

RESEARCH ARTICLE

Open Access



Atlantic salmon populations reveal adaptive divergence of immune related genes - a duplicated genome under selection

Erik Kjærner-Semb^{1,2*}, Fernando Ayllon¹, Tomasz Furmanek¹, Vidar Wennevik¹, Geir Dahle¹, Eero Niemelä³, Mikhail Ozerov⁴, Juha-Pekka Vähä^{4,6}, Kevin A. Glover^{1,2}, Carl J. Rubin⁵, Anna Wargelius¹ and Rolf B. Edvardsen^{1*}

Abstract

Background: Populations of Atlantic salmon display highly significant genetic differences with unresolved molecular basis. These differences may result from separate postglacial colonization patterns, diversifying natural selection and adaptation, or a combination. Adaptation could be influenced or even facilitated by the recent whole genome duplication in the salmonid lineage which resulted in a partly tetraploid species with duplicated genes and regions.

Results: In order to elucidate the genes and genomic regions underlying the genetic differences, we conducted a genome wide association study using whole genome resequencing data from eight populations from Northern and Southern Norway. From a total of ~4.5 million sequencing-derived SNPs, more than 10 % showed significant differentiation between populations from these two regions and ten selective sweeps on chromosomes 5, 10, 11, 13–15, 21, 24 and 25 were identified. These comprised 59 genes, of which 15 had one or more differentiated missense mutation. Our analysis showed that most sweeps have paralogous regions in the partially tetraploid genome, each lacking the high number of significant SNPs found in the sweeps. The most significant sweep was found on Chr 25 and carried several missense mutations in the antiviral *mx* genes, suggesting that these populations have experienced differing viral pressures. Interestingly the second most significant sweep, found on Chr 5, contains two genes involved in the NF-KB pathway (*nkap* and *nkrf*), which is also a known pathogen target that controls a large number of processes in animals.

Conclusion: Our results show that natural selection acting on immune related genes has contributed to genetic divergence between salmon populations in Norway. The differences between populations may have been facilitated by the plasticity of the salmon genome. The observed signatures of selection in duplicated genomic regions suggest that the recently duplicated genome has provided raw material for evolutionary adaptation.

Keywords: Whole genome duplication, Adaptation, Aquaculture, Immune system, GWAS, Resequencing, Selective sweep, SNPs, *Salmo salar*

Background

In addition to being one of the most highly prized freshwater fish for recreational fishing, the Atlantic salmon (*Salmo salar* L.) is one of the most economically important aquaculture species worldwide. Its natural distribution is throughout the North Atlantic, ranging from Long Island Sound to Ungava Bay in the west and from Northern Portugal to the Barents Sea in the east [1]. This

distribution is the result of postglacial colonization of ecosystems that became available when the glacial ice retreated about 10,000 years ago [2].

Atlantic salmon is characterised by highly significant, hierarchically structured population genetic divergence, with the largest differences observed between the European and North American lineages [3–5]. This divergence is also observed on a regional scale, presumably as a consequence of the colonization process associated with the retreat of the glacier [6, 7]. Moreover, local scale differentiation exists, for example between neighbouring rivers [8–10] and

* Correspondence: erikj@imr.no; rolfb@imr.no

¹Institute of Marine Research, Bergen, Norway

Full list of author information is available at the end of the article



among tributaries within the same river which might be explained by restricted gene flow, genetic drift and adaptation [11–13].

Atlantic salmon exhibit a relatively complex life history that includes spawning and juvenile rearing in freshwater followed by extended ocean migrations to the feeding grounds [14]. As a consequence, salmon go through several distinct transitions that are characterized by changes in behaviour and physiology [15]. They are also able to adapt to varying local conditions throughout their range of environments [16], exemplified by their ability to inhabit rivers with a wide range of temperatures, from Spain to the colder Arctic latitudes [17]. Previous studies have shown differences in temperature and climate to be associated with genetic differences between salmon populations [7, 18], and latitude also seems to be correlated with allele frequencies of markers relevant to immune response in American and European Atlantic salmon populations, possibly due to temperature induced differences in pathogen-driven selection or other environmental factors [19–21].

In the wild, Atlantic salmon are constantly confronted with a range of pathogens, and have consequently developed numerous innate and adaptive immune mechanisms to overcome infectious challenges [22]. Recent studies suggest that the prevalence of parasites and infectious diseases is increasing in wild populations partly due to global warming [23, 24]. Given the commercial relevance of Atlantic salmon and the recent release of a reference genome [25], particular effort should be made to identify genes targeted by natural selection in wild Atlantic salmon populations that ultimately can lead to optimized aquaculture practices. The potential relevance of these findings for the Atlantic salmon farming industry is exemplified by the identification of Infectious Pancreatic Necrosis (IPN) Virus resistance [26] and age at maturity associated genes [27, 28]. A relatively recent whole genome duplication occurred in the salmonid lineage some 80 million years ago [29], resulting in a partly tetraploid genome undergoing rediploidization. Consequently the genome contains many paralogous regions that could provide raw material for evolution as paralogous genes and regions can diversify and acquire new functions [30].

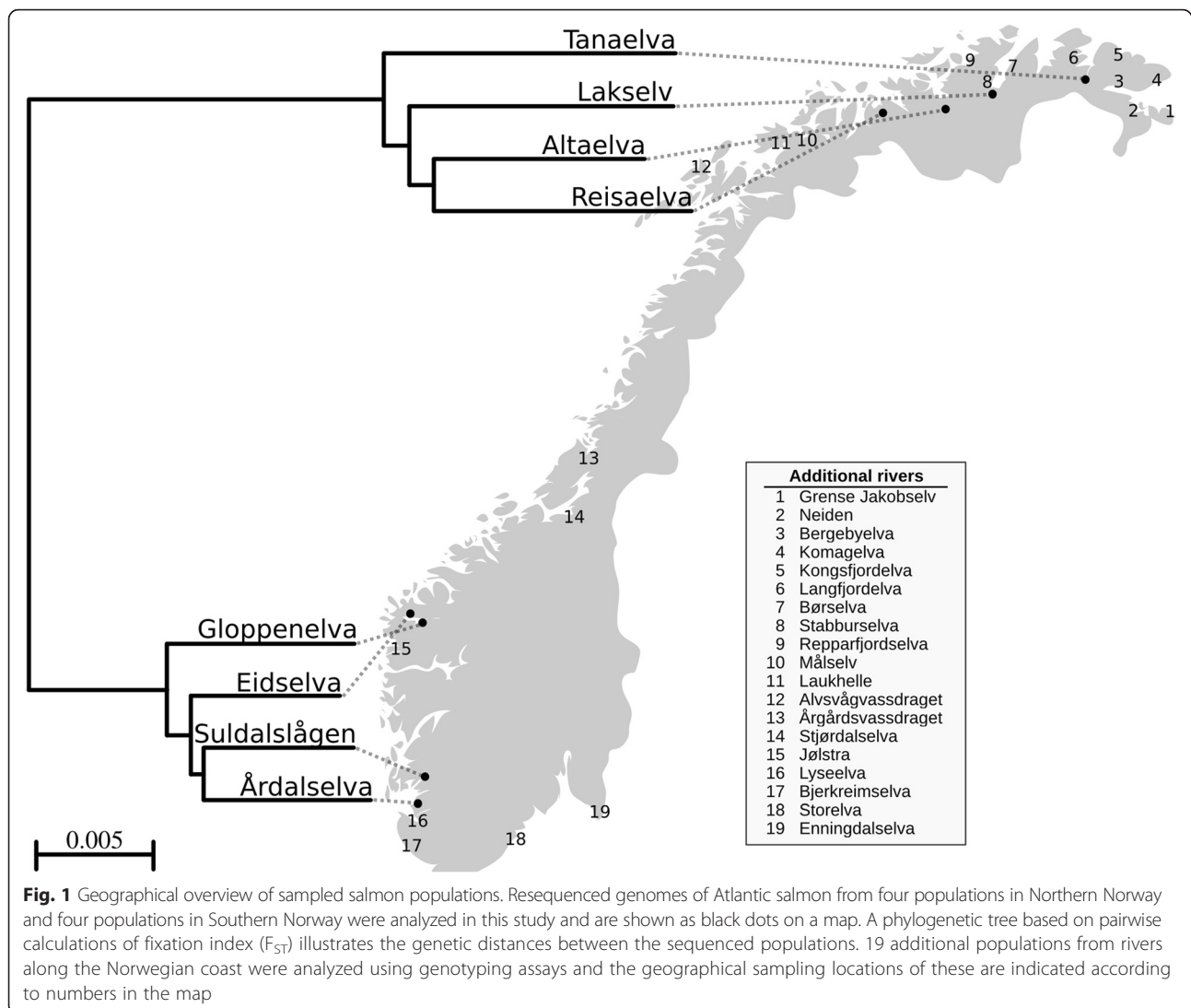
Based upon the analysis of microsatellite and SNP markers, several studies have demonstrated that there are highly significant genetic differences between Atlantic salmon populations located in the north and south of Norway [31–33]. However, the genomic regions and genes behind the differences have not been investigated in detail, and consequently, the potential adaptive significance of this genetic divergence remains elusive.

Recently, a genome wide association study (GWAS) based upon whole genome resequencing data revealed a

selective sweep in Atlantic salmon strongly associated with age of maturation [27]. Using a similar methodological approach, the present study aimed to identify genes and genomic regions diverging between Atlantic salmon populations in the north and south of Norway. In order to achieve this objective, salmon populations inhabiting the four rivers Tanaelva, Lakselv, Altaelva and Reisaelva from Northern Norway and the four rivers Gloppenelva, Eidselva, Suldalslågen and Årdalselva from Southern Norway were chosen for resequencing using DNA pools ($n = 30$ fish per river, Fig. 1). The major finding in this study was the observation that diversifying natural selection has acted on immune related genes causing adaptive divergence between populations in the north and south of Norway.

Results and discussion

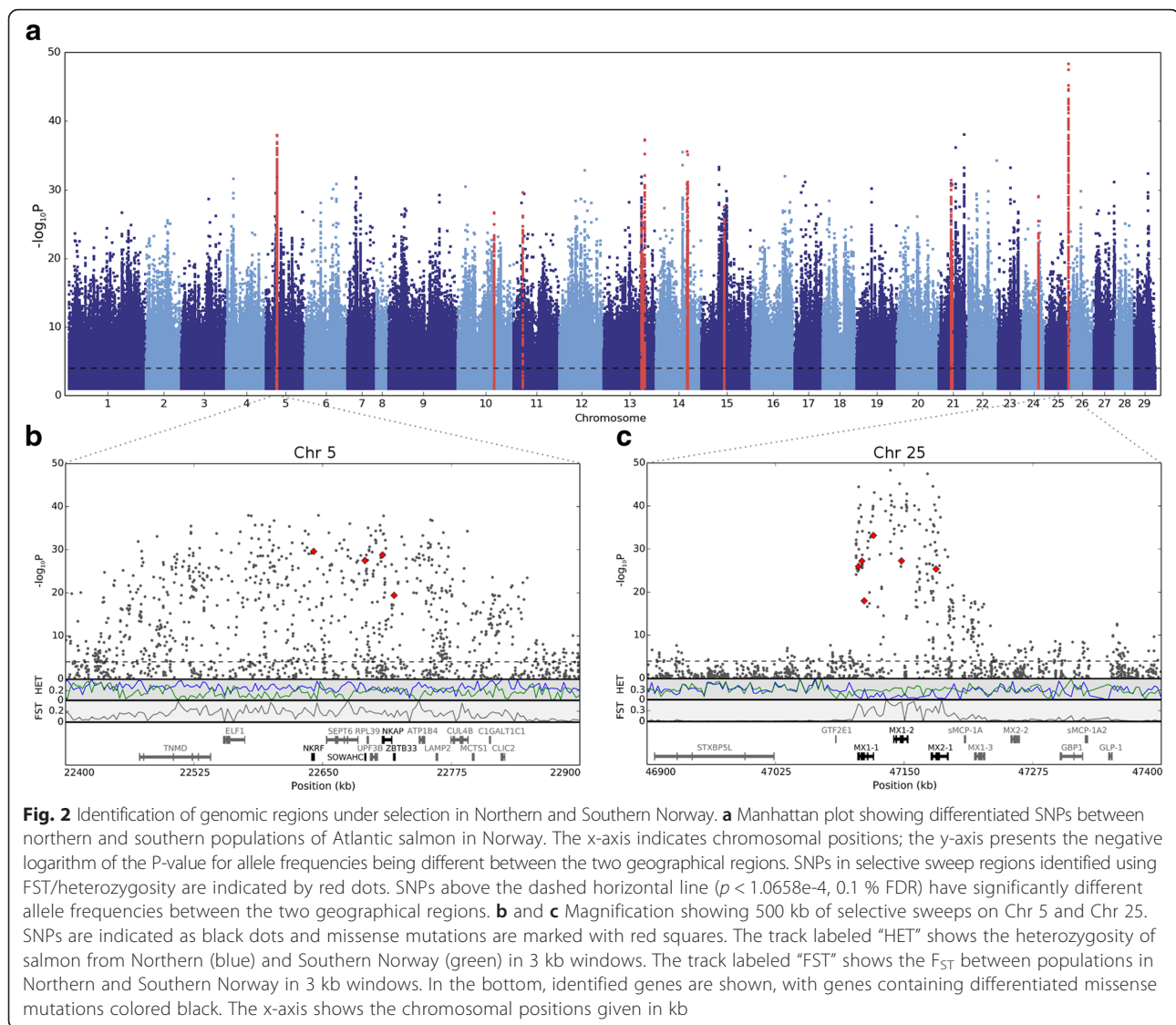
Whole genome sequence data from eight selected rivers along the Norwegian coast (Fig. 1) was mapped to the most recent Atlantic salmon reference genome (AKGD00000000.4). This yielded a 26.7× average depth of coverage of uniquely mapped reads per river. SNP calling revealed 4,450,990 high quality SNPs. To quantify the genetic difference between populations of the chosen rivers, Hudson's estimator for Wright's fixation index (F_{ST}) [34] was calculated (Additional file 1: Table S1). A phylogenetic tree was made using this distance matrix to illustrate and confirm the reported large genetic difference between the northern and southern populations of Atlantic salmon in Norway (Fig. 1). Statistical analysis using the Cochran-Mantel-Haenszel test for different allele frequencies between northern and southern Atlantic salmon populations revealed 474,410 SNPs with significantly different allele frequencies (0.1 % FDR, Fig. 2a). Genomic regions subjected to recent positive selection are expected to have lower heterozygosity than other regions, and if the selective pressure differs between populations, higher F_{ST} is observed [35]. An approach calculating F_{ST} and heterozygosity in 50 kb sliding windows has previously been used to identify genomic regions under selection (selective sweeps) [36]. This method was used to find selective sweeps which differ between northern and southern salmon populations in Norway (Additional file 1: Figure S1). The combined F_{ST} /heterozygosity approach suggested 10 selective sweeps that differed between the two geographical regions. The sweeps ranged from 75,000 to 575,000 bp in size, and were found in chromosomes 5, 10, 11, 13–15, 21, 24 and 25 (Table 1). These sweeps contained in total 59 genes involved in a number of different biological processes including cell division, cytokinesis, angiogenesis, development, transcriptional regulation and immune response. For a detailed list of gene ID and short description of function see Additional file 1: Table S2.



The high number of SNPs and genes in the selective sweeps complicates the task of pin-pointing the most important genetic differences. Therefore, we focused on missense mutations that induce amino acid changes in proteins, since these are more likely to confer a difference in biological function. Within the identified sweeps 20 significantly differentiated missense SNPs were found, comprising 15 different genes dispersed in 6 selective sweeps (Table 2). Three missense mutations were observed in the sweep on Chr 10, all in a single gene, *anln*, encoding an actin-binding protein required for cytokinesis. Three genes on Chr 13 harbor missense mutations: *trpc2*, involved in chemosensory transduction, and interestingly knockout mice display changes in their sexual, aggressive, and parenting behaviors [37]; *rrm1*, an enzyme essential for the production of deoxyribonucleotides; *rb1*, which promotes G0-G1 transition when phosphorylated by CDK3/cyclin-C acts as a transcriptional repressor of

E2F1 target genes. Also in the Chr 14 sweep there are three genes with missense mutations: *adnp*, a homeo-domain containing DNA binding transcription factor; *cpsf1*, encoding a component of the cleavage and polyadenylation specificity factor complex; *parp10*, encoding a ADP-ribosyltransferase involved in apoptosis, NF- κ B signaling, and DNA damage repair [38]. The sweep on Chr 21 contains one gene, *rnaseh2b*, which is linked to a chronic inflammatory disorder in humans [39].

The second most significant selective sweep was found on Chr 5 (Fig. 2b) and included the stress and immune response transcription factor genes *nkrf* and *nkap*; *zbtb33* encoding a transcriptional regulator binding to methylated CpG dinucleotides, and a gene with unknown function, *sowahc*. Both Nkrf and Nkap are transcription factors which regulate the NF- κ B pathway in which Nkap activates many cell processes including inflammation, immunity, differentiation, cell growth and apoptosis, while



Nkrf mediates transcriptional repression of certain Nkap responsive genes. Since NF- κ B signaling pathways activate the immune system in the host, these proteins are key targets for proteases expressed by invading pathogens [40]. Functional studies of the Nkap protein have revealed roles for this protein in T-cell maturation [41] and mRNA splicing [42]. To our knowledge, no previous studies have identified functionally significant SNPs associated with any of the four genes located within this sweep, however one of the SNPs found in *nkap* is located in a highly conserved region necessary for transcriptional repression. Here the valine is conserved in other species representing the ancestral variant while in Northern Norway methionine is most common (Additional file 1: Figure S2). This finding may be related to differences in immune defense between salmon from these two regions, a suggestion supported by the fact that the NF- κ B pathway is differently regulated in

IPN resistant salmon [43]. Further studies will reveal how these SNPs modulate the function of NF- κ B and virus response or if other functional properties are associated with the selective sweep on Chr 5.

The most significant sweep was found on Chr 25 and contained a cluster of five *mx* (myxovirus resistance) genes known to be involved in defense against viruses. Three of these *mx* genes contained missense mutations; *mx1-1*, *mx1-2* and *mx2-1* (Fig. 2c). These proteins are dynamin-like GTPases induced upon virus infection through the innate interferon system. It has been shown that they can act broadly against both DNA and RNA viruses and specifically against certain viruses [44] and studies in mouse, human and chicken have shown that single missense mutations in *Mx1* and *Mx2* can confer such specific responses [45–48]. It is possible that the identified missense SNPs in the *mx* genes reflect specific adjustments to different viral disease pressures between

Table 1 Selective sweeps. F_{ST} and heterozygosity estimations in sliding windows were used to identify differentiated loci undergoing selection in either northern or southern populations of Atlantic salmon in Norway. Genes carrying differentiated missense mutations are shown in bold

Chromosome	Chromosomal region	Sweep length (bp)	Genes in selective sweeps
5	22,475,000 – 22,800,000	325,000	<i>tnmd</i> , <i>elf1</i> , <i>nkrf</i> , <i>sept6</i> , <i>sowahc</i> , <i>rpl39</i> , <i>upf3b</i> , <i>nkap</i> , <i>zbtb33</i> , <i>atp1b4</i> , <i>lamp2</i> , <i>cul4b</i> , <i>mcts1</i> , <i>c1galt1c1</i> , <i>cltc2</i>
10	76,250,000 – 76,450,000	200,000	<i>sept7</i> , <i>anln</i> , <i>prmt7</i> , <i>coq9</i> , <i>polr2c</i> , <i>dgat1</i> , <i>nedd4</i> , <i>adat1</i> , <i>spire2</i>
11	19,225,000 – 19,300,000	75,000	<i>numa1</i>
13	78,325,000 – 78,425,000	100,000	<i>trpc2</i> , <i>rrm1</i> , <i>slc6a7</i> , <i>rb1</i> , <i>lpar6</i>
13	81,025,000 – 81,150,000	125,000	<i>eda2r</i> , <i>ar</i> , <i>ophn1</i>
14	64,700,000 – 65,275,000	575,000	<i>pard6g</i> , <i>bloc1s4</i> , <i>nutf2</i> , <i>adnp2</i> , <i>txn14a</i> , <i>pqlc1</i> , <i>kcng2</i> , <i>ctdp1</i> , <i>adck5</i> , <i>cpsf1</i> , <i>parp10</i> , <i>st3gal1</i> , <i>khdrbs3</i>
15	46,925,000 – 47,200,000	275,000	<i>mdga1</i>
21	24,850,000 – 25,075,000	225,000	<i>trim13</i> , <i>rnaseh2b</i> , <i>nr0b1</i> , <i>il1rap11</i>
24	34,225,000 – 34,525,000	300,000	<i>edil3</i>
25	47,075,000 – 47,225,000	150,000	<i>stxbp5l</i> , <i>gtf2e1</i> , <i>mx1-1</i> , <i>mx1-2</i> , <i>mx2-1</i> , <i>smcp-1a</i> , <i>mx1-3</i>

northern and southern populations of salmon. We identified missense SNPs in all regions of the protein including a SNP in the antiviral specificity domain in exon 13 (Additional file 1: Figure S3). This SNP represents a structurally relevant amino acid substitution, where

arginine seems to be the ancestral variant and cysteine the derived variant dominating in the northern population (Chr 25 position: 47,120,121). Likewise, SNPs in this domain have been associated with specific virus resistance in chicken [49, 50] and pig [51]. SNPs in *mx* genes

Table 2 Missense mutations in selective sweeps. Several missense mutations were discovered in the selective sweeps. The table lists resequencing derived reference alleles frequencies in the northern (N) and southern (S) populations of salmon in Norway. SNPs from Chr 5 and Chr 25 selected for genotyping are shown in bold

Chr	Position	$-\log_{10}P$	Reference allele frequency (S/N)	Ref/alt nucleotide	Ref/alt amino acid	Gene	Description of gene
5^a	22,641,277	29.6	1.00/0.34	A (G)	Phe (Ser)	<i>nkrf</i>	NF- κ B repressing factor.
5	22,691,092	27.5	0.95/0.29	G (T)	Ala (Ser)	<i>sowahc</i>	Unknown function.
5^a	22,708,279	28.8	1.00/0.33	C (T)	Val (Met)	<i>nkap</i>	NF- κ B activating protein.
5^a	22,719,453	19.4	0.79/0.27	T (A)	Val (Glu)	<i>zbtb33</i>	Transcriptional regulator.
10 ^b	76,272,006	18.7	0.94/0.46	T (G)	Asp (Glu)	<i>anln</i>	Actin binding protein.
10	76,272,245	18.4	0.96/0.49	C (A)	Pro (Gln)	<i>anln</i>	Same as above.
10	76,278,414	14.5	0.32/0.76	G (A)	Ala (Thr)	<i>anln</i>	Same as above.
13	78,341,469	24.1	0.97/0.43	G (A)	Gly (Arg)	<i>trpc2</i>	Transient Receptor Cation Channel.
13	78,349,376	20.3	0.98/0.40	T (G)	Glu (Ala)	<i>rrm1</i>	Ribonucleoside-diphosphate reductase.
13 ^b	78,413,379	20.7	0.98/0.48	A (T)	Gln (Leu)	<i>rb1</i>	Regulator of entry into cell division.
14	64,739,123	20.0	0.55/0.01	T (C)	Asn (Ile)	<i>adnp2</i>	Transcription factor.
14 ^b	64,988,859	11.9	0.36/0.79	A (T)	Asn (Lys)	<i>cpsf1</i>	Pre-mRNAs processing.
14 ^b	65,006,543	20.6	0.38/0.94	A (C)	Asp (Glu)	<i>parp10</i>	ADP-Ribosyltransferase.
21 ^b	24,974,056	19.7	0.98/0.46	T (A)	Met (Lys)	<i>rnaseh2b</i>	Non catalytic subunit of RNase H2.
25^a	47,105,598	26.0	0.84/0.03	A (G)	Thr (Ala)	<i>mx1-1</i>	Interferon-induced antiviral.
25	47,108,912	27.2	0.91/0.00	G (A)	Val (Ile)	<i>mx1-1</i>	Same as above.
25	47,111,137	18.0	0.79/0.06	G (A)	Val (Met)	<i>mx1-1</i>	Same as above.
25^a	47,120,121	33.1	0.83/0.06	C (T)	Arg (Cys)	<i>mx1-1</i>	Same as above.
25	47,147,348	27.2	0.22/0.98	G (T)	Pro (His)	<i>mx1-2</i>	Same as above.
25	47,181,001	25.4	0.85/0.00	T (C)	His (Arg)	<i>mx2-1</i>	Same as above.

^aAllele frequencies from genotyping are illustrated in Fig. 4^bAllele frequencies from genotyping are illustrated in Additional file 1: Figure S5

have also been investigated in another fish species, the turbot [52], however, properties related to protection against viruses were not investigated in this study. In rainbow trout (*Oncorhynchus mykiss*) genetic variation in *mx* between strains in exon 3–6, was correlated with susceptibility to infectious hematopoietic necrosis virus (IHNV) [53]. This virus also infects Atlantic salmon and our discovery of a missense mutation in exon 6 suggests that salmon could have adapted to the IHNV (Additional file 1: Figure S3). In addition, different strains of rainbow trout display variable susceptibility to this virus [54]. In this study we cannot elucidate the functional significance of the acquired SNPs in *mx* in Northern Norway, however, further studies will reveal whether any of these changes have been involved in host-virus adaptation [55].

We also investigated whether the selective sweeps on Chr 5 and Chr 25 had paralogous regions in the partially tetraploid salmon genome [56]. *In silico* analysis showed that both sweeps have paralogous regions located on other chromosomes. The Chr 5 sweep has a paralogous region on Chr 9 (Additional file 1: Figure S4, position 51,349,279 to 51,849,279), which did not contain any differentiated SNPs. The synteny is conserved in other species, and the existence of only one copy in zebrafish (*Danio rerio*), combined with the observation that missense mutations on Chr 5 are not present in the paralogous genes on Chr 9, indicate that the mutations arose after the salmonid specific whole genome duplication (WGD). Based upon this observation, it is possible to speculate that the WGD provided paralogous regions where one copy was free to sub- or neo-functionalize, much like the theory for duplicated genes [57] which has been suggested to be important for evolutionary adaptation and innovation in salmon [58], in teleosts [59] and in general [30]. A similar picture is seen for the sweep on Chr 25 where the paralogous region harbors a cluster of three *mx* genes on Chr 12 (position 66,552,602 to 67,052,602), but carries no differentiated SNPs or missense mutations. While the sweeps on Chrs 11, 15, 21 and 24 have no clear paralogous regions, the sweeps on Chr 10, Chr 14 and the two sweeps on Chr 13 also have paralogous regions with very few significantly differentiated SNPs, on Chr 16, 27 and 4, respectively (Fig. 3). Similarly, in our recent discovery of the loci in Chr 25 controlling age at maturity [27] we investigated the two paralogous regions in Chr 21, both of which were without SNPs associated with the trait. Together, these findings indicate that the partially tetraploid stage may be beneficial for adaptation, since one gene copy or gene cluster can keep the original function while the other can adapt to a new situation such as novel disease pressures.

In this study, the initial resequencing was based only upon males. This is because it allowed reusing sequence

data from our previous work [27]. The targeted SNP analysis, used to validate the results from resequencing in a larger independent set of rivers, was conducted using both males and females (Figs. 1 and 4). Genotyping of mixed sex salmon from 19 rivers ($n = 20$ salmon/river) along the Norwegian coast (Fig. 1) for five missense SNPs on Chr 5 and 25 confirmed strong genetic differentiation between salmon populations from the north and south of Norway (Fig. 4). Populations from northern rivers (1–9) displayed allele frequencies in the range 0–0.7, while those from southern rivers (11–19) were close to fixation for one allele at these two loci. Salmon from river 10, Målselv, shows intermediate frequencies, which corresponds well with what has been reported in the literature [31, 32]. These results also confirm allele frequency estimations from the pooled resequencing (Table 2). In addition, we designed Sequenom assays for five other missense SNPs in other regions; one SNP each for sweeps on Chrs 10, 13 and 21, and two SNPs in Chr 14. Genotyping was performed for all 19 rivers (Additional file 1: Figure S5). The allele frequencies showed the same clear difference between the northern and southern populations. For the SNPs on Chr 14 there appears to be an additional genetic shift between the rivers 14, Stjørdalselva and rivers south of this. In addition to the data produced within the present study, resequencing data from a recent publication was downloaded and compared to our results [28]. The downloaded data include three individually sequenced salmon from 4 southern and 3 northern salmon rivers in Norway. These data corroborate our resequencing and genotyping results (Additional file 1: Table S3). Our surveyed SNPs therefore also represent robust and good genetic markers for distinguishing northern and southern populations of Atlantic salmon in Norway. Future studies on an extended set of populations may reveal if these are also robust markers for detecting genetic structuring in other parts of the distribution range of the species.

Atlantic salmon aquaculture involves rearing domesticated fish that originate from commercial breeding programs. Forty wild populations from both the north and south of Norway were sampled when establishing the national breeding programs for salmon [60]. However, analyses of genetic markers demonstrate that there is a dominance of salmon from Southern Norway in the domesticated lines currently in production [61]. Genetic analyses of farmed salmon escapees in Norway have uncovered genetic introgression into native salmon populations in both Northern and Southern Norway, but the biological consequence remains unknown [32, 61, 62]. Consequently the results from the present study, where adaptive genetic divergence between wild salmon from populations located in the north and south of Norway was revealed, it is likely that the potential negative

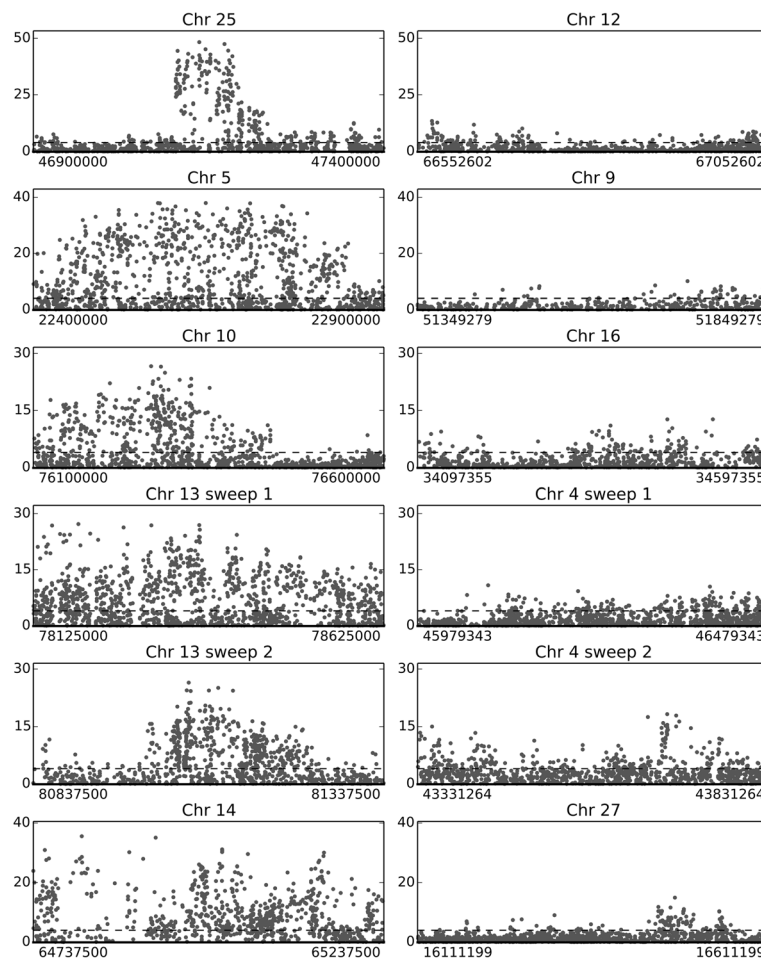


Fig. 3 Comparison of sweeps and paralogous regions. Representation of the selective sweeps having paralogous regions in the salmon genome, displaying a 500 kb overview of the SNPs in the selective sweeps (left side) and in the corresponding paralogous regions (right side). The y-axis shows $-\log_{10}P$ of the SNPs being different in populations between the north and south of Norway, and the x-axis represents the position in the given chromosome. The dashed lines indicate the genome wide significance threshold (0.1 % FDR)

genetic impact of domesticated salmon introgression is greater in populations located in northern regions, since the farmed fish originate mostly from wild Southern Norway populations.

Conclusion

In this study we performed a GWAS by genome resequencing with the aim to screen the Atlantic salmon genome for genetic differentiation between the northern and southern populations in Norway. By investigating eight rivers we uncovered ten particularly striking sweeps including two clusters of immune related genes harboring missense mutations. A feasible interpretation is that different populations of Atlantic salmon have historically been exposed to different selection pressures in the form of pathogens. Some of these adapted alleles could be advantageous for aquaculture production which is currently hampered by a number of diseases, including

virus infections [63]. Future studies should include gene editing of immune genes found in these selective sweeps [64, 65] in combination with viral exposure experiments. Within these experiments, viruses relevant to salmon aquaculture should be the primary focus since finding specific resistance alleles can be of significant value to the industry and could also be used for protecting wild fish against high disease pressures posed by open cage aquaculture [66]. Upon finding the protective alleles, selective breeding on individuals with beneficial haplotypes could lead to increased welfare for aquaculture salmon, decreased disease pressure on wild populations and could also be economically favorable for the industry. On the other hand, further studies should investigate the impact of genetic introgression from fertile aquaculture escapees on the adaptive genetic properties in wild populations. To reduce the risk of this unwanted loss of local adaptation and alteration of fitness-related traits, a

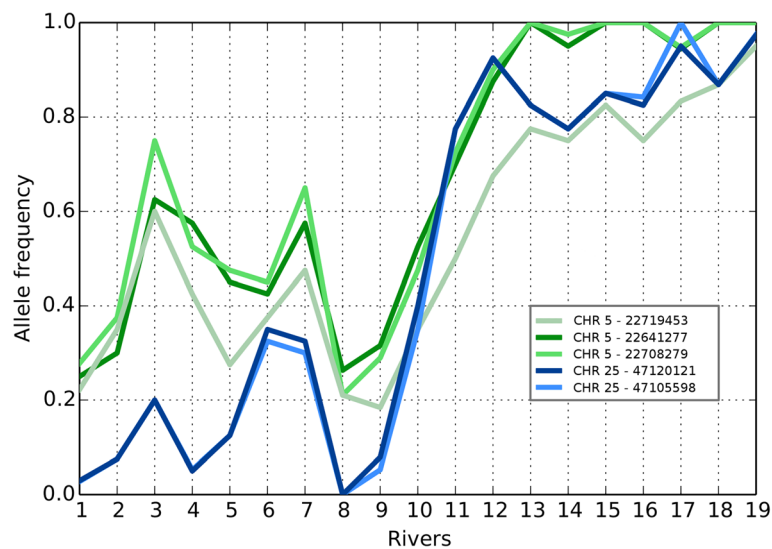


Fig. 4 Allele frequencies of five missense mutations in populations along the Norwegian coast. Three missense mutations identified in selective sweeps on Chr 5 and two on Chr 25 were used in a genotyping assay of Atlantic salmon populations along the Norwegian coast. The graphs show frequencies of the reference alleles. River numbers are explained in Fig. 1. SNPs on Chr 5 included missense mutations in the genes *nkrf*, *nkap* and *zbtb33* (green lines) and missense mutations on Chr 25 were located in the *mx1-1* gene (blue lines)

sustainable solution would be the use of sterile fish in aquaculture, especially in Northern Norway. Future studies should also investigate whether paralogous regions of selective sweeps have undergone positive selection or not, as the latter scenario would suggest an evolutionary mechanism which provides higher adaptive possibilities when a genome is partially tetraploid.

Methods

Samples and sampling

Scales from 30 Atlantic salmon males per river were selected from a sample set of 26,000 samples collected in coastal fisheries in Northern Norway. In the Kolarctic Salmon project (<http://prosjekt.fylkesmannen.no/Kolarcticsalmon>), the multilocus genotypes of all individuals were compared to a genetic baseline consisting of over 180 rivers from Northern Russia and Norway and were assigned to river of origin. Samples that were assigned with high probability to four rivers in Northern Norway; Altaelva, Reisaelva, Lakselv and Tanaelva were generously made available to this study. 30 salmon males from each of four different rivers in Southern Norway, including Årdalselva, Eidselva, Gloppenelva and Suldalslågen were sampled and resequenced in a recent study [27]. In addition to these, we also used male and female salmon DNA from 19 rivers along the Norwegian coast. These included 20 parr individuals from each of the rivers Grense Jakobselv, Neiden, Bergebyelva, Komagelva, Kongsfjordelva, Langfjordelva, Børselva, Staburselva, Repparfjordelva, Målselv, Laukhelle, Alvsvågvasdraget, Årgårdsvassdraget, Stjørdalselva, Jølstra, Lyseelva, Bjerkreimselva, Storelva and Enningdalselva (represented

by numbers in Fig. 1). With the exception of Enningdalselva where the sample was obtained from scales collected by recreational fisheries, these samples were obtained from fins collected by electrofishing of juvenile salmon from multiple locations in the rivers.

DNA extraction and sequencing

DNA from the 19 rivers for genotyping was extracted from scales or fin samples using Qiagen DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's recommendations. From salmon belonging to the four populations in Northern Norway total DNA was extracted from scales using Qiagen DNeasy Blood and Tissue Kit. Equal amounts of DNA from ten individuals were pooled to make three pools per river, totaling 30 individuals from each river. Paired-end libraries were constructed using the Genomic DNA Sample Preparation Kit (Illumina, CA, USA) according to manufacturer's instructions and sequenced on the Illumina HiSeq2000 platform (Illumina, CA, USA) at the Norwegian Sequencing center (<https://www.sequencing.uio.no>, Oslo, Norway) with each pool sequenced in separate lanes.

Sequence mapping and SNP calling

To ensure high quality sequences, sequenced reads were inspected with FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Adapter sequence removal and quality trimming was done with Cutadapt [67], resulting in 1,077,839,448 (SD 40,407,492) paired reads on average per river. Sequenced reads were mapped to the most recent release of the salmon genome

(AGKD000000.4) using Bowtie2 (v.2.1.0) [68] without soft clipping (end-to-end mode). To increase the sensitivity of the mapping, seed length ($-L$ parameter) was set to 18 and the interval between extracted seeds ($-i$ parameter) was set to $S, 1, 1.5$ corresponding to the function $f(L) = 1 + 1.5 \cdot \sqrt{L}$, where L is the read length. Additionally, the maximum number of mismatches per seed ($-N$ parameter) was set to $L, 0, 0.1$, corresponding to the function $f(L) = 0 + 0.1 \cdot L$, where L is the read length, and minimum alignment score ($--score-min$ parameter) was set to $L, -0.6, -0.4$, corresponding to the function $f(L) = -0.6 + -0.4 \cdot L$, where L is the read length. To remove ambiguously mapped reads the mapping quality threshold was set to 20. To obtain higher sequence coverage, the three sequenced pools per river were merged to a single BAM file using SAMtools merge. SNPs were called using SAMtools mpileup [69] and the output was parsed using the PoPoolation2 package (mpileup2sync.jar) [70] with a minimum base quality threshold of 20. For a SNP to be included in the final set of high quality SNPs, minimum coverage of 10 and maximum coverage of 50 (99 % percentile) was required for each river. In addition, the total number of observed minor alleles was required to be at least 8. Recently published whole genome resequencing data from individuals [28] was downloaded and mapped to the reference genome. The data included three salmon from each of the rivers Tanaelva, Repparfjordelva, Altaelva, Namsenelva, Årgårdsvassdraget, Nausta and Jølstra, where the first three represent populations in Northern Norway and the last four represent Southern Norway. Accession numbers for the samples are shown in the caption of Additional file 1: Table S3.

Statistical analysis

Pairwise fixation index (F_{ST}) between all eight sequenced populations was calculated for all high quality SNPs using Hudson's estimator for F_{ST} [34]. F_{ST} values were averaged over all SNPs in each population to generate a distance matrix using F_{ST} as genetic distance. This matrix was converted to a newick tree using NEIGHBOR from the Phylip package [71] and a phylogenetic tree was created with NJplot [72]. To find SNPs with significantly different allele frequencies (0.1 % FDR) between populations from Northern and Southern Norway the Cochran-Mantel-Haenszel test for repeated tests of independence from the PoPoolation2 package (cmh-test.pl) [70] was used. The FDR threshold was determined using the method described in [73]. Allele counts for each river were merged to get the total allele count per SNP in Northern and Southern Norway, corresponding to 120 individuals per geographical region. From this, F_{ST} values between the northern and southern populations were estimated using the F_{ST} calculation from the PoPoolation2 package (fst-sliding.pl) for each SNP, with $-pool-size$

parameter set to 120. Genomic regions with low values of heterozygosity may indicate SNPs under selection. Therefore heterozygosity values were estimated for north and south of Norway, separately, for each SNPs as $2 \cdot (\text{major allele frequency} \cdot \text{minor allele frequency})$. Sliding windows of 50 kb with steps of 25 kb was used to find genomic regions with high F_{ST} values and with low heterozygosity values in either Northern or Southern Norway. This approach is similar to one used to discover genomic regions under selection in other animals [36]. To identify putative selective sweeps it was required that the average F_{ST} value of the window was at least 0.17 (above 99.9 % percentile) and that average heterozygosity of the window in either Northern or Southern Norway was at most 0.15 (below 5 % percentile) (Additional file 1: Figure S1). The thresholds were chosen with focus on capturing the outliers in the F_{ST} and heterozygosity distributions. Putative sweeps were extended to the sides for as long as the neighboring windows had either average F_{ST} of at least 0.17 or heterozygosity of at most 0.15 in either Northern or Southern Norway. If identified sweeps were less than 50 kb apart these were joined to avoid fragmentation of the putative selective sweeps. Genomic windows containing more than 10 % ambiguous bases (Ns) in the reference assembly were discarded to exclude regions with high levels of uncertainty.

SNP annotation

Genes in the sweep regions were obtained from the official genome annotation (NCBI *Salmo salar* Annotation Release 100). Missense mutations in selective sweep regions were identified by manual inspection of the coding sequences. Amino acid sequences of five *mx* genes found in a selective sweep on Chr 25 were aligned to the homologs Mx1 and Mx2 from human and MxD and MxG from Zebrafish using BLASTP (default parameters). Functional domains in the Mx proteins were assigned using domain information for human Mx1 from UniProt. Amino acid sequences from four genes containing missense mutations in a selective sweep on Chr 5 (*nkrf*, *sowahc*, *nkap* and *zbtb33*) were aligned to homologous zebrafish and Northern Pike genes using BLASTP with default parameters. Synteny between genes in the sweep on Chr 5 and other animals was found using the UCSC genome browser (<https://genome.ucsc.edu>) to inspect the syntenic regions of zebrafish, human and mouse. Paralogous regions of the sweeps were identified using TBLASTN (default parameters) with the genes in the sweeps against the salmon genome.

Genotyping

Twenty salmon from 19 rivers along the Norwegian coastline ($n = 380$) were genotyped using ten of the most significant missense mutations on a Sequenom

MassARRAY iPLEX platform (San Diego, CA, USA). Primers and extension primers are listed in Additional file 1: Table S4. The genotyping primers were designed to not target any paralogous genes in the genome.

Additional file

Additional file 1: Tables S1 to S4, Figs S1 to S5. Supplemental Materials. (DOC 2321 kb)

Abbreviations

bp, base pair; Chr, chromosome; FDR, false discovery rate; FST, fixation index; GWAS, genome-wide association study; IHN, infectious hematopoietic necrosis virus; IPN, infectious pancreatic necrosis; kb, kilo bases; SNP, single nucleotide polymorphism; WGD, whole genome duplication

Acknowledgements

Samples were generously supplied by a number of agencies and persons. We wish to thank coastal fishermen in Troms and Finnmark, Norwegian Institute of Nature Research, Statens Naturoppsyn, Rådgivende Biologer AS, Fylkesmannen i Nord-Trøndelag and local fishermen in Enningdalselva for providing samples. We would also like to thank Anne Grethe Sørvik for expert technical assistance.

Funding

This project was financed by the Norwegian research council (NFR) and their HAVBRUK-BIOTEK 2021 program (project number 226221- SALMAT). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Availability of data and materials

Genomic sequences from all sequenced pools used in this study have been made available on SRA with Bioproject number PRJNA305872. A list of high quality SNPs has been deposited at <http://marinseq.imr.no/northsouth2016/>.

Authors' contributions

AW, KAG, VW, CJR and RBE conceived and designed the experiments. FA, EKS and GD conducted laboratory experiments. FA, EKS, TF, CJR, AW and RBE analyzed the data. VW, EN, MO and JPV provided samples for analysis. EKS, FA, KAG, AW and RBE wrote the first draft of the paper. All authors read and approved to the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Consent for publication

Not applicable.

Ethics approval and consent to participate

Scale samples from adult wild salmon in the rivers were collected by recreational anglers. Scale samples of salmon from four of the northern rivers were collected in commercial coastal fisheries for salmon. Thus no permits/licenses regarding the collection of these samples were required. Juvenile samples from other rivers were collected by own efforts or by several cooperating agencies with permits from the County Governor in the respective counties.

Author details

¹Institute of Marine Research, Bergen, Norway. ²Department of Biology, University of Bergen, Bergen, Norway. ³Natural Resources Institute Finland, Helsinki, Finland. ⁴Kevo Subarctic Research Institute, University of Turku, Turku, Finland. ⁵Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden. ⁶Association for Water and Environment of Western Uusimaa, Uusimaa, Finland.

References

- MacCrimmon HR, Gots BL. World distribution of Atlantic salmon, *salmo salar*. *J Fish Res Board Can.* 1979;36:422–57.
- Verspoor E, McCarthy EM, Knox D, Bourke EA, Cross TF. The phylogeography of European Atlantic salmon (*salmo salar* L.) based on RFLP analysis of the ND1/16sRNA region of the mtDNA. *Biol J Linn Soc.* 1999;68:129–46.
- King TL, Kalinowski ST, Schill WB, Spidle AP, Lubinski BA. Population structure of Atlantic salmon (*Salmo salar* L.): a range-wide perspective from microsatellite DNA variation. *Mol Ecol.* 2001;10:807–21.
- Taggart JB, Verspoor E, Galvin PT, Moran P, Ferguson A. A minisatellite DNA marker for discriminating between European and North American Atlantic salmon (*Salmo salar*). *Can J Fish Aquat Sci.* 1995;52:2305–11.
- Ståhl G. Genetic Population Structure of Atlantic Salmon, Population genetics & Fishery Management. Seattle: University of Washington Press; 1987.
- Tonteri A, Titov S, Veselov A, Zubchenko A, Koskinen MT, Lesbarreres D, Kaluzhin S, Bakhmet I, Lumme J, Primmer CR. Phylogeography of anadromous and non-anadromous Atlantic salmon (*Salmo salar*) from northern Europe. *Ann Zoologici Fennici.* 2005;42:1–22.
- Vincent B, Dionne M, Kent MP, Lien S, Bernatchez L. Landscape genomics in Atlantic salmon (*salmo salar*): searching for gene–environment interactions driving local adaptation. *Evolution.* 2013;67:3469–87.
- Koljonen M-L, Tähtinen J, Saisä M, Koskiniemi J. Maintenance of genetic diversity of Atlantic salmon (*Salmo salar*) by captive breeding programmes and the geographic distribution of microsatellite variation. *Aquaculture.* 2002;212:69–92.
- Ayllon F, Martinez JL, Garcia-Vazquez E. Loss of regional population structure in Atlantic salmon, *Salmo salar* L., following stocking. *ICES J Mar Sci.* 2006;63:1269–73.
- Olafsson K, Pampoulie C, Hjørleifsdóttir S, Gudjonsson S, Hreggvidsson GO. Present-Day genetic structure of Atlantic salmon (*salmo salar*) in Icelandic rivers and Ice-Cap retreat models. *PLoS One.* 2014;9(2):e86809.
- Primmer CR, Veselov AJ, Zubchenko A, Poututkin A, Bakhmet I, Koskinen MT. Isolation by distance within a river system: genetic population structuring of Atlantic salmon, *Salmo salar*, in tributaries of the Varzuga River in northwest Russia. *Mol Ecol.* 2006;15:653–66.
- Vähä J-P, Erkinaro J, Niemelä E, Primmer CR. Life-history and habitat features influence the within-river genetic structure of Atlantic salmon. *Mol Ecol.* 2007;16:2638–54.
- Dillane E, McGinnity P, Coughlan JP, Cross MC, de Eyto E, Kenchington E, Prodohl P, Cross TF. Demographics and landscape features determine intrariver population structure in Atlantic salmon (*Salmo salar* L.): the case of the River Moy in Ireland. *Mol Ecol.* 2008;17:4786–800.
- Taylor EB. A review of local adaptation in Salmonidac, with particular reference to Pacific and Atlantic salmon. *Aquaculture.* 1991;98:185–207.
- Ryman N, Utter F. Population genetics and fishery management. Seattle: University of Washington Press; 1987.
- de Leániz CG, Fleming IA, Einum S, Verspoor E, Consuegra S, Jordan WC, Aubin-Horth N, Lajus DL, Villanueva B, Ferguson A, et al.: Local Adaptation. In *The Atlantic Salmon: Genetics, Conservation and Management*. Blackwell Publishing Ltd; 2007: 195–235
- Thorstad EB, Whoriskey F, Rikardsen AH, Aarestrup K: Aquatic Nomads: The Life and Migrations of the Atlantic Salmon. In *Atlantic Salmon Ecology*. Wiley-Blackwell; 2010: 1–32
- Verspoor E, Jordan WC. Genetic variation at the Me-2 locus in the Atlantic salmon within and between rivers: evidence for its selective maintenance. *J Fish Biol.* 1989;35:205–13.
- Tonteri A, Vasemägi A, Lumme J, Primmer CR. Beyond MHC: signals of elevated selection pressure on Atlantic salmon (*Salmo salar*) immune-relevant loci. *Mol Ecol.* 2010;19:1273–82.
- Dionne M, Miller KM, Dodson JJ, Caron F, Bernatchez L. Clinal variation in MHC diversity with temperature: evidence for the role of host-pathogen interaction on local adaptation in Atlantic salmon. *Evolution.* 2007;61:2154–64.
- Dionne M, Miller KM, Dodson JJ, Bernatchez L. MHC standing genetic variation and pathogen resistance in wild Atlantic salmon. *Philos Trans R Soc Lon B.* 2009;364:1555–65.
- Dempsey PW, Vaidya SA, Cheng G. The Art of War: innate and adaptive immune responses. *Cell Mol Life Sci.* 2003;60:2604–21.
- Burge CA, Mark Eakin C, Friedman CS, Froelich B, Hershberger PK, Hofmann EE, Petes LE, Prager KC, Weil E, Willis BL, et al. Climate change influences on marine infectious diseases: implications for management and society. *Ann Rev Mar Sci.* 2014;6:249–77.

Received: 27 February 2016 Accepted: 30 June 2016

Published online: 11 August 2016

24. Ficke AD, Myrick CA, Hansen LJ. Potential impacts of global climate change on freshwater fisheries. *Rev Fish Biol Fisheries*. 2007;17:581–613.
25. Davidson WS, Koop BF, Jones SJ, Iturra P, Vidal R, Maass A, Jonassen I, Lien S, Omholt SW. Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biol*. 2010;11:403.
26. Moen T, Torgersen J, Santi N, Davidson WS, Baranski M, Ødegård J, Kjølglum S, Velle B, Kent M, Lubieniecki KP, et al. Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. *Genetics*. 2015;200:1313–26.
27. Ayllon F, Kjærner-Semb E, Furmanek T, Wennevik V, Solberg M, Dahle G, Taranger GL, Glover KA, Almén MS, Rubin CJ, et al. The *vgl3* locus controls age at maturity in wild and domesticated Atlantic salmon (*salmo salar* L.) males. *PLoS Genet*. 2015;11(11):e1005628.
28. Barson NJ, Aykanat T, Hindar K, Baranski M, Bolstad GH, Fiske P, Jacq C, Jensen AJ, Johnston SE, Karlsson S, et al. Sex-dependent dominance at a single locus maintains variation in age at maturity in salmon. *Nature*. 2015.
29. Near TJ, Eytan RI, Dornburg A, Kuhn KL, Moore JA, Davis MP, Wainwright PC, Friedman M, Smith WL. Resolution of ray-finned fish phylogeny and timing of diversification. *Proc Natl Acad Sci U S A*. 2012;109:13698–703.
30. Qian W, Zhang J. Genomic evidence for adaptation by gene duplication. *Genome Res*. 2014;24:1356–62.
31. Bourret V, Kent MP, Primmer CR, Vasemagi A, Karlsson S, Hindar K, McGinnity P, Verspoor E, Bernatchez L, Lien S. SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range of Atlantic salmon (*Salmo salar*). *Mol Ecol*. 2013;22:532–51.
32. Glover KA, Quintela M, Wennevik V, Besnier F, Sorvik AG, Skaala O. Three decades of farmed escapees in the wild: a spatio-temporal analysis of Atlantic salmon population genetic structure throughout Norway. *PLoS One*. 2012;7:e43129.
33. Ozerov M, Vasemagi A, Wennevik V, Diaz-Fernandez R, Kent M, Gilbey J, Prusov S, Niemela E, Vaha JP. Finding markers that make a difference: DNA pooling and SNP-arrays identify population informative markers for genetic stock identification. *PLoS One*. 2013;8:e82434.
34. Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting FST: the impact of rare variants. *Genome Res*. 2013;23:1514–21.
35. Smith JM, Haigh J. The hitch-hiking effect of a favourable gene. *Genet Res*. 1974;23:23–35.
36. Carneiro M, Rubin CJ, Di Palma F, Albert FW, Alfoldi J, Barrio AM, Pielberg G, Rafati N, Sayyab S, Turner-Maier J, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. *Science*. 2014;345:1074–9.
37. Omura M, Mombaerts P. Trpc2-expressing sensory neurons in the main olfactory epithelium of the mouse. *Cell Rep*. 2014;8:583–95.
38. Kaufmann M, Feijs KL, Luscher B. Function and regulation of the mono-ADP-ribosyltransferase ARTD10. *Curr Top Microbiol Immunol*. 2015;384:167–88.
39. Kind B, Muster B, Staroske W, Herce HD, Sachse R, Rapp A, Schmidt F, Koss S, Cardoso MC, Lee-Kirsch MA. Altered spatio-temporal dynamics of RNase H2 complex assembly at replication and repair sites in Aicardi-Goutieres syndrome. *Hum Mol Genet*. 2014;23:5950–60.
40. Hodgson A, Wan F. Interference with nuclear factor kappaB signaling pathway by pathogen-encoded proteases: global and selective inhibition. *Mol Microbiol*. 2015.
41. Thapa P, Das J, McWilliams D, Shapiro M, Sundsbak R, Nelson-Holte M, Tangen S, Anderson J, Desiderio S, Hiebert S, et al. The transcriptional repressor NKAP is required for the development of iNKT cells. *Nat Commun*. 2013;4:1582.
42. Burgute BD, Peche VS, Steckelberg AL, Glockner G, Gassen B, Gehring NH, Noegel AA. NKAP is a novel RS-related protein that interacts with RNA and RNA binding proteins. *Nucleic Acids Res*. 2014;42:3177–93.
43. Cofre C, Gonzalez R, Moya J, Vidal R. Phenotype gene expression differences between resistant and susceptible salmon families to IPNV. *Fish Physiol Biochem*. 2014;40:887–96.
44. Mitchell PS, Emerman M, Malik HS. An evolutionary perspective on the broad antiviral specificity of MxA. *Curr Opin Microbiol*. 2013;16:493–9.
45. Lee SH, Vidal SM. Functional diversity of Mx proteins: variations on a theme of host resistance to infection. *Genome Res*. 2002;12:527–30.
46. Goujon C, Moncorge O, Bauby H, Doyle T, Ward CC, Schaller T, Hue S, Barclay WS, Schulz R, Malim MH. Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature*. 2013;502:559–62.
47. Sironi M, Biasin M, Cagliari R, Gnudi F, Saulle I, Ibba S, Filippi G, Yahyaei S, Tresoldi C, Riva S, et al. Evolutionary analysis identifies an MX2 haplotype associated with natural resistance to HIV-1 infection. *Mol Biol Evol*. 2014;31:2402–14.
48. Mitchell PS, Patzina C, Emerman M, Haller O, Malik HS, Kochs G. Evolution-guided identification of antiviral specificity determinants in the broadly acting interferon-induced innate immunity factor MxA. *Cell Host Microbe*. 2012;12:598–604.
49. Sasaki K, Yoneda A, Ninomiya A, Kawahara M, Watanabe T. Both antiviral activity and intracellular localization of chicken Mx protein depend on a polymorphism at amino acid position 631. *Biochem Biophys Res Commun*. 2013;430:161–6.
50. Ko JH, Takada A, Mitsuhashi T, Agui T, Watanabe T. Native antiviral specificity of chicken Mx protein depends on amino acid variation at position 631. *Anim Genet*. 2004;35:119–22.
51. Nakajima E, Morozumi T, Tsukamoto K, Watanabe T, Plastow G, Mitsuhashi T. A naturally occurring variant of porcine Mx1 associated with increased susceptibility to influenza virus in vitro. *Biochem Genet*. 2007;45:11–24.
52. Abollo E, Ordas C, Dios S, Figueras A, Novoa B. Molecular characterisation of a turbot Mx cDNA. *Fish Shellfish Immunol*. 2005;19:185–90.
53. Trobridge GD, LaPatra SE, Kim CH, Leong JC. Mx mRNA expression and RFLP analysis of rainbow trout *Oncorhynchus mykiss* genetic crosses selected for susceptibility or resistance to IHN. *Dis Aquat Organ*. 2000;40:1–7.
54. Purcell MK, Lapatra SE, Woodson JC, Kurath G, Winton JR. Early viral replication and induced or constitutive immunity in rainbow trout families with differential resistance to Infectious hematopoietic necrosis virus (IHN). *Fish Shellfish Immunol*. 2010;28:98–105.
55. Daugherty MD, Malik HS. Rules of engagement: molecular insights from host-virus arms races. *Annu Rev Genet*. 2012;46:677–700.
56. Gidskehaug L, Kent M, Hayes BJ, Lien S. Genotype calling and mapping of multisite variants using an Atlantic salmon iSelect SNP array. *Bioinformatics*. 2011;27:303–10.
57. Ohno S, Wolf U, Atkin NB. Evolution from fish to mammals by gene duplication. *Hereditas*. 1968;59:169–87.
58. Warren IA, Ciborowski KL, Casadei E, Hazlerigg DG, Martin S, Jordan WC, Sumner S. Extensive local gene duplication and functional divergence among paralogs in Atlantic salmon. *Genome Biol Evol*. 2014;6:1790–805.
59. Glasauer SM, Neuhauss SC. Whole-genome duplication in teleost fishes and its evolutionary consequences. *Mol Genet Genomics*. 2014;289:1045–60.
60. Gjedrem T, Gjoen HM, Gjerde B. Genetic-origin of Norwegian farmed Atlantic salmon. *Aquaculture*. 1991;98:41–50.
61. Glover KA, Pertoldi C, Besnier F, Wennevik V, Kent M, Skaala Ø. Atlantic salmon populations invaded by farmed escapees: quantifying genetic introgression with a Bayesian approach and SNPs. *BMC Genet*. 2013;14:4.
62. Skaala O, Wennevik V, Glover KA. Evidence of temporal genetic change in wild Atlantic salmon, *Salmo salar* L., populations affected by farm escapees. *Ices J Mar Sci*. 2006;63:1224–33.
63. Collet B. Innate immune responses of salmonid fish to viral infections. *Dev Comp Immunol*. 2014;43:160–73.
64. Edvardsen RB, Leininger S, Kleppe L, Skaftnesmo KO, Wargelius A. Targeted mutagenesis in Atlantic salmon (*Salmo salar* L.) using the CRISPR/Cas9 system induces complete knockout individuals in the F0 generation. *PLoS One*. 2014;9:e108622.
65. Wargelius A, Leininger S, Skaftnesmo KO, Kleppe L, Andersson E, Taranger GL, Schulz RW, Edvardsen RB. Dnd knockout ablates germ cells and demonstrates germ cell independent sex differentiation in Atlantic salmon. *Sci Rep*. 2016;6:21284.
66. Taranger GL, Karlson O, Bannister RJ, Glover KA, Husa V, Karlsbakk E, Kvamme BO, Boxaspen KK, Bjorn PA, Finstad B, et al. Risk assessment of the environmental impact of Norwegian Atlantic salmon farming. *Ices J Mar Sci*. 2015;72:997–1021.
67. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17:10–2.
68. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods*. 2012;9:357–9.
69. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. Genome project data processing 5: the sequence alignment/Map format and SAMtools. *Bioinformatics*. 2009;25:2078–9.
70. Kofler R, Pandey RV, Schlotterer C. PoPoolation2: identifying differentiation between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics*. 2011;27:3435–6.

71. Fink WL. Microcomputers and phylogenetic analysis. *Science*. 1986;234:1135–9.
72. Perriere G, Gouy M. WWW-query: an on-line retrieval system for biological sequence banks. *Biochimie*. 1996;78:364–9.
73. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. *J R Stat Soc B Met*. 1995;57:289–300.

Submit your next manuscript to BioMed Central
and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

