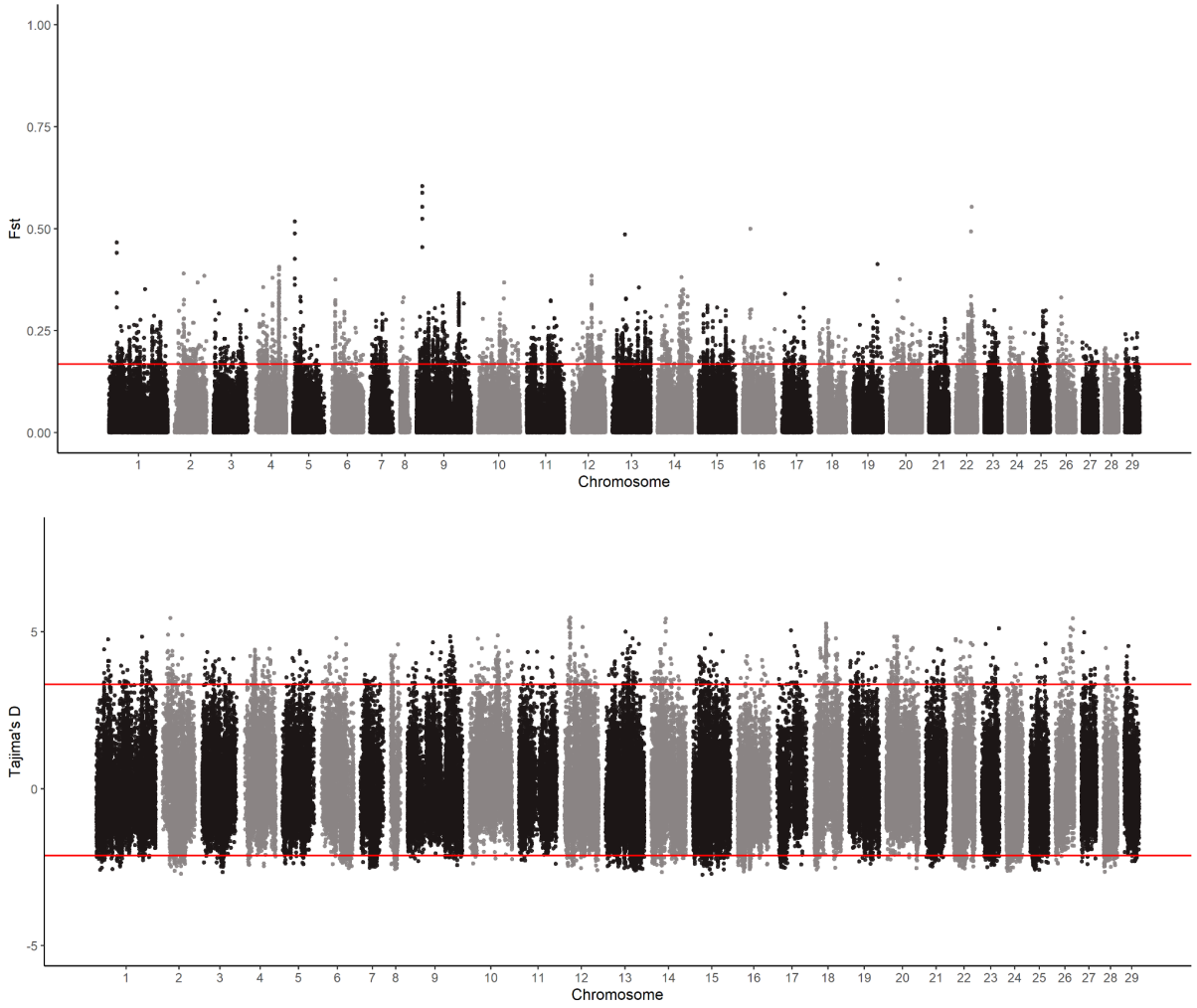


Figure 4. Manhattan plots showing the (a) F_{ST} distribution and (b) Tajima's D distribution alongside the chromosomes of the Canadian populations.

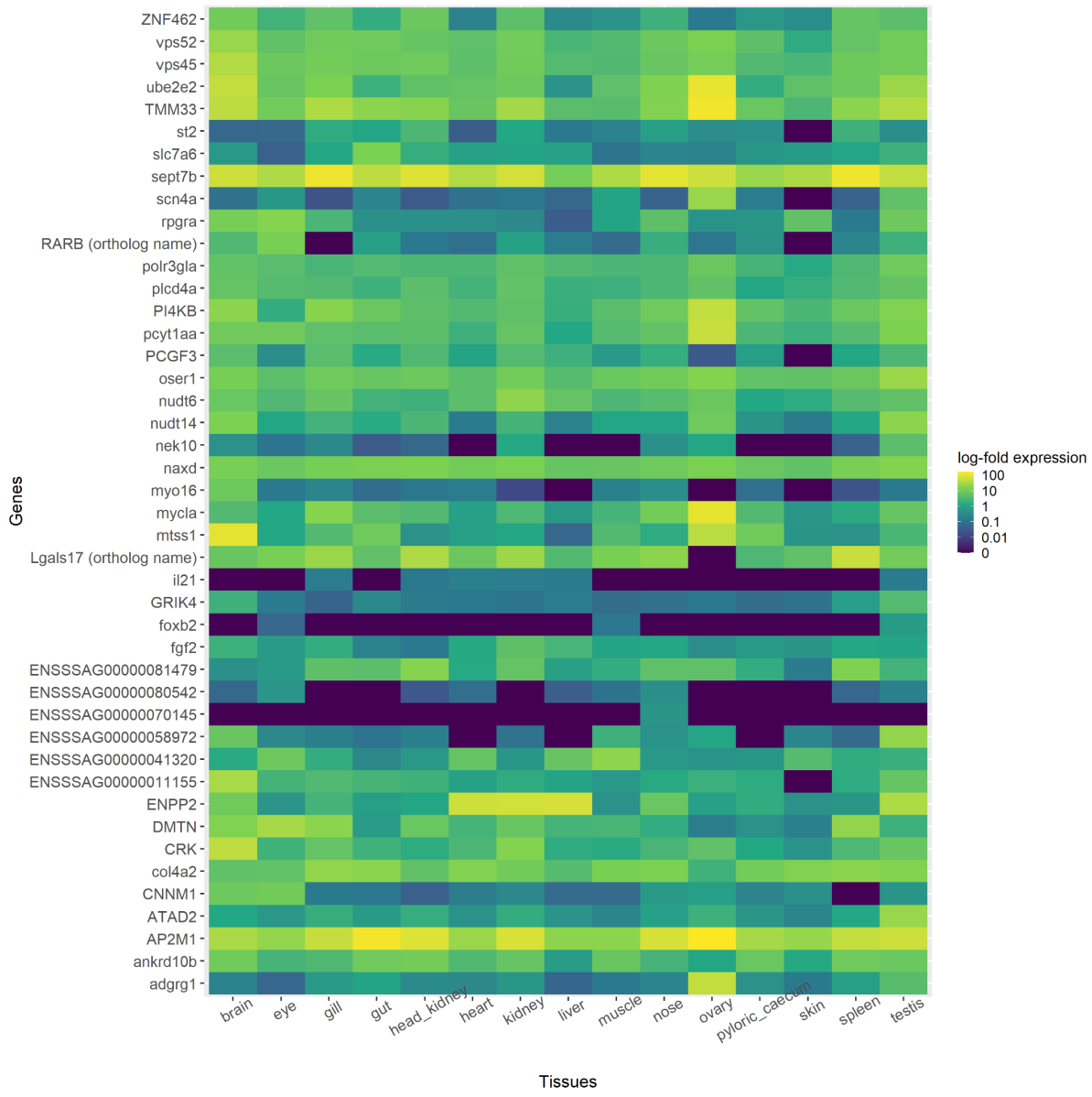


Figure 5. Expression level of all the genes overlapping with significant variants from the three intersect (F_{ST}/T_D, F_{ST}/XP-EHH) in Norwegian populations.

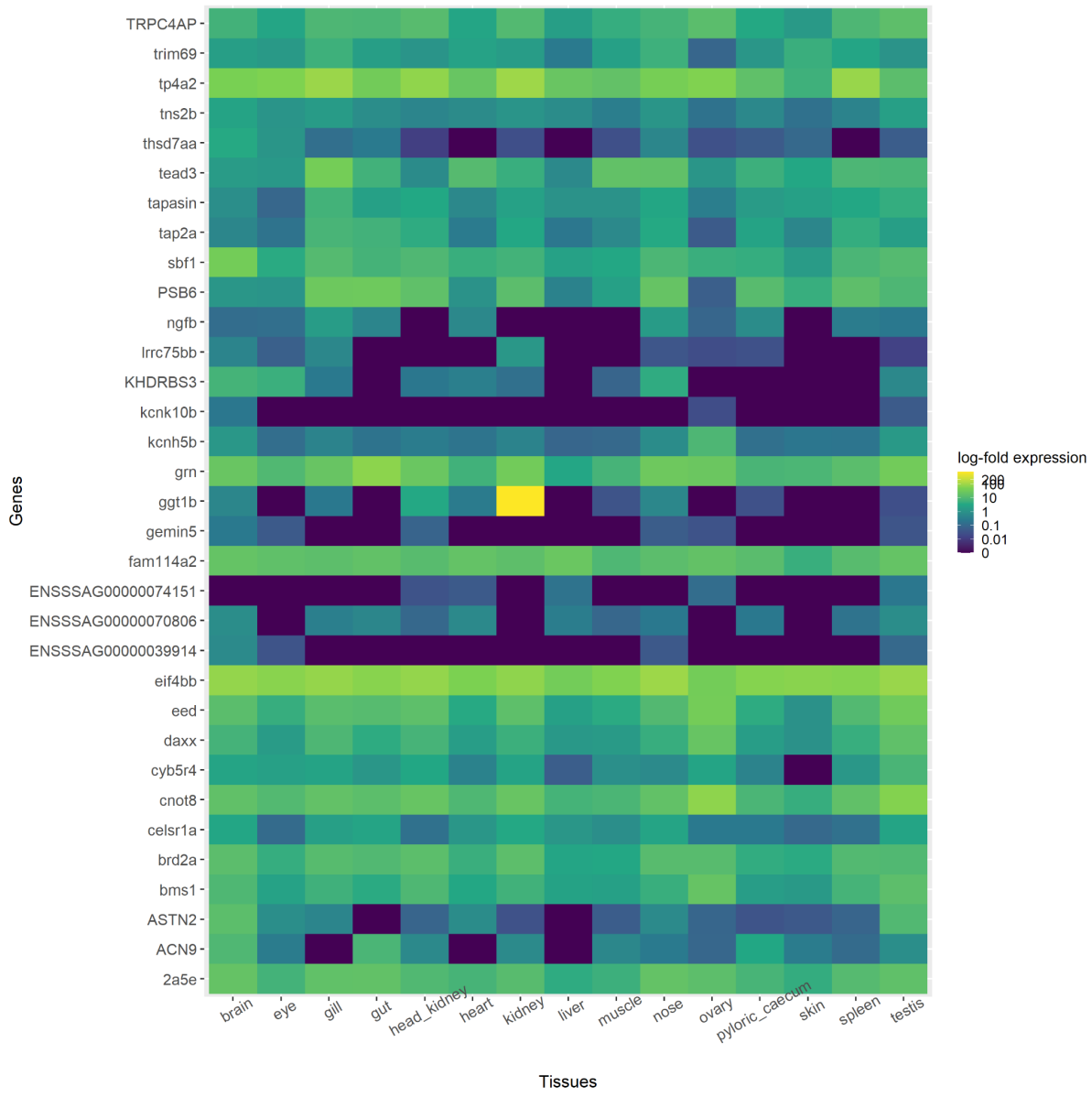


Figure 6. Expression level of all the genes overlapping with significant variants from the two intersects (F_{ST}/T_D, F_{ST}/XP-EHH) in Canadian populations.

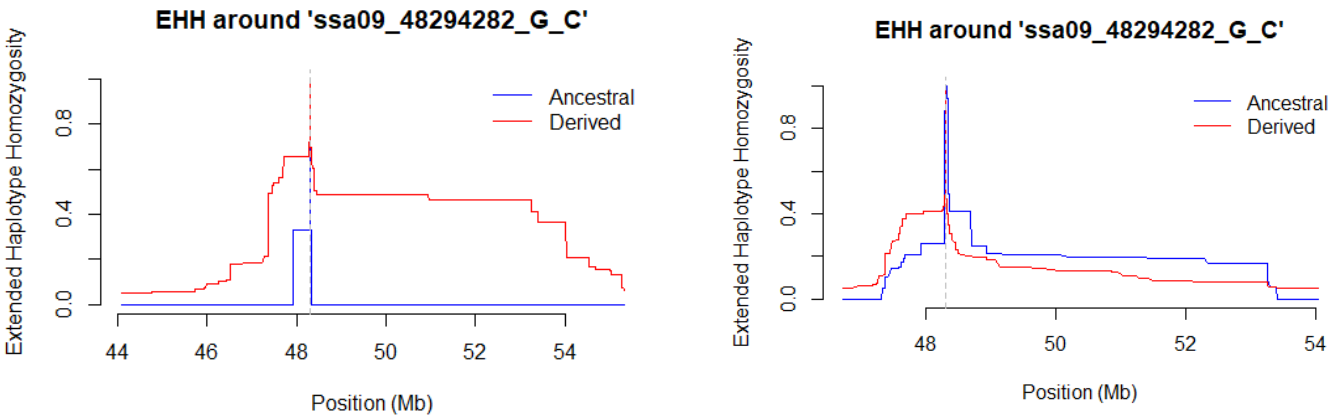


Figure 7. Extended Haplotype Homozygosity (EHH) patterns in farmed (on the left) and wild strain (on the right) around the locus ssa09:48294282 identified in the Canadian populations. The plots show which allele (ancestral or derived) possesses the extended haplotype and which therefore is under selection. In this example, the derived allele in the breeding population corresponds to the extended haplotype, whereas in the wild, both alleles seem to have similar patterns.

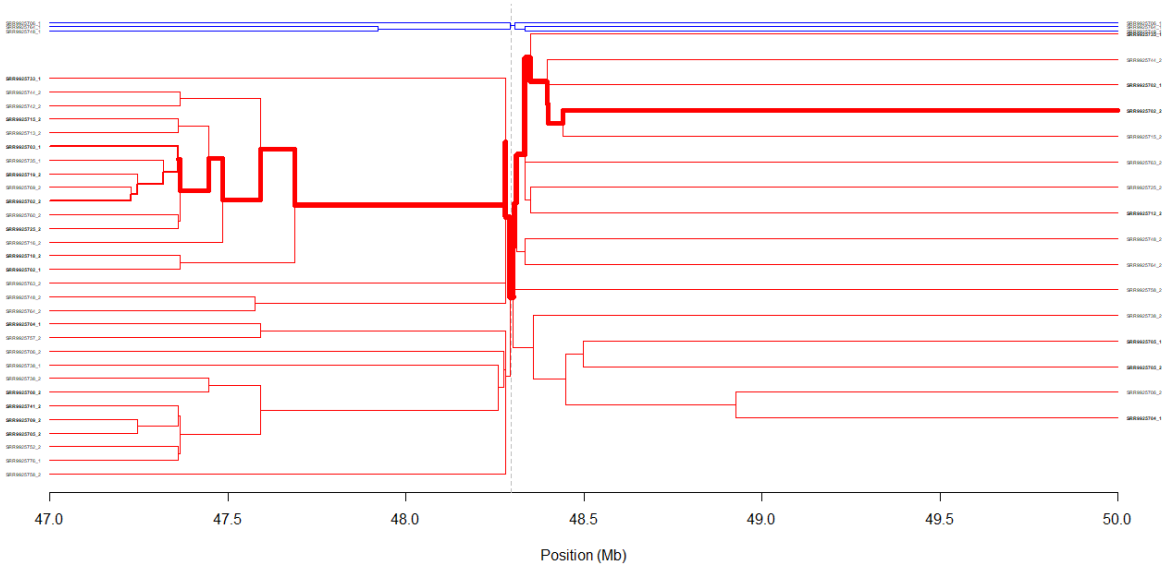


Figure 8. Bifurcation plot of the marker ssa09:48294282 in the Canadian farmed population. The plot shows the pattern of haplotype splitting as it moves from the focal allele. The fatter the line the more individuals share the haplotype. The selected allele will have a fatter line and will split less than the alternative allele that is not selected. The red corresponds to the pattern of the derived allele, the blue to that of the ancestral allele. Here we see that the derived allele is close to fixation and that there is very low trace of the reference allele. This is in agreement with the EHH plots in Figure 7 where the extended haplotype corresponds to the derived allele in the farmed population. This is also in agreement with the high frequency of the derived allele in the farmed population (see Table 2 of the report). This suggests that the selection occurs on the derived allele.

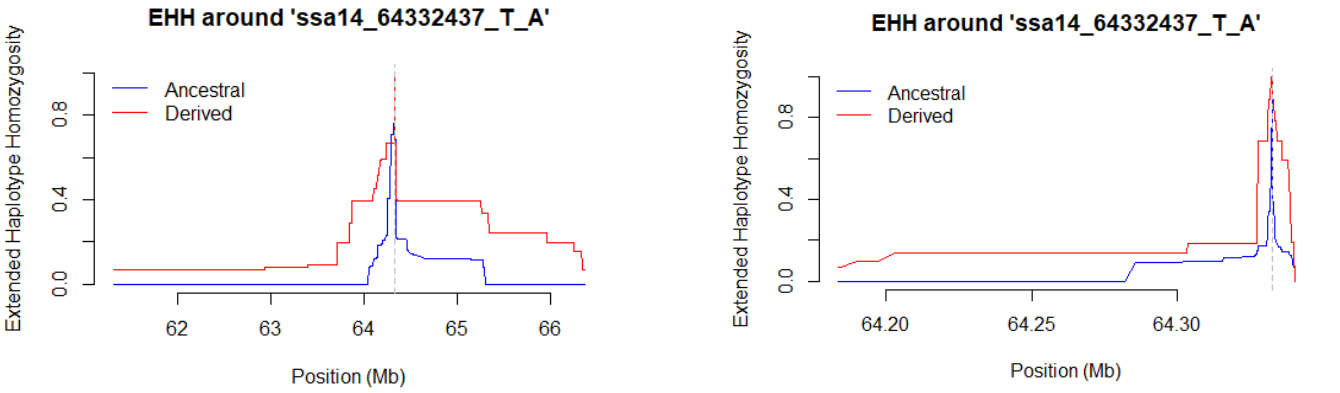


Figure 9. Extended Haplotype Homozygosity (EHH) patterns in farmed (on the left) and wild strain (on the right) around the locus ssa14:64332437 identified in the Canadian populations. The plots show which allele (ancestral or derived) possesses the extended haplotype and which therefore is under selection.
Here, in the wild both derived and ancestral are very tight peaks, in the farmed both look more extended, with the derived looking more so. If both alleles are extended in the farmed, it could be explained by bottleneck and/or genetic drift. Derived allele is extra extended – maybe because it is rarer and so more heavily impacted by drift and lower effective population size.

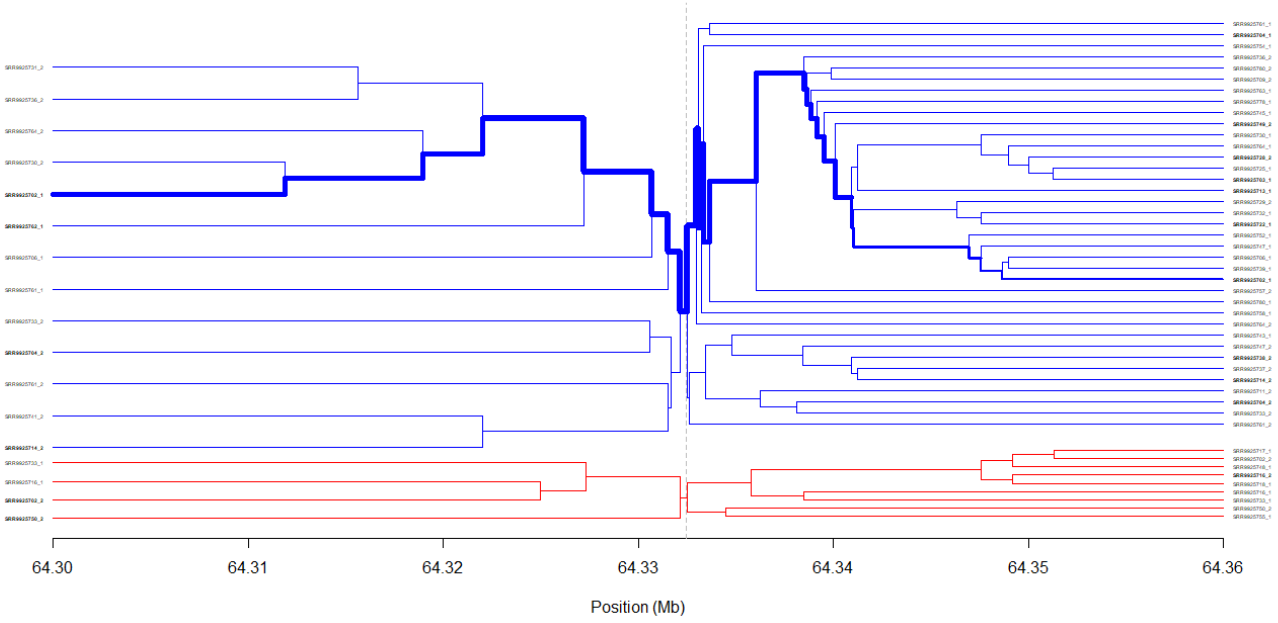


Figure 10. Bifurcation plot of the marker ssa14:64332437 in the Canadian farmed population. The plot shows the pattern of haplotype splitting as it moves from the focal allele. The fatter the line the more individuals share the haplotype. The selected allele will have a fatter line and will split less than the alternative allele that is not selected. The red corresponds to the pattern of the derived allele, the blue to that of the ancestral allele. Here we see that the ancestral allele is close to fixation and that there is very low trace of the derived allele. The frequency of the ancestral is higher (see Table 2 of the report). In the bifurcation

plots, the impact is the majority if individuals are extended for the ancestral allele even though the EHH plots in Figure 9. show that both alleles are extended.

Gene ID	Gene name	Chr	Position (Mbp)	Overlapping position	Description	Intersect	Predicted role
ENSSSAG00000003879	mycla	ssa01	21.4890	1	protein L-Myc-1b-like, DNA-binding transcription factor activity. Regulation of transcription by RNA polymerase II. (zebrafish)	FST / Tajima	RNA transcription regulation
ENSSSAG00000003216	ZNF462	ssa01	109.3549	1	zinc finger protein Z62-like, neuromast	FST / Tajima	sensing of mechanical changes in water
ENSSSAG00000080542	NA	ssa01	115.1140	1	cytochrome P450 2D15-like	FST / Tajima	uncharacterized
ENSSSAG00000011155	NA	ssa01	153.3515	1	WD repeat-containing protein 7-like	FST / Tajima	uncharacterized
ENSSSAG00000058972	NA	ssa02	22.1746	1	collagen alpha-2(IX) chain-like, Focal adhesion, ECM-receptor interaction	FST / Tajima	cellular process
ENSSSAG00000040220	GRIK4	ssa04	63.6860	3	glutamate receptor ionotropic, kainate 4-like	FST / Tajima	signalling pathway
ENSSSAG00000112902	U7	ssa04	69.236306	1	U7 small nuclear RNA	FST / Tajima	RNA
ENSSSAG00000003872	sept7b	ssa05	54.9231	4	septin-7	FST / XP-EHH	uncharacterized
ENSSSAG00000004894	nek10	ssa05	54.9356	8	NIMA related kinase 10, mitosis regulation, MAP kinase activity.	FST / XP-EHH	cellular process
ENSSSAG00000004993	rarb (ortholog name)	ssa05	54.9541	14	retinoic acid receptor beta-like	FST / XP-EHH	cellular signaling in embryonic morphogenesis, cell growth and differentiation
ENSSSAG00000110907	ube2e2	ssa05	57.610828	3	ubiquitin-conjugating enzyme E2E2	FST / XP-EHH	uncharacterized

ENSSSAG00000017268	TMM33	ssa09	53.1165	2	transmembrane protein 33-like	FST / XP-EHH	cellular process, signaling pathway
ENSSSAG00000073600	il21	ssa09	58.6459	1	interleukin 21, cytokine receptor binding activity. Predicted to act upstream of or within immune response	FST / XP-EHH	immune system
ENSSSAG00000073609	nudt6	ssa09	58.6893	1	nudix hydrolase 6, ADP-ribose diphosphatase activity; NAD binding activity; and NADH pyrophosphatase activity	FST / XP-EHH	uncharacterized
ENSSSAG00000048732	PCGF3	ssa09	59.5914	1	polycomb group RING finger protein 3	FST / Tajima	energy regulation
ENSSSAG00000114007	fgf2	ssa09	64.958717	1	fibroblast growth factor receptor binding activity	FST / XP-EHH	fibroblast
ENSSSAG00000067569	adgrg1	ssa11	14.4545	1	receptor-interacting serine/threonine-protein kinase 2-like, Predicted to enable G protein-coupled receptor activity. Acts upstream of or within glial cell development and regulation of oligodendrocyte progenitor proliferation (zebrafish)	FST / Tajima	uncharacterized
ENSSSAG00000067592	slc7a6	ssa11	14.4689	1	Y+L amino acid transporter 2-like, amino acid transmembrane transport.	FST / Tajima	signalling pathway
ENSSSAG00000080076	CRK	ssa13	60.7220	1	adapter molecule crk-like, Focal adhesion, ECM-receptor interaction	FST / Tajima	cellular process
ENSSSAG00000075127	AP2M1	ssa14	13.9406	1	AP-2 complex subunit mu-A-like	FST / XP-EHH	signalling pathway
ENSSSAG00000075137	pcyt1aa	ssa14	13.9474	1	choline-phosphate cytidylyltransferase A-like, Phosphonate and phosphinate metabolism	FST / XP-EHH	metabolism
ENSSSAG00000068404	PI4KB	ssa14	60.8018	11	phosphatidylinositol 4-kinase beta-like, Inositol phosphate metabolism, phosphatidylinositol signaling system	FST / XP-EHH	metabolism
ENSSSAG00000041320	rarb (ortholog name)	ssa14	14.4728	1	retinoic acid receptor RXR-gamma-A-like, cellular signaling in embryonic morphogenesis, cell growth and differentiation	FST / Tajima	cellular signaling in embryonic morphogenesis, cell growth and

							differentiation
ENSSSAG00000064498	nudt14	ssa15	39.4884	2	uridine diphosphate glucose pyrophosphatase-like, UDP-sugar diphosphatase activity, nucleoside phosphate metabolic process and ribose phosphate metabolic process	FST / Tajima	metabolism
ENSSSAG00000076833	plcd4a	ssa16	64.4833	1	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase delta-4-like, Inositol phosphate metabolism, phosphatidylinositol signaling system	FST / Tajima	metabolism
ENSSSAG00000070145	NA	ssa19	31.2381	9	NA	FST / XP-EHH	uncharacterized
ENSSSAG00000063232	scn4a	ssa19	57.6034	1	sodium voltage-gated channel alpha subunit 4, membrane depolarization during action potential; neuronal action potential; sodium transmembrane transport	FST / Tajima	nervous system, signaling pathway
ENSSSAG00000026781	ankrd10b	ssa21	26.8790	5	ankyrin repeat domain-containing protein 10-like, family of proteins that mediate the attachment of integral membrane proteins to the spectrin-actin based membrane cytoskeleton	FST / Tajima	cellular process
ENSSSAG00000027563	naxd	ssa21	26.9314	2	ATP-dependent (S)-NAD(P)H-hydrate dehydratase	FST / Tajima	metabolism
ENSSSAG00000029049	col4a2	ssa21	26.9507	4	collagen alpha-2(IV) chain-like, Focal adhesion, ECM-receptor interaction, encodes one of the six subunits of type IV collagen, the major structural component of basement membranes	FST / Tajima	cellular process
ENSSSAG00000031736	myo16	ssa21	27.3415	4	myosin XVI, actin filament binding activity and protein phosphatase binding activity	FST / Tajima	cellular process
ENSSSAG00000098075	rpgra	ssa21	26.758532	1	retinitis pigmentosa GTPase regulator A	FST / Tajima	view
ENSSSAG00000050524	oser1	ssa22	3.3568	1	oxidative stress responsive serine rich 1	FST / Tajima	cellular process
ENSSSAG00000069177	foxb2	ssa24	30.8314	1	forkhead box B2, DNA-binding transcription factor activity, anatomical structure morphogenesis; cell differentiation; and regulation of transcription by RNA polymerase (zebrafish)	FST / Tajima	uncharacterized
ENSSSAG00000081479	NA	ssa25	16.4799	1	NA	FST / Tajima	uncharacterized
ENSSSAG00000081480	st2	ssa25	16.5100	1	ST2L protein-like	FST / Tajima	uncharacterized

ENSSSAG00000099637	DMTN	ssa26	2.235293	1	dematin actin binding protein, actin filament binding activity, cellular skeleton	FST / Tajima	cellular process
ENSSSAG00000078638	vps52	ssa27	10.3993	8	VPS52 subunit of GARP complex	FST / XP-EHH	signalling pathway
ENSSSAG00000078726	Lgals17 (ortholog name)	ssa27	10.4101	8	cell surface receptor signaling pathway; cellular response to lipopolysaccharide; and regulation of T cell activation	FST / XP-EHH	signaling pathway, immune system
ENSSSAG00000078859	ATAD2	ssa27	10.7061	9	ATPase family AAA domain containing 2	FST / XP-EHH	metabolism
ENSSSAG00000079724	ENPP2	ssa27	10.8859	1	ectonucleotide pyrophosphatase/phosphodiesterase family member 2-like	FST / XP-EHH	metabolism
ENSSSAG00000019850	mtss1	ssa27	11.0584	7	secretagogin-like, actin binding activity and phospholipid binding activity, membrane structure	FST / XP-EHH	cellular process
ENSSSAG00000040623	vps45	ssa27	11.8719	1	vacuolar protein sorting-associated protein 45-like	FST / XP-EHH	cellular process
ENSSSAG00000063338	polr3gla	ssa27	14.7651	2	DNA-directed RNA polymerase III subunit RPC7-like	FST / XP-EHH	RNA transcription regulation
ENSSSAG00000075572	CNNM1	ssa28	25.5312	1	metal transporter CNNM4-like	FST / Tajima	metabolism, signaling pathway

Table S1. Predicted function of the genes detected from the F_{ST}/T_D and F_{ST}/XP-EHH intersects in the Norwegian farmed population.

Gene ID	Gene name	Chr	Overlapping Position	Position (Mpb)	Description	Intersect	Predicted Role
ENSSSAG00000055271	cyb5r4	ssa02	1	18.5453	cytochrome b5 reductase 4-like, heme binding activity, generation of precursor metabolites and energy and superoxide metabolic process.	FST/XP-EHH	metabolism
ENSSSAG00000059605	bms1	ssa02	1	22.4105	ribosome biogenesis protein BMS1 homolog, GTP binding activity; GTPase activity; and U3 snoRNA binding activity. Acts upstream of or within liver development; pancreas development; and rRNA processing (zebrafish).	FST/Tajima	cellular process
ENSSSAG00000008083	celsr1a	ssa07	1	32.2079	cadherin EGF LAG seven-pass G-type receptor 1-like, G protein-coupled receptor activity and calcium ion binding activity. Acts upstream of or within several processes, including cell migration involved in gastrulation; neural tube development; and stem cell population maintenance (zebrafish).	FST/Tajima	cellular process
ENSSSAG00000071635	sbf1	ssa07	2	41.3447	myotubularin-related protein 5-like, guanyl-nucleotide exchange factor activity and phosphatase regulator activity (zebrafish). Predicted to act upstream of or within regulation of GTPase activity.	FST/Tajima	metabolism
ENSSSAG00000005294	NA	ssa09	5	45.8712	NA	FST/XP-EHH	uncharacterized
ENSSSAG00000008640	kcnk10b	ssa09	1	29.8722	potassium channel, subfamily K, member 10b, stabilization of membrane potential	FST/XP-EHH	cellular process, signaling pathway

ENSSSAG00000044391	2a5e	ssa09	10	25.3318	Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit epsilon isoform, Intracellular protein-containing complex	FST/XP-EHH	uncharacterized
ENSSSAG00000088169	kcnh5b	ssa09	2	24.4131	potassium voltage-gated channel, potassium ion transmembrane transport and regulation of membrane potential.	FST/XP-EHH	cellular process, signaling pathway
ENSSSAG00000055525	ASTN2	ssa11	32	66.5480	astrotactin-2-like	FST/XP-EHH	uncharacterized
ENSSSAG00000071165	tead3	ssa12	3	79.0473	TEA domain transcription factor 3, Predicted to enable DNA-binding transcription factor activity, RNA polymerase II-specific and RNA polymerase II cis-regulatory region sequence-specific DNA binding activity. Acts upstream of or within Kupffer's vesicle development; cilium assembly; and determination of heart left/right asymmetry (zebrafish).	FST/XP-EHH	early development, RNA-transcription factor
ENSSSAG00000058310	cnot8	ssa13	7	72.0391	CCR4-NOT transcription complex subunit 8, Intracellular protein-containing complex (orthologs Rainbow trout and Coho salmon)	FST/XP-EHH	transcription factor
ENSSSAG00000058323	fam114a2	ssa13	2	72.0425	family with sequence similarity 114 member A2	FST/XP-EHH	uncharacterized
ENSSSAG00000072499	TRPC4AP	ssa13	1	9.5893	short transient receptor potential channel 4-associated protein-like, cell control factor (mouse), phosphatase binding activity. Predicted to be involved in ubiquitin-dependent protein catabolic process (zebrafish)	FST/XP-EHH	Intracellular protein-containing complex
ENSSSAG00000077251	trim69	ssa13	1	85.8300	zinc-binding protein A33-like	FST/Tajima	signaling pathway
ENSSSAG00000079278	ggt1b	ssa13	1	86.3237	gamma-glutamyl transpeptidase 1-like, Threonine-type peptidase activity.	FST/Tajima	

ENSSSAG00000101250	lrrc75bb	ssa13	1	90.930	leucine rich repeat containing 75Bb	FST/Tajima	uncharacterized
ENSSSAG00000102970	gemin5	ssa13	4	76.636	gem (nuclear organelle) associated protein 5, Predicted to enable mRNA 3'-UTR binding activity. Acts upstream of or within T cell differentiation and neuromast regeneration (zebrafish)	FST/XP-EHH	RNA transcription factor, immune system
ENSSSAG00000009401	KHDRBS3	ssa14	22	65.1712	KH domain-containing, RNA-binding, signal transduction-associated protein 3-like	FST/XP-EHH	RNA binding, signaling pathway
ENSSSAG00000040701	daxx	ssa14	83	59.0265	death domain-associated protein 6-like, apoptose, immune response	FST/XP-EHH	apoptose, immune response
ENSSSAG00000040723	tapasin	ssa14	52	59.0405	tapasin-like, transmembrane glycoprotein that mediates interaction between newly assembled major histocompatibility complex (MHC) class I molecules and the transporter associated with antigen processing (TAP), which is required for the transport of antigenic peptides across the endoplasmic reticulum membrane. (human, mouse, zebrafish)	FST/XP-EHH	Immune system
ENSSSAG00000041924	PSB6	ssa14	7	59.1031	proteasome subunit beta type-6-B like protein, Threonine-type peptidase activity, Intracellular protein-containing complex	FST/XP-EHH	uncharacterized
ENSSSAG00000042614	brd2a	ssa14	7	59.1627	bromodomain-containing protein 2-like, lysine-acetylated histone binding activity. Acts upstream of or within central nervous system development; regulation of apoptotic process; and thrombocyte differentiation (zebrafish)	FST/XP-EHH	uncharacterized
ENSSSAG00000046480	tp4a2	ssa14	27	81.6219	tyrosine phosphatase type IVA 2	FST/XP-EHH	uncharacterized
ENSSSAG00000070806	NA	ssa14	1	10.3424	NA	FST/XP-EHH	uncharacterized
ENSSSAG00000073182	ACN9	ssa14	4	69.0405	protachykinin-like	FST/XP-EHH	uncharacterized

ENSSSAG00000073406	thsd7aa	ssa14	48	70.0418	thrombospondin type-1 domain-containing protein 7A-like, angiogenic intersegmental vessels	FST/XP-EHH	early development
ENSSSAG00000093386	tap2a	ssa14	5	64.476	transporter associated with antigen processing	FST/XP-EHH	immune system
ENSSSAG00000074151	NA	ssa15	1	35.6154	NA	FST/Tajima	uncharacterized
ENSSSAG00000039914	NA	ssa16	6	59.3227	cholecystokinin receptor-like, peptides with important roles in gastrointestinal and neural physiology.	FST/XP-EHH	metabolism, nervous system
ENSSSAG00000067620	eif4bb	ssa22	3	47.1487	eukaryotic translation initiation factor 4B, RNA binding activity	FST/XP-EHH	RNA binding activity
ENSSSAG00000067634	NA	ssa22	2	47.1556	NA	FST/XP-EHH	uncharacterized
ENSSSAG00000077182	ngfb	ssa22	1	8.7198	nerve growth factor receptor binding activity	FST/XP-EHH	nervous system
ENSSSAG00000120547	tns2b	ssa22	4	47.198	tensin 2b, focal adhesion	FST/XP-EHH	cellular process
ENSSSAG00000076429	grn	ssa28	1	2.8275	Granulins, Acts upstream of or within several processes, including neuron differentiation; sensory system development; and skeletal muscle acetylcholine-gated channel clustering (zebra fish)	FST/Tajima	nervous system, muscle contraction
ENSSSAG00000114830	eed	ssa17	2	26.750417	ectoderm development, regulation of kisspeptin pathway	FST/XP-EHH	early development, reproduction

Table S2. Predicted function of the genes detected from the F_{ST}/T_D and F_{ST}/XP-EHH intersects in the Canadian farmed population.