Norwegian University of Life Sciences
Faculty of Biosciences (Biovit)
Department of Animal and Aquacultural Sciences

Philosophiae Doctor (PhD)
Thesis 2025:77

Genomic Basis of Pathogen Response and Immune Gene Evolution in Atlantic Salmon

Genomisk grunnlag for patogenrespons og evolusjon av immun-gener i atlanterhavslaks

Domniki Manousi

Genomic Basis of Pathogen Response and Immune Gene Evolution in Atlantic Salmon

Genomisk grunnlag for patogenrespons og evolusjon av immun-gener i atlanterhavslaks

Philosophiae Doctor (PhD) Thesis

Domniki Manousi

Norwegian University of Life Sciences
Faculty of Biosciences (Biovit)
Department of Animal and Aquacultural Sciences

Ås 2025

Thesis number 2025:77
ISSN 1894-6402
ISBN 978-82-575-2285-8

Supervisors and Evaluation Committee

PhD Supervision team

Dr. Marie Saitou
Faculty of Biosciences
Department of Animal
and Aquacultural sciences, CIGENE,
Norwegian University of Life Sciences,
P.O. Box 5003 NMBU,
1432 Ås, Norway.
marie.saitou@nmbu.no

Prof. Simen Sandve
Faculty of Biosciences
Department of Animal
and Aquacultural sciences, CIGENE,
Norwegian University of Life Sciences,
P.O. Box 5003 NMBU,
1432 Ås, Norway.
simen.sandve@nmbu.no

Prof. Sigbjørn Lien
Faculty of Biosciences
Department of Animal
and Aquacultural sciences, CIGENE,
Norwegian University of Life Sciences,
P.O. Box 5003 NMBU,
1432 Ås, Norway.
sigbjorn.lien@nmbu.no

PhD Evaluation Committee

Prof. Andrea Doeschl-Wilson
The Roslin Institute,
Easter Bush Campus,
University of Edinburgh,
Midlothian,
EH25 9RG, United Kingdom.
andrea.wilson@roslin.ed.ac.uk

Senior Researcher, PhD.
Victoria Louise Pritchard
Institute for Biodiversity and
Freshwater Conservation,
UHI, Inverness,
IV1 1SA, United Kingdom.
Victoria-
Louise.Pritchard.ic@uhi.ac.uk

Dr. Gareth Frank Difford
Faculty of Biosciences
Department of Animal
and Aquacultural sciences,
Norwegian University of Life
Sciences,
P.O. Box 5003 NMBU,
1432 Ås, Norway.
gareth.difford@nmbu.no

Acknowledgements

First, I would like to thank my main supervisor, Marie Saitou, for giving me the opportunity to join this PhD project. Throughout this journey, you have consistently provided support, guidance, and out-of-the-box ideas that broadened both my skills and perspective. You have treated me more as a colleague than a student, allowing me to develop autonomy and confidence in my work.

I would also like to thank my co-supervisors, Sigbjørn Lien and Simen Sandve, for their invaluable support and guidance throughout my PhD. Every discussion and spirited scientific "argument" with you has been a valuable learning experience. Your passion and dedication to research is truly inspiring and I am grateful to have had the opportunity to work with you.

Furthermore, I would like to thank all the people who I had the pleasure of collaborating with during my PhD. Special thanks to Tim Knutsen, Fabian Grammes and Thomas Moen from AquaGen, who — beyond providing key resources for my analyses — offered invaluable insights into Atlantic salmon aquaculture. Big thanks also to Valia Oikonomou and Cathrine Brekke for brainstorming with me.

In addition, a heartfelt thanks to the folks at CIGENE. Thank you not only for the impeccable professional environment you provided but also for establishing a community where I could present myself beyond my scientific work. Thanks to my MSLab group members (Célian, Jun, Pauline, Erik, Akira, Maëlys) for contributing to a vibrant and diverse scientific group, and especially to my office mates Mari, Prabin, and Dorota — it was a pleasure to share knowledge, frustrations, and the ups and downs of the PhD journey together.

I would also like to express my deepest gratitude to my family for their unwavering love and support. Your understanding and encouragement mean the world to me. To my chosen family, two and four-legged members included, thank you for always supporting me and making me feel at home when home was far away.

Finally, I extend my gratitude to the Atlantic salmon. The challenges presented by its duplicated genome have continuously pushed me to think deeper and broader. I am grateful to have contributed, in a small way, to the understanding of this remarkable species.

Here's to *Salmo* — and here's to Ohno.

Table of Contents

1 List of papers

Paper 1

Domniki Manousi, Dorota M. Jaskula, Fabian Grammes, Tim M. Knutsen Shahmir Naseer, Samuel AM Martin, Thomas Moen, Marie Saitou, Sigbjørn Lien (2025). **Dissecting the genetic basis of response to salmonid alphavirus in Atlantic salmon.** *Accepted for publication in BMC genomics.* https://doi.org/10.21203/rs.3.rs-6005887/v1

Paper 2

Domniki Manousi, Shahmir Naseer, Samuel AM Martin, Simen R. Sandve, Marie Saitou. **Evolution of gig immune genes in teleosts shaped by gene gain and loss: regulatory and functional insights from Atlantic salmon.** *Manuscript.*

Paper 3

Pauline Buso, Célian Diblasi, Domniki Manousi, Jun Soung Kwak, Arturo Vera-Ponce de Leon, Kristina Stenløkk, Nicola Barson, Marie Saitou (2025). **Parallel Selection in Domesticated Atlantic Salmon from Divergent Founders Including Whole-Genome Duplication-derived Homeologous Regions.** *Genome Biology and Evolution*, Volume 17, Issue 4, April 2025, evaf063, https://doi.org/10.1093/gbe/evaf063

2 Abstract

Resistance against environmental and pathogenic pressures is imperative for the welfare and sustainability of the aquaculture industry. As a result, enhancing and optimizing immune responses of farmed populations is a priority for aquaculture. Atlantic salmon (Salmo salar) is an important fish for the aquaculture industry due to its contribution to the worldwide food supply chain. At the same time the species holds additional evolutionary significance due to its unique life history; Experience of a relatively recent, salmonid-specific whole-genome duplication (WGD) event that was followed by incomplete rediploidization. Differences in rediploidization timing increased the complexity of the Atlantic salmon genome, particularly in genomic regions experiencing delayed rediploidization. This thesis focused on understanding how the intricate genomic structure, evolutionary history and selection pressures influence expression of key aquaculture traits such as disease response in farmed Atlantic salmon.

Paper 1 investigates the genetic architecture of response against pancreas disease (PD), a viral pathogen with serious implications for aquaculture populations. In this study we used the latest advances in Atlantic salmon genomics, transcriptomics and epigenetics to detect genetic variation with functional impact on disease resistance. We used genomics and transcriptomics data from virally infected individuals to identify likely gene candidates encompassing the disease associated genomic region and detect genetic variants with putative impact on gene expression. Our analyses confirmed a major effect locus on chromosome Ssa03, harboring six genes, including three tandemly duplicated copies of the *gig1-like* gene. We applied long-read based genomics and transcriptomics methods to manually annotate the three gene copies and detect complex genomic variation overlapping the disease associated locus. Finally, by integrating epigenetics information from an experiment of viral stimulation, we identified simple as well as larger, hyper-polymorphic, variants with putative functional impact on PD response.

In paper 2 we explored the evolutionary history of the *gig1* and *gig2* antiviral gene families in farmed Atlantic salmon. Comparison of *gig* gene gain/loss trees revealed large gene expansions in gig families, particularly in species experiencing multiple

rounds of WGD. Increased gig gene abundance in Atlantic salmon and distribution of *gig* genes across homeologous genomic regions indicated a potential link between genome duplication and evolution of antiviral responses. We investigated the functional divergence of *gig* paralogs by performing functional domain prediction and repeat sequence detection analyses. Increased abundance of repeat sequence overlap with predicted functional domains indicated genomic rearrangements as a potential driver of divergence in gig genes. Finally, we used multiple transcriptomics and epigenetics information to investigate the functional divergence of paralog genes across different cells and tissues. Our findings revealed an overall generalized response of *gig* families, linked to transcriptional activity of type I Interferon (IFN-I)-related regulatory elements. In addition, differences in expression and regulation patterns between tandemly duplicated genes suggested potential exchange of regulatory machinery between segmentally duplicated gene copies.

Lastly, paper 3 investigated the impact of WGD in the domestication of two geographically isolated Atlantic salmon lineages (West Norwegian and North American). We used multiple population-based approaches including allele and site frequency as well as haplotype differentiation and explored selection signatures between farmed and wild populations of the two lineages. Comparison of selection signatures between the two lineages of farmed Atlantic salmon revealed parallel selection of SNPs overlapping gene members of the Major histocompatibility complex (MHC) locus located on chromosome Ssa14. In addition, inspection of selected haplotypes within the two aquaculture populations identified lineage-specific selection signatures on WGD-duplicated gene, specifically on the homeologous regions of chromosomes Ssa13 and Ssa04 and chromosomes Ssa14 and Ssa27. Haplotype analysis of the MHC region under common selection on Ssa14 revealed long-term balancing selection as well as independent sweeps in the two isolated populations, suggesting that domestication and selection based on common environmental pressures has likely shaped adaptive immunity in farmed Atlantic salmon. In addition, gene expression analyses of the duplicated genes under selection showed common as well as different functional features, indicating a possible role of WGD in facilitation of immunogenetic diversity and reduction of pleiotropy through sub-functionalization of the duplicated genes.

3 Norsk sammendrag

Motstand mot miljømessige og patogene påkjenninger er avgjørende for velferd og bærekraft i havbruksnæringen. Dermed er forbedring av immunresponsen hos oppdrettspopulasjoner en prioritet for havbruksnæringen. Atlanterhavslaks (Salmo salar) er en viktig art for havbruksnæringen på grunn av dens bidrag til den globale matforsyningskjeden. I tillegg har arten ytterligere evolusjonær interesse da laksefisk familien har gjennomgått en relativt nylig helgenomduplisering (HGD), etterfulgt av ufullstendig rediploidisering. Forskjeller i rediploidiseringstidspunkt har økt kompleksiteten i atlanterhavslaksens genom, spesielt innen genomiske regioner som opplevde forsinket rediploidisering. Denne avhandlingen fokuserer på å forstå hvordan genomstruktur, evolusjonær historie og seleksjonspress påvirker viktige egenskaper for havbruksnæringen som for eksempel sykdomsrespons hos oppdrettslaks.

Artikkel 1 undersøker den genetiske arkitekturen til respons mot pankreassykdom (PD), et viruspatogen med alvorlige konsekvenser for havbrukspopulasjoner. I denne studien brukte vi de nyeste metodene innen genomikk, transkriptomikk og epigenetikk på atlanterhavslaks for å finne genetiske varianter med funksjonell effekt på sykdomsresistens. Vi brukte genomikk- og transkriptomikkdata fra viralt infiserte individer for å identifisere sannsynlige kandidatgener i den sykdomsassosierte genomregionen og for å oppdage genetiske varianter med antatt innvirkning på genuttrykk. Våre analyser bekreftet et lokus med stor effekt på sykdomsresistens på kromosom Ssa03, dette lociet har seks gener, inkludert tre tandemdupliserte kopier av gig1-liknende gen. Vi brukte genomikk- og transkriptomikkmetoder basert på long-read sekvensering for å manuelt annotere de tre genkopiene og identifisere kompleks genomisk variasjon som overlappet det sykdoms-assosierte lokuset. Avslutningsvis, ved å integrere epigenetisk informasjon fra et eksperiment med viral stimulering, identifiserte vi både enkle og større, hyper-polymorfe varianter med antatt funksjonell innvirkning på PD-respons.

Artikkel 2 utforsket den evolusjonære historien til gig1- og gig2-antivirale genfamilier i oppdrettslaks. Sammenlikning av gen gevinst/tap-trær på tvers av

flere arter avslørte store genutvidelser i gig-familier, spesielt hos arter som har opplevd ytterligere runder med WGD. Økt mengde gig-gener i atlanterhavslaks og fordelingen av disse genene over homeologe genomregioner indikerte en potensiell kobling mellom genomduplisering og utviklingen av antivirale responser. Vi undersøkte funksjonell divergens av gig-paraloger ved å utføre prediksjon av funksjonelle domener og analyser for påvisning av repeterte motiver. Økt forekomst av repeterte sekvenser og overlapp mellom repeterte motiver og predikerte funksjonelle domener indikerte genomiske omorganiseringer som en potensiell drivkraft for divergens i gig-gener. Videre brukte vi flere transkriptomikk datasett samt epigenetisk informasjon for å undersøke funksjonell divergens av paraloge gener på tvers av ulike celler og vev. Våre funn avslørte en generell respons i gig-familiene, koblet til transkripsjonell aktivitet av type I Interferon (IFN-I)-relaterte regulatoriske elementer. I tillegg antydet forskjeller i uttrykk og reguleringsmønstre mellom tandemdupliserte gener en potensiell utveksling av regulatorisk maskineri mellom segmentalt dupliserte genkopier.

Til slutt utforsket artikkel 3 påvirkning av HGD på domestiseringen av to geografisk isolerte atlanterhavslaks-linjer (vestnorsk og nordamerikansk). Vi brukte flere populasjonsbaserte metoder inkludert allel- og stedfrekvens samt haplotype-differensiering og undersøkte seleksjonssignaturer mellom oppdrettspopulasjoner og ville populasjoner av de to linjene. Sammenlikning av seleksjonssignaturer mellom de to linjene med oppdrettslaks avslørte parallell seleksjon av SNP-er som overlapper genmedlemmer av Major histocompatibility complex (MHC)-lokuset lokalisert på kromosom Ssa14. I tillegg identifiserte inspeksjon av utvalgte haplotyper innenfor de to havbrukspopulasjonene linjespesifikke seleksjonssignaturer dupliserte gener, spesielt på de homeologe regionene av kromosomene Ssa13 og Ssa04 og kromosomene Ssa14 og Ssa27. Haplotypeanalyse av MHC-regionen under felles seleksjon på Ssa14 avslørte langtidsbalanserende seleksjon samt uavhengige seleksjonssveip i de to isolerte populasjonene, som antyder at domestisering og seleksjon basert på felles miljøpress sannsynligvis har formet

adaptiv immunitet i oppdrettet atlanterhavslaks. I tillegg viste genuttrykksanalyser av de dupliserte genene under seleksjon både felles og individuelle funksjonelle egenskaper, som indikerer en mulig rolle for WGD i å legge til rette for immunogenetisk mangfold og reduksjon av pleiotropi gjennom sub-funksjonalisering av de dupliserte genene.

4 Synopsis

4.1 Introduction

4.1.1 The Atlantic salmon

Atlantic salmon (*Salmo salar*) is a member of the Salmonid family of ray-finned fishes (*Actinopterygii*). The salmonid family can be distinguished into three monophyletic subfamilies including the Salmoninae (*Salmo, Oncorhynchus, Salvelinus, Brachymystax, Parahucho and Hucho spp*), Coregoninae (*Coregonus, Prosopium and Stenodus spp*) and Thymallinae (*Thymallus spp.*) families, shown in **Figure 1** (Lecaudey et al., 2018). Across salmonids, the Salmoninae subfamily exhibits interesting life-history and behavioral traits; species such as trouts, salmons and charrs (*Oncorhynchus, Salmo, and Salvelinus spp*, respectively) are characterized by anadromy, the ability to migrate from the natal river towards the sea in order to mature and, once matured, return back to freshwater to spawn (Dowall, 2001). Atlantic salmon in particular, spends the first developmental stages in the natal river (1-4 years). After reaching a critical size threshold, immature fish (*parr*) undergo complex physiological, morphological and biochemical remodelling in order to prepare for seawater migration, thus transforming into *smolt* (West et al., 2021). This process of parr-smolt transformation is collectively described by the term "smoltification".

Naturally, Atlantic salmon populations are distributed across the Northern hemisphere (Rikardsen et al., 2021) and can be distinguished into two main lineages; the European and the North American lineage. In addition, further lineages can be identified within the European group (Bourret et al., 2013; Rougemont & Bernatchez, 2018). Divergence of the North American lineage from the ancestral European population dates approximately 1.56-1.76 million years ago (mya), with secondary contacts between the two lineages occurring ~10,000–15,000 years ago, towards the end of the last glacial maxima (Rougemont & Bernatchez, 2018). Geographical isolation of Atlantic salmon lineages facilitated the accumulation of genetic incompatibility, including differences in chromosomal organization between North American and European populations (Hartley, 1987; Lien et al., 2016). As a result, the two lineages are currently considered partially isolated, while substantial

gene flow can be observed between and within these lineages and sub-populations (Rougemont & Bernatchez, 2018).

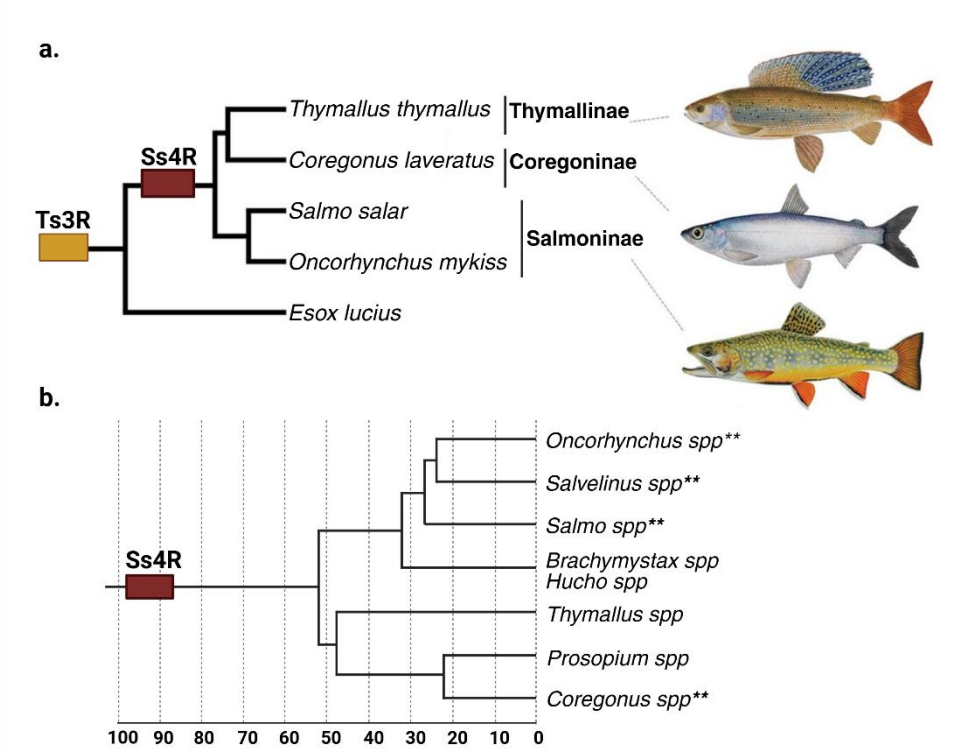


Figure 1. Phylogeny of the salmonid family. *(a.) Phylogenetic tree of salmonids including the salmonid ancestor Esox lucius. Yellow and red rectangles indicate the teleost-specific (Ts3R) and salmonid-specific (Ss4R) WGD events, respectively. Figure adapted from* (Macqueen & Johnston, 2014)*. (b.) Phylogenetic tree of the Salmonidae family. Phylogeny drawn after* (Macqueen & Johnston, 2014)*. Scale shows the chronological divergence of species in million years (mya). Asterisks denote anadromy, as retrieved from (Alexandrou 2013). Similarly to (a.), red rectangle indicates the Ss4R duplication event. Image created with BioRender.*

Despite the distribution of natural/wild populations across the northern hemisphere, commercial farming of Atlantic salmon is nowadays encountered at a global level (Ceballos-Concha et al., 2025; Kim et al., 2024; Leith et al., 2014). To date, almost 95% of farmed salmon is produced in Europe, United kingdom, Chile and North America (R. D. Houston & Macqueen, 2019; A. Iversen et al., 2020), whereas annual salmon production accounts for more than one third of global finfish production (FAO 2024). Due to the socio-economic importance of farmed Atlantic salmon, great research efforts have been put towards the improvement and

optimization of reared stocks, with particular focus on economically important traits such as growth, maturation and disease resistance (R. D. Houston & Macqueen, 2019).

Disease resistance constitutes a major challenge for Atlantic salmon aquaculture. The high stocking density of fish within rearing pens facilitates pathogen transmission, rendering farmed populations susceptible to disease. In turn, this results in high mortality rates and consequently large economic losses for production while at the same time the welfare of farmed populations is severely compromised (Lafferty et al., 2015). Besides harming aquaculture, pathogens accumulating within farming enclosures pose additional threat to wild populations; Farm escapees and wild fish that bypass high infestation farming sites can potentially come in contact with pathogens and transmit infections towards wild stocks (Mordecai et al., 2021).

To limit losses and improve the welfare and biosecurity of farmed populations, several disease prevention measures are currently implemented, including prophylactic vaccinations and antibiotic treatments (Mondal & Thomas, 2022). In addition to these treatments, selection and breeding of individuals whose genetic profile shows improved immune responses are broadly applied in aquaculture (Ross D. Houston et al., 2020).

4.1.2 The Atlantic salmon genome

The architecture of the salmonid genome, shaped by a recent whole-genome duplication (WGD) event, adds another layer of complexity into understanding the genomic basis of aquaculture traits in Atlantic salmon. This complexity is mainly attributed to salmon experiencing a salmonid-specific (Ss4R) WGD event approximately 85-106 mya (Gundappa et al., 2022; Lien et al., 2016; Macqueen & Johnston, 2014) (**Figure 2**). Variability in rediploidization timing following the Ss4R event resulted in retention of a large portion of duplicated genome in Atlantic salmon; approximately 73% of the Atlantic salmon genome experienced early rediploidization (~100 mya, referred to as ancestral ohnolog resolution, AORe) while 27% of genomic regions underwent rediploidization during the last 50 mya (referred to as lineage-specific ohnolog resolution, LORe) (Gundappa et al., 2022; Lien et al., 2016; Macqueen & Johnston, 2014; Robertson et al., 2017). The difference in rediploidization timing increased the genomic complexity of the Atlantic salmon genome with LORe regions expressing over 95% sequence

similarity (Lien et al., 2016). At the same time, retention of duplicated genome resulted in large gene expansions, driving the evolution and functional diversification of gene families (Robertsen & Greiner-Tollersrud, 2024). Retention, loss and divergence of whole-genome and tandemly duplicated genes had an impact in the evolution and diversification of gene repertoires linked to immune response (Robertsen & Greiner-Tollersrud, 2024; Chris J. Secombes & Zou, 2017; J. Zou et al., 2014). In addition to gene evolution, earlier research has also speculated the contribution of the Ss4R duplication event in the development of novel abilities such as anadromy, however there is little empirical evidence supporting such hypothesis (Alexandrou et al., 2013; Robertson et al., 2017)

4.2 Application of genomics methods in Atlantic salmon research

Investigation of the impact of WGD on the genome structure and function is crucial for understanding the biology and evolution of Atlantic salmon. The following section describes key areas of genomics research including genetic structure, gene expression and regulation, drawing examples from application of these methods in Atlantic salmon research.

4.2.1 Study of genome structure and genetic variation

Understanding the genomic basis of phenotypic expression is important for the development of effective strategies towards improvement of aquaculture populations. Changes and/or rearrangements in the genome sequence of organisms -described collectively as genetic variation- play a key role in shaping the genomic landscape of aquaculture traits, including capacity of the organism to respond to environmental stimuli (Mukiibi et al., 2025). Genetic variation can differ in size and complexity, ranging from a single nucleotide polymorphism to millions of base pairs (Mbp). Depending on their structural characteristics (such as size, nucleotide repetition and orientation), genetic variants can be classified into different categories, including positional mutations (SNPs), short sequence insertions or deletions (INDELs), tandem repeats of one, two or more nucleotides (microsatellites or STRs, respectively) and large structural variants (SVs, conventionally defined as > 50 bp) (Mahmoud et al., 2019). Examples of different types of genetic variation are illustrated in **Figure 2**. Based on their structural features, SVs can be further distinguished into sequence insertions, deletions, duplications, inversions and translocations. Albeit less abundant in the genome, due to their size, SVs can affect more base pairs than SNPs (Catanach et al., 2019; Hämälä et al., 2021).

Genetic variants are encountered across the entire genome; however, overlap of genetic variation with functional genomic regions -such as protein-coding regions and regulatory elements- can directly impact phenotypic expression. Functional alterations led by genomic variants can affect the expression level of genes as well as the product and/or efficiency of transcribed proteins, ultimately impacting phenotypic responses (Haunshi et al., 2018; Minafra et al., 2023; Noreen et al., 2012).

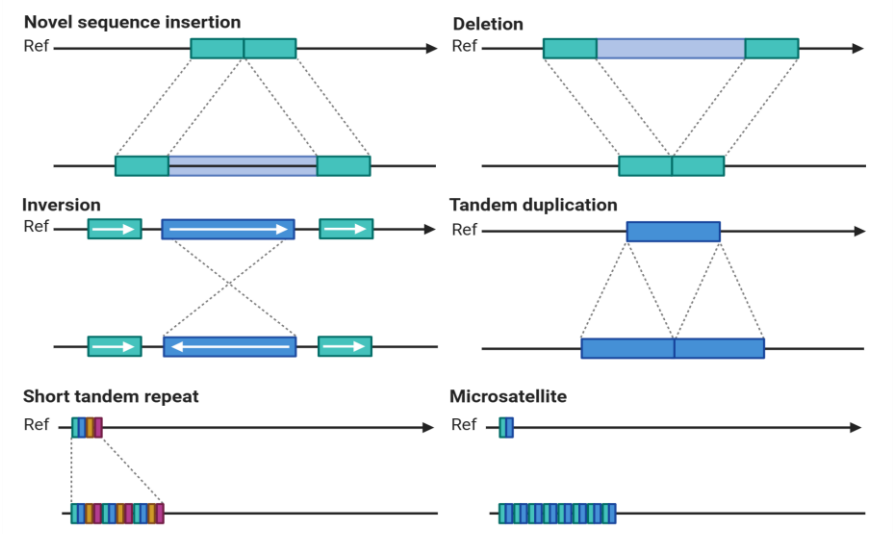


Figure 2. Schematic representation of different types of genetic variation. Large dark and light blue highlighted regions demonstrate types of structural variation (genomic rearrangements >50 bp). Small multi-colored regions represent tandem repeats, distinguished based on the number of repeated nucleotides.

In recent years, application of next generation sequencing (NGS) technologies has enabled the detection and association of genetic variants with major challenges in Atlantic salmon aquaculture, including disease resistance (Fraslin et al., 2020; Robledo, Palaiokostas, et al., 2018). Short-read sequencing has been instrumental in the identification of simple genetic variants such as SNPs and small INDELs. Refinement of detected variants and establishment of genotyping arrays -often including strategically selected genetic variation that is associated with aquaculture traits (QTLs)- has facilitated cost-efficient genotyping of large aquaculture populations. Furthermore, implementation of this information into breeding programs (either through marker assisted - or genomic selection) has improved

drastically the resistance of farmed populations against a broad range of pathogenic threats (Fraslin et al., 2020; Ross D. Houston et al., 2020; Moen et al., 2009).

Although application of short-read sequencing played a key role in the investigation of the Atlantic salmon genome, the read-span limitations of these technologies (typically 150–250 bp) face significant challenges in respect to genomic complexity. In particular, the length of produced sequences is not suitable for detection of complex genetic variation such as SVs and repeat elements (Mahmoud et al., 2019). Furthermore, alignment of short reads against complex and/or repetitive genomic regions is challenging as short sequences are unable to be accurately placed (Cechova, 2020; Phan et al., 2015). Considering these weaknesses, the genomic complexity of the Atlantic salmon genome (Bertolotti et al., 2020; Lien et al., 2016) significantly impacted the quality, contiguity and genomic resolution of the first, short read-based genome reference (accession number GCF_000233375.1). As a result, the first Atlantic salmon genome reference encountered gaps and errors, affecting the detection as well as interpretation of findings from downstream analyses (Hillestad et al., 2020).

In response to the limitations of short read sequencing, the emergence of long and ultra-long read sequencing technologies (including Oxford Nanopore technologies and Pacific Biosciences) has assisted in the assembly of complex genomic regions, leading to a dramatic increase in genomic resolution and accuracy (Hai et al., 2022). As a result of the application of such technologies in Atlantic salmon, a superior genome reference is now available (Ssal_v3.1, accession GCA_905237065.2) (Stenløkk, 2023). Use of this genome assembly has revealed novel genetic variation (Monsen, 2022; Stenløkk, 2023), unlocking new possibilities in the study of complex genetic variation with causal impact on immune response (Bai et al., 2020; Jobson & Roberts, 2022; Wu et al., 2021). In addition, development of a functional annotation based on this high quality reference genome has improved the accuracy of predicted gene models (Johnston et al., 2024), allowing for discovery and in-depth study of genes, particularly those impacted by the salmonid-specific WGD event.

4.2.2 Investigation of the functional landscape

While the study of Atlantic salmon genomics offers a comprehensive view of the organism's genetic architecture, the field of functional genomics investigates how changes in the genetic landscape can affect the regulation of biological functions. This is achieved through the study of gene expression and regulation -how, when,

and to what extent the coding regions of the genome are translated into functional products. Together, the different aspects of functional genomics can be applied to map the functional landscape of the genome: the network of genes, regulatory elements, and molecular processes that drive biological activity.

4.2.2.1 Genes and gene expression

Gene expression can be described as the process of translating the genomic information packed inside genes into functional products such as proteins and RNA molecules. Genes are therefore regarded as heritable units of DNA sequence that are able to contribute to phenotypic expression in the organism (Portin & Wilkins, 2017). In that sense, the study of gene expression and regulation plays a key role in understanding the molecular mechanisms modulating important biological processes (Robledo et al., 2016).

Based on the central dogma of molecular biology, the genetic information flows from DNA to RNA and finally to protein (Crick, 1970); the DNA sequence contained inside genes is transcribed to messenger (m)RNA and translated into amino acid chains, ultimately resulting in production of functional protein products. Initiation of RNA transcription begins with recognition of the promoter region of a gene by RNA polymerase (**Figure 3**). Recruitment of RNA polymerase on the promoter opens the DNA duplex and initiates RNA synthesis, ultimately elongating the transcribed RNA as it progresses through the DNA template (Cramer, 2019). Termination of the RNA transcription occurs when RNA polymerase ceases synthesis; both the enzyme and the newly transcribed (nascent) RNA are then released from the DNA template.

In addition to the transcription of prtotein-coding genes, exponential increase in our ability to retrieve information from DNA sequence and analyse transcribed RNA has led to the discovery of genes whose transcription does not code for functional proteins but is involved in regulation of gene expression (ncRNAs, **Figure 3**). Based on features such as functionality and length, ncRNAs can be classified into several categories (Poliseno et al., 2024). For eukaryotes, nuclear RNA polymerase II (Pol II) is responsible for synthesizing all protein-coding RNA as well as different types of ncRNA (Kuehner et al., 2011).

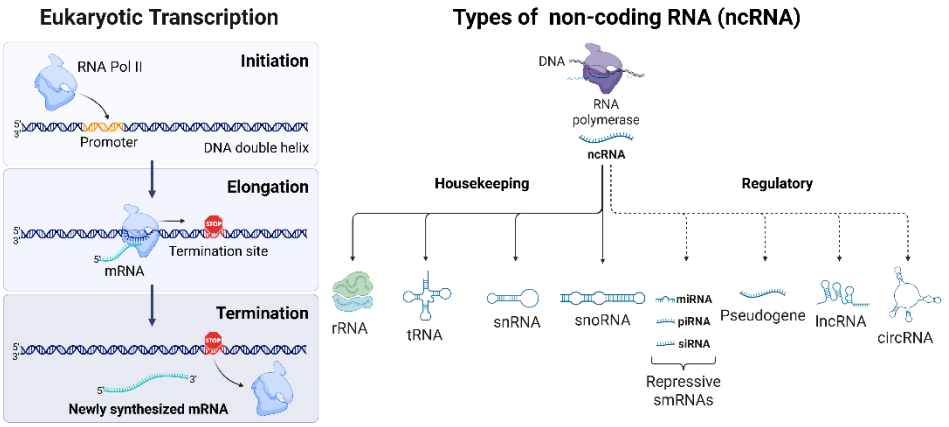


Figure 3. (left) Simplistic schematic of the process of gene transcription. (right) Different types of transcribed RNA that does not result in production of functional protein profuct (ncRNA). Based on their functionality ncRNAs can be distinguished between Housekeeping and regulatory with further classification focusing on other features such as transcript length. Image created on BioRender.

4.2.2.2 Gene regulation and epigenetic changes

Genes may or may not be expressed at all times. Regulation of genes' expression state is vitally important as it allows for the orchestration of cellular and molecular processes required for the survival and growth of the organism, as well as for response against variable environmental cues (Peter & Davidson, 2017; C. Zou et al., 2011). Since gene expression relies heavily on the transcription of a DNA template, presence and integrity of genetic regulatory elements is essential for controlling gene activity. In that sense, changes occurring on the DNA sequence can have serious implications in both the transcribed product as well as in the regulation of gene transcription (Chatterjee & Ahituv, 2017; Rojano et al., 2019; J. Wang et al., 2024).

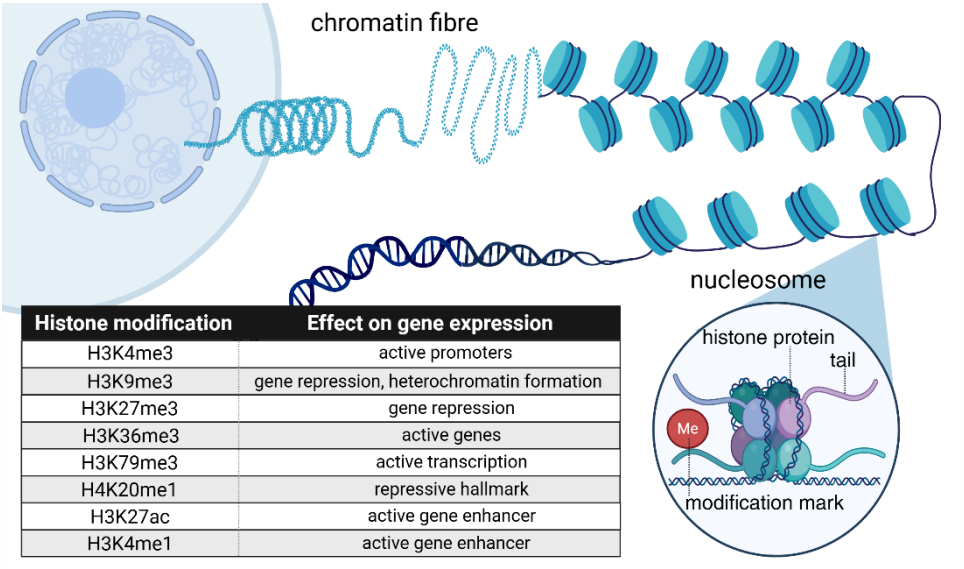
Besides structural changes, presence of heritable yet DNA-independent (epigenetic) alterations can also impact gene expression. These alterations, led by non-genetic factors such as temperature, stress and pathogenic stimuli, can influence the phenotypic plasticity of individuals (Boltana et al., 2018; Hamon & Cossart, 2008). The study of the epigenetic landscape focuses mainly on changes in the chromatin structure, DNA methylation, post-translational modifications of histone proteins and ncRNAs (Y. Li, 2021). In the present thesis, the term epigenetics is used

exclusively to describe functional changes imposed by alterations of the chromatin structure and histone protein modifications.

4.2.2.3 Chromatin structure and the impact of organization changes

In eukaryotic genomes, the DNA is organized into nucleosomes, the fundamental units of chromatin. Each nucleosome consists of approximately 147 base pairs of DNA wrapped around an octamer of histone proteins (Kouzarides, 2007) (**Figure 4**). The histone octamer is composed of a pair of four core histone proteins (H2A, H2B, H3 and H4) and each protein contains two distinct domains: (1) a central globular domain and (2) an amino-terminal residue extension (conventionally referred to as 'tail'). The globular domain, resides within the nucleosome, regulating histone-histone interactions as well as the arrangement of nucleosome-wrapped DNA. On the other hand, the histone tail protrudes from the nucleosome surface, serving as key site for the accumulation of post-translational modifications (PTMs) that regulate chromatin dynamics and gene expression (Peterson & Laniel, 2004; T. Zhang et al., 2015). One after another, nucleosomes are packed together, forming the chromatin fibre (**Figure 4**).

Variability in nucleosome density across the chromatin fibre can result in regions of loosely packed— known as "open" or "accessible"— or densely packed, and therefore non-accessible, chromatin. Depending on the level of compaction, chromatin can be defined as *Euchromatin* or *Heterochromatin*, describing the respective state of accessibility or non-accessibility (Hamon & Cossart, 2008). Accessible chromatin regions allow the transcriptional machinery to access exposed DNA sites in the vicinity of genes and interfere with their transcription. This is achieved through binding of transcription factors (TFs) with accessible DNA binding sites (TFBSs) and engagement of transcription by RNA polymerase II (Gottesfeld & Carey, 2018; Luo et al., 2022). Although the TFBS sites are required to be proximal to the transcription start site of a gene in order to impact its transcription, the actual distance can vary as far as > 1 Mbp (Leviyang, 2021). Due to its influence on the transcription of protein-coding elements, chromatin accessibility is regarded as the primary component of gene regulation and expression.



Histone modification	Effect on gene expression
H3K4me3	active promoters
H3K9me3	gene repression, heterochromatin formation
H3K27me3	gene repression
H3K36me3	active genes
H3K79me3	active transcription
H4K20me1	repressive hallmark
H3K27ac	active gene enhancer
H3K4me1	active gene enhancer

Figure 4. Schematic representation of the chromatin structure. DNA sequence is wrapped around eight histone proteins, forming the nucleosome structure. Densely packed nucleosomes construct the chromatin fibre. Post-translational modifications (PTMs) that accumulate on the histone tail can lead to changes in the regulation of genes. Examples of histone modifications and their impact on gene regulation are retrieved from (Liu et al., 2022). Figure was created with BioRender.

A key feature of chromatin is its highly dynamic nature; during various biological processes the accessibility state of chromatin can alternate, modulating accordingly the accessibility of TFBS sites. In turn, changes in TFBS accessibility can directly impact transcriptional activity, up- or down-regulating gene expression (Q. Zhang & Cao, 2019). These chromatin dynamics are orchestrated by alterations of the chromatin structure, including nucleosome remodelling as well as the activity of histone-modifying enzymes (Allis & Jenuwein, 2016; Cairns, 2009; X.-Y. Li et al., 2011).

Post-translational modifications (PTMs) play a central role in the modulation of the chromatin structure. The intrinsic flexibility of the histone 'tail' enables the acquisition of a wide range of PTMs (Peterson & Laniel, 2004; T. Zhang et al., 2015). In turn, PTM accumulation on the histone tail can affect nucleosome-DNA interactions, orchestrate chromatin structure remodelling complexes to influence transcription but also regulate functions such as DNA repair, recombination, and

chromosome segregation (Talbert & Henikoff, 2021). To date, over 100 different types of PTMs have been identified, while some of the most extensively studied PTMs that influence transcription include histone acetylation, methylation, ubiquitylation and phosphorylation (Ku et al., 2011). Of these, PTMs that are associated with transcriptional activity include lysine acetylation on the tails of histones H3 and H4 (H3k[]ac and H4k[]ac, respectively), tri-methylation of H3 histone lysines 4, 36 or 79 (H3k[4, 36, 79]me3) and ubiquitylation of the H2B histone. Conversely, PTMs associated with transcriptional repression include trimethylation of H3 histone lysines 9 or 27 (H3k[9,27]me3) and ubiquitylation of H2A on lysine 119 (**Figure 5**) (T. Zhang et al., 2015). Although in the past only certain PTMs were defined as dynamic, nowadays most if not all of the major histone modifications are considered to be at least to some extent reversible (Millán-Zambrano et al., 2022; Talbert & Henikoff, 2021).

4.2.2.4 Study of functional genomics in Atlantic salmon

Beyond the detection of genetic variation, next generation sequencing (NGS) technologies have been instrumental in the study of the Atlantic salmon functional landscape. In particular, use of NGS methods for the study of gene expression (mainly through mRNA sequencing, RNA-seq) has led to detection and transcriptomic profiling of genes regulating core biological processes (M. Iversen et al., 2021; Moen et al., 2015; Sveen et al., 2017). In addition, investigation of the Atlantic salmon transcriptome has also revealed the impact of the Ss4R duplication event on the salmonid functional landscape; transcriptomic profiling of WGD-duplicated genes has allowed provides insight regarding the evolutionary history of gene duplicates, evolving through functional redundancy, specialization and/or novelty (Sandve et al., 2018)

Similar to transcriptomics analysis, in recent years, study of epigenetic changes is also performed using NGS methods. In particular, analysis of chromatin accessibility and post-translational modifications is performed through *assay for transposase-accessible chromatin with sequencing* (ATAC-seq) and *chromatin immuno-precipitation with sequencing* (ChIP-seq) methods, respectively (Buenrostro et al., 2013, 2015; Nakato & Sakata, 2021; Park, 2009). Analysis of epigenetics information facilitated the investigation of DNA-dependent and independent changes influencing the regulation of biological functions, including important aquaculture traits (Beemelmanns et al., 2021; Shengnan Gao et al., 2024; Koch et al.,

2023; Uren Webster et al., 2018; Y. Wang et al., 2024). However, study of epigenetics in Atlantic salmon is currently limited.

4.3 The Salmonid immune system

The immune system plays a key role in the health and survival of organisms, offering protection against pathogen infections and assisting in the retention of physiological balance (homeostasis). Because of this, the immune system has a direct impact on the development, growth as well as the successful adaptation of aquatic species to demanding environmental conditions (Magnadottir, 2010).

Similarly to higher animals, the salmonid immune system can be subdivided into the innate (non-specific) and the adaptive (acquired) systems, with substantial crosstalk occurring between the two components (Magnadóttir, 2006). The innate and adaptive arms of immunity differ in terms of response timing and memory; innate immunity responds rapidly -yet with no immune memory- towards the detection of non-self-pathogenic organisms and then coordinates the induction of highly specific -but slower- adaptive immune responses (Collet, 2014; Smith et al., 2019). The innate and adaptive immunity components are briefly described below, focusing particularly on the aspect of antiviral immunity.

4.3.1 Innate immune response

The innate immune system constitutes the first line of the host defence mechanism and is relatively conserved across teleost fish and higher vertebrates (Mokhtar et al., 2023). However, lack in diversity of the adaptive system of teleost fish (compared to that of mammals), but also higher pathogenic diversity of aquatic environments has increased the dependency of teleost species on innate immunity, even from the early embryonic stages (Rombout et al., 2005). For anadromous fish such as Atlantic salmon, experience of additional challenges related to energy allocation for intensive physiological processes during migration as well as transition through highly diverse environments, increases the importance of maintaining strong acute responses to pathogenic stimuli, resulting in marked immunogenetic differences between taxa (Aoki et al., 2013; Colgan et al., 2021).

In fish, the innate immune system can be subdivided into three layers: (1) physical barriers, (2) cellular components, and (3) humoral components. Physical barriers include the scales and mucosal layer as well as the epithelial layers of the gills and the gastrointestinal tract (Mokhtar et al., 2023; Smith et al., 2019). The ccellular

components of innate immunity are composed of a range of leukocytes including macrophages, monocytes, granulocytes, dendritic cells and natural killer cells (Smith et al., 2019). As fish lack bone marrow, production of these leukocytes occurs primarily in the anterior (or head) kidney, the thymus, spleen and gills, establishing these organs as critical components of fish immunity (Bjørgen et al., 2024; Haugarvoll et al., 2008; Lulijwa et al., 2019). Based on their cell type, pathogen-activated leukocytes can take part in different processes such as phagocytosis, cytokine production or stimulation of the adaptive immune response through antigen presentation (Smith et al., 2019). Finally, the humoral components consist of macromolecules released by the pathogen activated cells; these molecules regulate functions such as macrophage activation and phagocytosis, as well as inflammatory and bactericidal responses. Some of the most studied humoral parameters include growth inhibitors, lytic enzymes (such as lysozyme), components of the complement system, antibacterial peptides, cytokines and acute phase proteins that are secreted into the plasma during stress or pathogen infection (Collet, 2014; Magnadóttir, 2006; Smith et al., 2019).

Initiation of the innate immune system starts with recognition of non-self. Molecular patterns associated with pathogens or host tissue damage (PAMPs and DAMPs, respectively) bind into germline-encoded pattern recognition receptors (PRRs) of the host, triggering a cellular response (Bianchi, 2007; Mokhtar et al., 2023). Several groups of PRRs exist in fish, including Toll-like receptors (TLRs), retinoic acid inducible gene I (RIG-I) - like receptors (RLRs), c-type lectins (CLRs), Melanoma-Differentiation-Associated gene 5 (MDA5) and nucleotide-binding oligomerization domain-like receptors (NLRs), reviewed in (Smith et al., 2019). Activation of receptors depends on binding specificity as well as the nature of the released pathogenic molecules. Binding of ligands with the appropriate receptor results in stimulation of the innate immune system, followed by acute inflammation and an early protective response that includes leukocyte and cytokine production, phagocytosis and ultimately pathogen elimination and establishment of immune memory (Soliman & Barreda, 2023). Alternatively, upon stimulation of the innate response the adaptive immune system can be activated through antigen presentation and use previously acquired immune memory to confer highly pathogen-specific immunity and prohibit prolonged infection (Soliman & Barreda, 2023).

Due to its quick response timing during pathogen infection, the innate immune system is essential for salmonids —particularly in response against viral pathogens (Verrier et al., 2013). Earlier research suggested that rapid innate immune responses during viral infection could suffice for inhibition of viral replication and elimination of viral pathogens without inducing adaptive immunity (Collet, 2014; Takeuchi & Akira, 2009). Under this hypothesis, the innate immune system poses as a powerful mechanism for pathogen resistance with a key role in antiviral immunity.

4.3.1.1 Cytokines and the role of Interferons (IFNs) in antiviral responses

Among the different components of innate immunity, one with central role in antiviral responses is that of cytokines. Cytokines play a key role in the regulation of innate immune system by responding to inflammation signalling during pathogen invasion and regulating inflammatory events (C. J. Secombes et al., 2001; Soliman & Barreda, 2023). These protein molecules participate in intercellular communication and signal transmission and are involved in regulatory processes of both the innate as well as the adaptive immune system (Morris et al., 2018; Sakai et al., 2021).

Over 100 different types of cytokines have been identified in humans, most of which are also found in fish. In addition, multiple paralogs and cytokine family expansions are encountered in teleosts -including salmonids-, likely due to these species experiencing additional rounds of WGD (J. Zou & Secombes, 2016). Fish cytokines comprise large protein families, consisting of interleukins (ILs), interferons (IFNs), chemokines and Tumor Necrosis Factors (TNFs). Family members can be further distinguished based on their structure, involvement in innate or adaptive immune responses but also based on their pro- or anti-inflammatory role (Sakai et al., 2021; Soliman & Barreda, 2023). The present thesis focuses exclusively on the IFN family of cytokines which constitutes a hallmark for antiviral immunity.

In mammals, IFNs -together with some IL families are members of the class II α-helical cytokines (Redmond et al., 2019; J. Zou & Secombes, 2016). Based on their binding receptors, genomic organization, and structural homology, three types of IFNs are encountered in fish, namely IFN-I (analogous to human IFN-α, β, κ), IFN-II (analogous to IFN-γ), and IFN-III (analogous to IFN-λs) (Manry et al., 2011; Pestka et al., 2004). Types I and III IFNs are both part of the innate immune system, however they are related to antiviral responses with different levels of specificity;

type I IFNs are present in all nucleated cells whereas type III IFNs are normally found in high infection risk tissues (Gan et al., 2020). On the other hand, type II IFNs are more versatile and are involved in cell-mediated immunity against a broad range of pathogens during adaptive immune responses (Mokhtar et al., 2023; T. Wang & Secombes, 2013). Although in the past the origins of type III IFNs in teleost species have been greatly debated (Chang et al., 2009; Lutfalla et al., 2003; Robertsen, 2006), recent phylogenetic analyses revealed that the fish type III IFNs constitute a sister clade of IFN-I (Redmond et al., 2019; Z. Wang et al., 2022).

Across teleost fish, Atlantic salmon appears to have acquired one of the largest and most complex repertoires of type I IFNs and their respective receptors (Robertsen & Greiner-Tollersrud, 2024; Chris J. Secombes & Zou, 2017; J. Zou et al., 2014). Complexity of the IFN-I gene and receptor repertoires is likely attributed to the effect of the salmonid-specific WGD that -together with multiple rounds of tandem duplication- led to high turnover rates for these genes, including large gene gains/losses (Chris J. Secombes & Zou, 2017; B. Sun et al., 2014).

Following the teleost-specific (Ts3R) whole genome duplication event, the expansion of the IFN-I gene family across teleost species led to significant differences in the induction of IFN signalling between fish and mammals (Aoki et al., 2013). Yet, the IFN-I downstream signalling pathway appears to be rather conserved across clades, mediated via the Janus kinase and Signal transducer and activator of transcription (JAK-STAT) signalling pathway (Collet, 2014; Sobhkhez et al., 2014). Specifically, interaction between type I IFNs and their corresponding receptors (referred to as CRFBs, the teleost equivalents of mammalian IFNARs) activates the receptor-associated Janus family tyrosine kinases Tyk2 and Jak1 and binds them to the same receptor complex (**Figure 5**). The activated kinases then promote phosphorylation of the STAT1 and STAT2 transcription factors which - together with the IFN regulatory factor IRF9 - form the IFN regulatory gene factor 3 (ISGF3) complex. Translocation of the ISGF3 complex from the cell membrane to the nucleus and binding with IFN-stimulated response elements (ISREs) induces transcription of a large cluster of IFN-stimulated genes (ISGs), triggering the induction of a cascade of antiviral responses (Gan et al., 2020; Schneider et al., 2014).

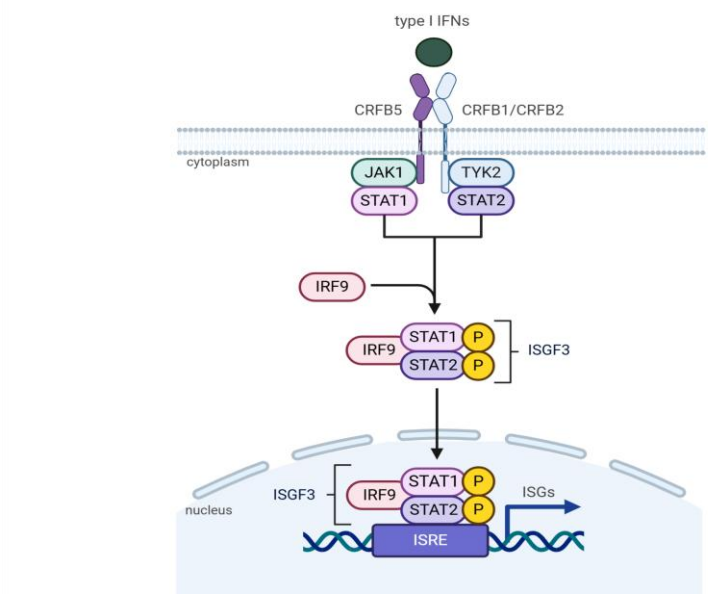


Figure 5. The IFN-I downstream signalling, mediated via the Janus kinase and Signal transducer and activator of transcription (JAK-STAT) signalling pathway. Visualization adjusted from (Gan et al., 2020; Robertsen, 2018), created with BioRender.

4.3.1.2 IFN-stimulated genes (ISGs)

Described above, activation of IFN-mediated responses leads to transcriptional initiation of a large group of antiviral genes known as IFN-stimulated genes (ISGs). Although many of these genes are normally expressed at baseline levels within the cell, upon IFN stimulation, ISGs work synergistically in order to limit viral replication and thus viral infection in the cell (Poynter & DeWitte-Orr, 2016). Furthermore, the large abundance of ISGs allows these genes to not only participate in the regulation of innate and adaptive immune responses, but to also obtain very diverse roles, specializing in function across every possible stage of the viral life-cycle (Grimholt et al., 2020; Schneider et al., 2014). Activation of ISGs occurs via binding of transcription factors (TFs) -such as the ISGF3 complex- to the ISRE elements located close to the ISG transcription start site (TSS). Engagement of the ISRE site with the ISGF3 complex induces ISG transcription, activating the antiviral response (Ivashkiv & Donlin, 2014).

Hundreds of ISGs have been identified across mammals, several of which are also found in fish (Clark et al., 2023; Schoggins et al., 2011). Recent studies have revealed that core ISGs are rather conserved across human and teleosts (Clark et al., 2023). For Atlantic salmon, the genes Mx, ISG15 and viperin are among the most extensively studied ISGs (Robertsen, 2008; Verrier et al., 2011). Besides conserved ISGs, the presence of fish-specific ISGs has also been reported (Ortega-Villaizan et al., 2022; Y.-B. Zhang et al., 2013). Although the life history and functional impact of fish-specific ISGs is not fully understood yet, their role in response to important aquaculture diseases makes them key components in the establishment of disease management strategies for farmed populations (Krasnov et al., 2021).

4.3.2 Adaptive immune responses

Adaptive immunity follows the acute responses of the innate system. Adaptive immune responses can be triggered either immediately through antigen presentation and use of the organism's immune memory or later on in the immune response, following the induction of innate immunity (Mokhtar et al., 2023). Several components of the adaptive system can be traced back to jawless vertebrates - including B-cell and T-cell activation (Smith et al., 2019). However, integral elements of adaptive immunity such as the T-cell receptors (TCRs), Major Histocompatibility Complex (MHC) proteins and immunoglobulins (Igs) are believed to have evolved later in bony fish and other jawed vertebrates (Flajnik, 2018). The latter elements are particularly important for the adaptive immune system as they're involved in highly specific immune responses but also in the establishment of host immune memory (Wilson, 2017).

Similarly to the innate immune system, the adaptive response mechanism comprises both humoral and cellular components, with B-cells and T-cells each playing central roles on the first and the second component, respectively (Smith et al., 2019). In particular, B-cells are responsible for the production of high affinity Immunoglobulins (Igs) against detected antigens and the presentation of these peptides to T-cells. Main vehicle for antigen presentation constitutes the group of MHC proteins, protein complexes involved in detection, process and presentation of molecules to the membrane-bound TCRs (Smith et al., 2019). Interaction between the presented antigens and TCRs activates the helper (CD4+), regulatory (CD4+) or cytotoxic (CD8+) T-cells, navigating the immune response (Nakanishi et al., 2015). During antigen presentation, activation of the appropriate T-cell type relies heavily

on stimulation of the appropriate TCR receptor and its ability to recognize antigens presented through the MHC proteins (Nakanishi et al., 2015).

4.3.2.1 The Major Histocompatibility Complex (MHC)

MHC proteins are important for the induction and regulation of adaptive immunity (Wilson, 2017). Role of these intracellular peptides is to discriminate between self and non-self proteins and interact with T-cell receptors to initiate T-cell mediated responses. The MHC loci show impressively high levels of polymorphism, leading to acquisition of a diverse repertoire of represented antigens required for pathogen-specific immune responses (Smith et al., 2019). Based on their structure, function, and antigen presentation, MHC molecules can be subdivided into the MHC-I and MHC-II classes, whereas distinction based on their variability can further dissect each MHC class proteins into highly polymorphic and highly conserved groups termed classical and non-classical, respectively (Flajnik, 2018; Wilson, 2017). Classical MHC-I proteins can be found on most cell types and activate CD8+ T-cells, thus initiating a protective immune response. On the other hand, classical MHC-II molecules are present on specialized antigen-presenting cells and interact with CD4+ T-cells through presentation of antigens originating from foreign endocytosed proteins (Grimholt et al., 2003; Sundaram et al., 2020).

Across vertebrates the two MHC complex classes are found within the same genomic region and share substantial genetic linkage. Contrary to that, in teleost fish these loci appear to be segregated across different chromosomes while in certain species the MHC-II locus seems to be completely lost (Grimholt, 2016). Breaking of genetic linkage and segregation of the MHC loci in teleost fish has been attributed to rapid increase of immune diversity caused by a WGD event in the ancestor of the teleost group (Dijkstra et al., 2013). For Atlantic salmon, experience of the lineage-specific WGD resulted in acquisition of additional MHC-I and MHC-II gene clusters, located across homologous regions in the salmonid genome (Harstad et al., 2008). Retention of the duplicated MHC loci in Atlantic salmon has further increased the immunogenetic diversity and functional complexity of MHC molecules and their respective pathways (Grimholt et al., 2020; Harstad et al., 2008; Lukacs et al., 2007).

Despite the increased genomic complexity of the Atlantic salmon MHC loci, association of these loci with resistance against important aquaculture pathogens (Kjøglum et al., 2006; Moen et al., 2009) highlights the need for thorough

investigation of these molecules. In addition, selection of farmed populations for the particular loci makes the MHC complexes prime targets for the improvement of disease resistance for farmed Atlantic salmon (Fraslin et al., 2020).

4.4 Advances in the study of genomic architecture for immune traits in farmed Atlantic salmon

Discussed previously, Application of NGS technologies has played a pivotal role in our understanding of the Atlantic salmon functional landscape. In particular, use of mRNA sequencing has allowed the transcriptomic profiling of molecular mechanisms associated with response against aquaculture diseases, revealing genes with key role in pathogen resistance (Aslam et al., 2020; Eslamloo et al., 2020; Valenzuela-Miranda et al., 2015). In addition, comparison of the transcriptomic profile between healthy and infected individuals has helped identify genes with regulatory differences during host-pathogen interactions (Robledo, Gutiérrez, et al., 2018). Identification of such differentially regulated genes can serve as indicators of immune fitness and provide an effective method of monitoring the resilience of farmed populations against pathogens. At the same time these genes can also be used to assess the effectiveness of applied preventative measures, such as antibiotic treatments and vaccinations (Krasnov et al., 2021; Thorarinsson et al., 2021). Furthermore, detection and association of genetic variation with the expression profile of immune-related genes offers valuable insights regarding how changes in the DNA sequence can functionally influence pathogen resistance (Fraslin et al., 2020). Implementation of such information in aquaculture breeding programs - either through marker-assisted or genomic-selection- has the potential to improve the resilience of farmed populations and increase the overall welfare and sustainability of the aquaculture industry (Ross D. Houston et al., 2020; Mukiibi et al., 2025).

Although well established in humans (X. Wang et al., 2021), the impact of epigenetic changes on disease response has only recently been a topic of study in Atlantic salmon. As a result, there has been limited epigenetics research in respect to Atlantic salmon immunity; previous studies focused on the effect of chromatin remodelling on important biological processes such as metabolism, maturation and smoltification (Harvey et al., 2024; Holen et al., 2023; Mohamed et al., 2022). Similarly to the study of chromatin structure, not much is known about the effect of histone modifications on Atlantic salmon immunity. However, studies in other

salmonid species have provided evidence regarding the epigenetic reprogramming of IFN responses during viral infection, whereas similar patterns of epigenetic control during antiviral signalling have also been observed in other teleost species (Medina-Gali et al., 2018; Mokhtar et al., 2023). Taken together, the study of epigenetic influences is becoming increasingly important in understanding the regulatory mechanisms underlying immune responses in aquaculture (Liu et al., 2022).

Integration of the epigenetic information into genomics studies offers an exciting and highly promising opportunity to enhance robustness against aquaculture diseases and further explore the genetic architecture of immune; for example through detection of non-coding genetic variation with regulatory impact on disease resistance (Johnston et al., 2024; Mukiibi et al., 2025). Embracing this opportunity, collective efforts in recent years through the AQUA-FAANG consortium have managed to provide large and in-depth catalogs of genome-wide epigenetics datasets for several aquaculture species, including Atlantic salmon (Johnston et al., 2024). These publicly available catalogs, including marks of epigenetic regulation during immuno-stimulation, open new horizons in the study of gene regulation during disease infection.

4.5 Aims and Outline of the thesis

Main objective of this thesis was to investigate the impact of genomic architecture, evolutionary history and selection pressures in shaping important aquaculture traits in farmed Atlantic salmon. Three papers outline this work:

Paper 1: ***Dissecting the genetic basis of response to salmonid alphavirus in Atlantic salmon***

Study 1 focused on uncovering the genomic landscape underlying response to pancreas disease (PD), a major challenge for Atlantic salmon aquaculture. To achieve this, the study leveraged the latest advances in Atlantic salmon genomics including short and long-read based genomics and transcriptomics data, epigenetics information as well as the latest version of the Atlantic salmon reference genome. Objective of this study was to therefore use genomic, functional and regulatory information in order to fine map the genetic architecture of PD resistance and identify simple and complex genetic variants with putative functional impact on disease response.

Paper 2: ***Evolution of gig immune genes in teleosts shaped by gene gain and loss: regulatory and functional insights from Atlantic salmon***

Study 2 builds upon the findings of Study 1 and investigates the evolutionary trajectory of two gene families with large antiviral impact on Atlantic salmon immune response, *Gig1* and *Gig2*. Through the scope of the salmonid-specific WGD event, this study explored how gene duplication, functional divergence, and repeat content collectively shape the evolutionary fate of immune genes with high importance for aquaculture.

Paper 3: ***Parallel selection in domesticated Atlantic salmon from divergent founders including parallel selection on WGD derived homeologous regions***

Study 3 provides a population-level perspective of how whole genome duplications can provide the genetic redundancy and plasticity required for the evolution of adaptive immunity (MHC loci), particularly in response to artificial selection and domestication pressures.

Papers 1 and 2 were fully conceptualized and conducted in the scope of the present thesis. On the other hand, paper 3 represents a collaborative effort to which I contributed specifically to the preparation and haplotyping of genotype datasets as well as in the analysis of gene expression patterns.

4.6 Brief summary of paper results

Paper 1: Dissecting the genetic basis of response to salmonid alphavirus in Atlantic salmon

In paper 1 we investigated the genomic architecture underlying pancreas disease (PD) resistance in Atlantic salmon. We performed a GWAS association analysis using survival records from multiple populations of fish infected with salmonid alphavirus (SAV3), either via intraperitoneal injection (IP) or through infectious cohabitation (CH) with injected fish. Association analysis identified two genomic regions with significant association to PD resistance, specifically a major effect locus on chromosome Ssa03 and a secondary infection model-specific region on chromosome ssa07. Focusing on the Ssa03 region, we identified a narrow haplotype block (54,000 bp) that included the most highly associated SNP to PD resistance (top SNP). This haplotype block harboured six genes including *nlrp3*, three tandemly duplicated copies of the *gig1-like* gene, *abcc6b* and an unknown gene (named *sspd*). Transcriptomics analysis of fish infected with SAV3 virus revealed significant responses only for the three *gig1-like* gene copies within the Ssa03 haplotype block, suggesting these genes as likely causal candidates for PD resistance. Further analysis of gene expression profiles revealed an association between poor fish immunity and higher *gig1-like* expression. In addition, performance of an allele-specific expression analysis using genotype and gene expression data from samples with heterozygous genotype for the top SNP on Ssa03 detected multiple SNPs with putative functional impact on the expression of the first and the third *gig1-like* copies. Further investigation using whole-genome level short and long-read sequencing information revealed several SNPs and complex variants that segregated together with the top SNP for PD response. Finally, integration of regulatory information (ATAC-seq and ChIP-seq data) from fish infected with the viral mimic poly(I:C) uncovered additional variants with putative functional impact on PD resistance, including SNPs and two (AC)n microsatellites overlapping regulatory regions of the first and third *gig1-like* gene copies.

Paper 2: Evolution of gig immune genes in teleosts shaped by gene gain and loss: regulatory and functional insights from Atlantic salmon

Paper 2 focused on the gig1 and gig2 fish-specific antiviral gene families and explored how multiple rounds of whole-genome and segmental duplication have impacted the evolution and functional divergence of gig genes in Atlantic salmon.

Inspection of the *gig1* and *gig2* gain/loss gene trees across 11 species revealed large gene expansions in both families. Phylogenetic analysis of *gig* genes across species revealed high divergence among *gig* orthologs, indicating rapid evolution and diversification. Focusing on Atlantic salmon *gig* genes, functional domain prediction identified shared as well as family-specific functional domains in *gig* genes; presence, absence but also elongation/duplication of predicted domains between *gig* copies provided further evidence regarding the functional diversification of WGD-derived and tandemly duplicated paralogs. In addition to functional domains, repeat content was also detected in the coding sequence of *gig* paralogs. Overlap of repeats with coding regions was linked to duplication/elongation of functional domains as well as to regions of low protein stability. Further investigation using transcriptomics data from a SAV3 viral challenge as well as from in vitro and in vivo experiments of Interferon (IFN) stimulation showed a generalized antiviral response across paralogs, validating the genes' role in Atlantic salmon immunity. Additionally, genes located in homeologous regions of the salmonid genome showed similar expression profiles whereas the expression patterns of tandemly duplicated genes were more complicated. Detection of cis regulatory elements of *gig* genes using epigenetics information (ATAC-seq and ChIP-seq) from the same in vivo experiment of IFN stimulation, identified IFN-related *cis* regulators nearby most *gig* genes, confirming their role as classical IFN-stimulated genes (ISGs). In addition to the observed cis-regulators, potential interplay was suggested between segmentally duplicated genes, where the absence of an active regulator from one or more tandemly duplicated copies did not prohibit the respective genes' expression. While most gig genes showed similar immune responses across different tissues and experimental conditions, the regulatory and expression profiles of tandemly duplicated paralogs did not follow uniform patterns, suggesting possible evolution and functional divergence of these genes either through regulatory sharing or degeneration.

Paper 3: Parallel selection in domesticated Atlantic salmon from divergent founders including parallel selection on WGD derived homeologous regions

Paper 3 investigated parallel selection signatures across geographically isolated populations (West Norwegian and North American) of wild and farmed Atlantic salmon.

In paper 3, three population-based approaches (F_{st}, Tajima D and haplotype differentiation) were used across wild and aquaculture populations of West Norwegian and North American (east and west Atlantic, respectively) populations of Atlantic salmon to detect cross-population selective sweeps. Combination of analyses between the two geographic lineages revealed signatures of parallel selection across aquaculture populations of both European and North American origin; shared selective sweeps were identified for identical SNPs overlapping genes members of the Major histocompatibility complex (MHC) locus on chromosome Ssa14 (*daxx* and *tapbp)*. Further comparison of selection signatures between aquaculture populations identified additional SNPs under selection, overlapping gene pairs located in WGD-derived homeologous regions of the salmonid genome. Interestingly, for these gene pairs, a different copy of each gene was being selected in each of the two aquaculture populations. Gene expression profiling for the WGD-duplicated genes under selection revealed shared as well as diverged expression patterns, suggesting that similar selective pressures likely shape the phenotypic plasticity of aquaculture populations across lineages.

4.7 Discussion

4.7.1 The impact of highly repetitive or duplicated regions on mapping quality

The work in this thesis utilized genomics and transcriptomics data produced through long-read sequencing technologies as well as a highly improved long-read based reference genome (GCA_905237065.2). Use of long-read data played key role in downstream analyses; in paper 1 these data allowed in-depth investigation of the PD associated QTL locus on ssa03, uncovering complex variation with putative causal impact on disease resistance and revealing the presence and structure of three tandemly duplicated *gig1-like* genes associated with PD response (**Figure 5 in paper 1**). In paper 2, long-read transcriptomics data were used to identify the presence and structure of Atlantic salmon *gig1* and *gig2* paralogs, particularly those residing within LORe regions of the salmonid genome. Taken together, Use of DNA and mRNA data produced through long-read sequencing technologies increased significantly the genomic resolution of the analyses performed in this thesis, particularly regarding the LORe regions of the salmonid genome (**Figure 3 in paper 2**).

The complexity of LORe regions (Lien et al., 2016) has a negative impact on the genomic resolution of analyses, as it introduces ambiguity in the alignment and assembly of sequencing reads (Phan et al., 2015). Errors in the mapping of sequencing reads against highly repetitive and/or duplicated genomic regions can affect the results and interpretation of downstream analyses in a multi-faceted aspect. Taking examples from the impact of errors and gaps in the genome assembly, use of an older version of the Atlantic salmon reference genome (GCA_000233375.4) to study PD resistance had previously failed to detect the three *gig1-like* gene copies residing within the putative QTL region on chromosome Ssa03 (Aslam et al., 2020). In addition, earlier studies struggled to identify the precise location of the disease associated locus likely due to the earlier version of the Atlantic salmon reference genome lacking the required physical information for chromosome Ssa03. Furthermore, due to the increased sequence similarity shared across the q-arm of Ssa03 and a duplicated region on the p-arm of Ssa06 (**Figure 3 in paper 2**), erroneous disease-associated regions had also previously appeared on the ohnolog region of chromosome Ssa06 (Hillestad et al., 2020).

Besides errors in genome assembly, highly repetitive and duplicated genomic content can also challenge the alignment of sequencing reads against a given reference genome and/or transcriptome. Inability to align short sequencing reads against complex genomic regions limits the detection of putatively important genetic variation, particularly in the case of complex genetic variation such as structural variants, large indels and tandem repeats (Mahmoud et al., 2019). From the viewpoint of transcriptomics analyses, mapping errors caused by high sequence similarity of WGD as well as tandem-duplicated genes can also affect downstream analyses by introducing bias in the quantification of gene expression. In particular, increased sequence similarity shared across gene duplicates can reduce the mapping accuracy of mRNA sequencing reads and spuriously assign the same reads across multiple copies (Lan & Pritchard, 2016). Collectively, errors due to genomic complexity have a large impact on the accuracy of downstream analyses and consequently on the interpretation of underlying mechanisms that regulate important biological processes. Although at the moment, use of long-read sequencing technologies remains prohibitively expensive for the analysis of large aquaculture populations at the whole-genome level, integration of these technologies can provide major benefits in the study of organisms with similarly complicated genomes.

4.7.2 Disparities across functional annotation methods hinder the study of novel genes

Analyses performed for papers 1 and 2 identified large discrepancies between the publicly available versions of the Atlantic salmon functional annotation provided by the NCBI and Ensembl genome databases (Cunningham et al., 2022; O'Leary et al., 2016). Specifically in paper 1, the first exon of two tandemly duplicated *gig1-like* copies on Ssa03 appeared only in the functional annotation version of the NCBI database, whereas the gene *sspd* (ENSSSAG00000108304) was only found in the gene annotation provided by Ensembl (**Figure 6 in paper 1**). Similarly to the case of *sspd*, some of the Atlantic salmon *gig1* and *gig2* paralogs in paper 2 were annotated in only one of the two databases. To resolve such inconsistencies, analyses in papers 1 and 2 used a combination of short and long-read sequencing transcriptomics data from virally stimulated and healthy individuals, respectively. Combination of short and long read data and manual inspection of sequencing read coverage on the functionally annotated regions confirmed the structure of *gig* genes on paper 1. The same approach was followed in paper two -using the Ensembl

functional annotation as guide- thus establishing a consensus annotation for most of Atlantic salmon *gig1* and *gig2* paralogs.

Discrepancies in the functional annotation provided by publicly available databases is problematic for the study of genes (Vaattovaara et al., 2019). Annotation errors - like for example partial or absent gene models- can impact the interpretation of findings, from concealing the presence of genes to introducing errors in gene structure and therefore functional prediction (Cheng et al., 2017; Jupe et al., 2013). Such errors can directly affect the inference of phylogenetic relationships and ultimately the interpretation of genes' evolutionary history (Chisanga et al., 2022; R. Zhang et al., 2017). From the standpoint of paper 2, a recent study reported the presence of a *gig1* paralog in the spotted gar (*Lepisosteus oculatus*) (An et al., 2025). However, based on the Ensembl annotation used in paper 2, the particular *gig1* paralog (ENSLOCG00000006921) belonged to a different phylogenetic group and was therefore excluded from all following analyses. As a result, while other studies indicate the presence of a *gig1* orthologs in the Holostei family (and even as early as lobe-finned fish), paper 2 observed ortholog genes only among teleost species.

In addition to evolutionary studies, discrepancies in gene annotation can also impact quantitative genomics analyses. Errors in the quantification of gene expression can over- or underestimate gene responses and result in interpretation errors of functional findings. Furthermore, annotation discrepancies can also affect the detection of regulatory mechanisms and alternative transcripts (Chisanga et al., 2022; R. Zhang et al., 2017) as well as conceal the impact of putatively causal genetic variation. Taken from paper 1, use of the Ensembl-derived gene annotation would conceal the overlap of complex variants with functional regions of the first and third copies of gig1-like gene on chromosome ssa03 (**Figure 6 in paper 1**).

A possible source of discrepancies in the functional annotation of different public databases (such as the Ensembl and NCBI platforms) could stem from differences in the methodology used by each database during the prediction of gene models (Chisanga et al., 2022). Methodological differences during automated but also manual curation processes, coupled with lack in frequent maintenance and update of the available annotation files could explain the observed differences between public databases (Vaattovaara et al., 2019). While other factors can additionally impact the accuracy of gene detection and annotation, the analyses performed for the work of this thesis highlight the importance in presence of a consensus annotation that is constructed using appropriate genetic material and processed

using advanced molecular (sequencing technologies) and bioinformatics (annotation) methods.

4.7.3 The rapid evolution of immune genes complicates the inference of phylogenetic relationships

Paper 2 identified large expansions within *gig1* and *gig2* gene families, with most gene copies detected in species experiencing multiple rounds of whole-genome duplication (**Figure 1 in paper 2**). In addition, high amino acid sequence divergence was observed in both *gig* families, in line with previous reports (An et al., 2025; Y.-B. Zhang et al., 2013). To minimize systematic errors during the phylogenetic analyses in paper 2, the amino acid sequences of *gig* orthologs -obtained from 11 aquatic species- were first aligned and then filtered for residue inconsistencies (**Materials and methods in paper 2**). Based on the filtered alignments, best-fit substitution models were implemented in phylogenetic analysis of each *gig* family and bootstrap values were calculated to support the inferred phylogenetic relationships. Filtering of amino acid sequence alignments discarded a sizeable portion of sequence, retaining 31% and 49% of gig1 and gig2 of total amino acid sequence. The low amino acid conservation across alignments of each family confirmed high sequence divergence for these genes as well as high sensitivity regarding the inference of phylogenetic relationships.

In line with these findings, earlier phylogenetic analyses of antiviral type I interferon (IFN-I) genes showed that dissection of the evolutionary relationships is particularly difficult and sensitive to errors (Redmond et al., 2019). Based on the findings of the same study, the observed sensitivity of phylogenetic inference for IFN-I orthologs was attributed -among other factors- to the short length and rapid evolution of these genes. Similarly "messy" evolutionary signals have also been observed across gene families of the innate and the adaptive immune systems, including members of the MHC and Ig families in higher vertebrates and members of the IFN gene families in salmonids (M. Nei et al., 1997; J. Zou et al., 2014). Collectively, these findings showcase the particularly fast evolution of immune genes.

Several models have been suggested in order to characterize the evolution of immune genes, as these genes undergo more dynamic evolution compared to other functional groups -likely due to the genes' role in host-pathogen evolution (Vinkler et al., 2023). Among suggested models, the evolutionary model of gene birth-and-

death was proposed, describing a case where tandem gene duplications increase the number of gene paralogs; some of the duplicated copies diverge functionally and are therefore maintained, whereas others accumulate mutations and ultimately become pseudogenes or completely disappear. As a result of the birth-and-death model, large gene families are formed, sharing complex orthologous relationships across species and including high proportions of pseudogenes (M. Nei et al., 1997; Masatoshi Nei & Rooney, 2005).

Gene evolution shaped by the birth-and-death process -particularly over long term evolutionary processes- can have profound impact on the diversification of multigene families, allowing for rapid adaptation to environmental pressures and pathogenic challenges (Ota & Nei, 1994). Based on the low amino-acid sequence retention rate and the observed phylogenetic relationships of *gig* orthologs in paper 2, the evolution of the *gig* families could likely be explained through a process of gene birth-and-death (**Figure 2 and Supplementary figures 1 and 2 in paper 2**). In corroboration, similar observations were reported in phylogenetic relationships of IFN-I genes across aquatic species (Redmond et al., 2019). Previous studies reported the earliest appearance of *gig2 and gig1* genes in jawless vertebrates and lobe-finned fish, respectively (An et al., 2025; Y.-B. Zhang et al., 2013), allowing birth-and-death evolution to take place over long evolutionary timescales for *gig* genes to expand and functionally diverge. In addition, observed in the case of salmonids and cyprinids, experience of multiple WGD rounds have further facilitated gene duplication, potentially providing the genetic resources required to accelerate the evolution and functional diversification of *gig* paralogs.

Our findings therefore support the notion that the exceptionally fast evolution of immune-related genes makes phylogenetic inference particularly sensitive. Consequently, use of inferred phylogenetic relationships to interpret the functional divergence of large immune-related gene families should be taken with considerable caution.

4.7.4 Complex regulatory relationships observed between segmentally duplicated genes

Analyses in papers 1 and 2 revealed dissimilarities in the transcriptional activity and expression profile of tandemly duplicated paralogs (**Figures 5 and 6 in paper 1 and Figure 6 in paper 2**). Particularly, absence of cis-regulators in some tandemly duplicated genes like gig2A_ssa04b, gig2A_ssa04e and the likely gene

candidate for PD resistance gig1B_ssa03b did not seem to prohibit gene expression and antiviral responses (**Figure 6 in paper 2**). Paper 1 revealed the association of three tandemly duplicated *gig1* genes in pancreas disease resistance, yet the role of each duplicate in antiviral response is not clear. As *gig* genes are involved in severe aquaculture diseases including infectious salmonid anemia virus (ISAV) and salmonid alphavirus (PD) (Aslam et al., 2020; Krasnov et al., 2021), it is important to decipher the relationship between regulatory activity and gene expression patterns of tandemly duplicated paralogs.

As discussed previously, high sequence similarity of duplicated genes can bias gene expression quantification due to the possibility of short mRNA sequencing reads aligning multiple times across duplicated genomic regions. Considering the possibility of including similar artefacts in downstream analyses, in paper 1, *gig1* gene expression was quantified using genotype information to filter mRNA reads that mapped uniquely against the Atlantic salmon transcriptome (**Materials and methods and figure 7 in paper 1**) (van de Geijn et al., 2015). Notably, gene expression analysis using only the uniquely mapping reads revealed that all three *gig1* copies on ssa03 were expressed during viral infection and that expression of the second duplicate gene was genuine - albeit lower than the other two copies. Absence of transcriptional activity in the likely promoter region of the specific *gig1* paralog on Ssa03 (**Figure 6 in paper 1**), combined with the observed response to viral infection, supports a hypothesis of potential exchange or sharing of regulatory mechanisms between adjacent paralogs.

Based on the hypothesis that over-expression of tandemly duplicated *gig* genes could possibly interfere with regulation of IFN response and lead to harmful over-expression (Reyes-López et al., 2015; Robledo et al., 2016), evolution of reduced or shared responses of tandemly duplicated copies could explain the observations in paper 1. In corroboration, a previous study on mammals reported that closely located gene paralogs (> 1Mbp) tended to be co-regulated more often than non-duplicated genes and were also associated with a common regulator more often than expected by chance (Ibn-Salem et al., 2017). Such case of shared regulatory expression could possibly explain the observed activity on the tandemly duplicated *gig1* copies on chromosome Ssa03 but also the observed expression patterns of *gig2* paralogs located on Ssa04 (gig2A_ssa04a- gig2A_ssa04e, **Figure 6 in paper 2**)

However, not all tandemly duplicated genes shared consistently the same expression patterns, suggesting potential exchange of regulatory elements between

adjacent paralogs. Following this hypothesis, a Duplication–Degeneration–Complementation (DCD) evolutionary scenario -where degeneration occurs in the regulatory machinery of gene duplicates in specific tissues or cells- could also provide a possible explanation for the functional diversification of *gig* paralogs (Force et al., 1999). However, the epigenetics information used in analyses of papers 1 and 2 (ChIP-seq and ATAC-seq peaks) were obtained only from head kidney tissue, prohibiting further investigation of tissue/cell-specific regulation of *gig* expression.

While different evolutionary models could potentially fit the functional diversification patterns of tandemly duplicated paralogs, consistently low expression between tissues and experimental methods for certain genes could also be the outcome of pseudogenization and functional redundancy following multiple rounds of whole-genome and tandem duplication. Taken together, deciphering of the regulatory relationships between tandemly duplicated genes is particularly challenging. In the case of gig genes, further study –possibly facilitated by transcriptomics and epigenetics analysis of multiple tissue samples or even targeted gene editing methods- is required in order to clarify the functional specialization of gig paralogs and their potential use in aquaculture disease management strategies.

4.8 Identified gaps for future study

4.8.1 Broader application of long-read sequencing is required for in-depth investigation of immune responses in aquaculture

In the present thesis we utilized long-read sequencing data to elucidate the genomic architecture of pancreas disease (PD) resistance and explore the evolutionary history of two antiviral gene families. Use of long read-based datasets enhanced the genomic and transcriptomic resolution of the PD associated region, uncovering complex genetic variation with putative role in disease resistance. In addition, these data provided functional evidence regarding the structure of *gig* genes and particularly for paralogs overlapping complex genomic (LORe) regions. Collectively, our findings highlight the importance and benefit of using advanced sequencing technologies. However, certain limitations of the datasets used in papers 1 and 2 indicate the need for application of long read sequencing at a larger scale in order to fully explore the functional role of *gig* paralogs in Atlantic salmon as well as their implications in response to aquaculture diseases -including resistance to *Salmonid alphavirus* (SAV3).

Paper 1 utilized information obtained from seven aquaculture individuals to identify complex genetic variation underlying the disease-associated region. Use of a larger sample size of long read sequenced fish could improve the resolution of detected variants and decipher putative genotyping errors from genuine variation - particularly regarding hyper-polymorphic and tandemly repeated variants. In addition, inclusion of more log-read sequenced individuals would allow the detection of genetic linkage between complex genetic variants and SNPs with high association to PD resistance (ssa03_95086730). Although use of long-read sequencing for the analysis of commercial populations at the whole-genome level is not economically feasible yet, application of targeted sequencing on selected, disease-associated, regions using long-read methods could provide a cost-efficient alternative of genotyping large populations (Iyer et al., 2024). Application of the latter approach on large populations of fish that participate in disease challenge trials could allow the inclusion of complex genetic variation in association studies, ultimately enabling the detection of complex variants associated with pathogen resistance.

In addition to long read-based genomics data, papers 1 and 2 in this thesis utilized a long-read sequencing transcriptomics dataset sampled from multiple issues of healthy individuals (Sandholm, 2022). Use of the latter dataset allowed the investigation of *gig* genes' structure; however, it prohibited any analysis related to immune gene responses. Under that assumption, use of long read transcriptomics data from multiple tissues of virally infected individuals in paper 2 could provide solid evidence regarding the presence, structure and expression profile as well as functional divergence of *gig* genes—especially for paralogs complex regions of the salmonid genome (Gundappa et al., 2022; Lien et al., 2016). In addition, overcoming the limitations of short read sequencing technologies -such as read length and multiple alignment errors-, use of long read-based expression data could also increase the accuracy of estimated gene abundance, thus providing valuable insight regarding the expression and functionality of tandem duplicates. Long reads could therefore enhance the transcriptomic resolution and provide accurate estimates of the expression profile and functional divergence of *gig1* and *gig2* genes. Since both *gig* families are involved in antiviral response against several aquaculture diseases (Krasnov et al., 2011; Madushani et al., 2021; C. Sun et al., 2013), thorough investigation of the functional divergence and impact of these genes could prove invaluable for the development of disease management strategies in aquaculture.

Lastly, although analyses in paper 3 focused solely on short-read sequencing data and SNP variants, use of long read data could largely improve the genomic resolution of the WGD-duplicated MHC loci in Atlantic salmon. In particular, many SNPs overlapping MHC-related genes were identified in the duplicated regions of chromosomes ssa14 and ssa27, including non-synonymous mutations. Given the high variability of the MHC region which fuels immunogenetic diversity (Grimholt et al., 2020; Sundaram et al., 2020), use of long read sequencing methods for MHC-targeted genotyping could shed light upon the functional impact of non-synonymous variants overlapping the particular immune-related region. In addition, application of long reads could further explore the genetic variation landscape underlying adaptive immunity beyond the spectrum of SNP mutations.

4.8.2 Functional validation using gene editing approaches can provide insight into the role of immune genes and the impact of putatively causal variation

Use of functional information and primarily *in silico* approaches in this thesis has provided valuable insight regarding the role and functional divergence of duplicated genes, as well as the putative functional impact of genetic variation on gene expression and consequently disease response. Yet, performance of functional validation *in vitro* could provide solid evidence regarding the impact and biological significance of the findings in each study. Taken from the results in paper 2, both *gig* families showed similar immune responses between poly(I:C) infected head kidney cells and tissue, as well as between different tissues (SAV3 infected heart) (**Figure 6 in study 2**). These findings therefore indicate that it is possible to investigate the impact of gig genes using cell lines and obtain results that are highly comparable to those retrieved from *in vivo* viral challenges. Under this assumption, application of gene editing experiments in virally stimulated cell lines provides a high-promising opportunity for the study of *gig1* and *gig2* immune responses (Reza et al., 2025).

For study 1, application of a series of knock-out (KO) experiments in SAV3-stimulated cells or in combination with *gig1* over-expression could provide valuable insight regarding the causality and regulation for each of the three tandemly duplicated *gig1-like* copies overlapping the Ssa03 QTL for PD resistance. In addition to that, functional disruption of the duplicated genes could shed light upon the possibility of regulatory trade-offs between gene copies and provide evidence regarding potential gene dosage compensation between ohnolog genes (Lan &

Pritchard, 2016; Luzuriaga-Neira et al., 2022). Finally, additional experimentation could be performed using knock in (KI) or knock out (KO) approaches in order to test the functional impact of putatively causal genetic variation; particularly the effect of (AC)n repeat length on *gig1-like* gene expression (Lee et al., 2013; Raudstein et al., 2023). The latter experiment could unveil the functional impact of microsatellite repeat expansion on gig1-like gene expression and validate the putative causal impact of the suggested variants on PD response.

For paper 2, a KO experimental approach could also be implemented in order to identify the role of *gig* paralogs within the immune signaling cascade. Performance of such an experiment and disrupting core genes involved in IFN-mediated responses, including IFNs, could validate the dependency of *gig1* and *gig2* genes in the IFN-mediated antiviral response (Reza et al., 2025). Furthermore, taking advantage of the functional domain findings in paper 2 (**Figure 3**), experimentation via immunostaining methods could additionally shed light upon the subcellular localization of *gig* proteins in cell mediated response, elucidating their role within the antiviral signaling cascade (F. Sun et al., 2014).

Lastly, paper 3 identified parallel selection signatures overlapping the two WGD-duplicated MHC class I loci between geographically distinct lineages Atlantic salmon (**Figures 2 and 3 in study 3**). In particular, WGD-duplicated copies of the same MHC genes (ohnolog genes) were under selection in aquaculture populations of North American and West Norwegian origin, namely the genes *daxx, tapbp, brd2a, psmb13a and psmb8(a-b)*. Implementation of gene editing approaches and inactivation of gene duplicates could therefore elucidate the functional diversification of selected gene copies between lineages and additionally reveal possible gene compensation mechanisms between gene duplicates. At the same time, application of these methods under a range of infectious experimental setups (for example strain or lineage -specific pathogens) could potentially provide further insight regarding differential ohnolog selection in response to adaptation of populations to different aquaculture environmental conditions and pathogenic stimuli.

5 References

Alexandrou, M. A., Swartz, B. A., Matzke, N. J., & Oakley, T. H. (2013). Genome duplication and multiple evolutionary origins of complex migratory behavior in Salmonidae. *Molecular Phylogenetics and Evolution*, *69*(3), 514–523. https://doi.org/10.1016/j.ympev.2013.07.026

Allis, C. D., & Jenuwein, T. (2016). The molecular hallmarks of epigenetic control. *Nature Reviews. Genetics*, *17*(8), 487–500. https://doi.org/10.1038/nrg.2016.59

An, L.-L., Gong, X.-Y., Dan, C., Sun, H.-Y., Guo, W.-H., Luan, H.-Y., Wu, M.-Y., Yu, J.-C., & Zhang, Y.-B. (2025). Family evolution and functional divergence of bony fish-specific Gig1 homologs. *Water Biology and Security*, *100382*, 100382. https://doi.org/10.1016/j.watbs.2025.100382

Aoki, T., Hikima, J.-I., Hwang, S. D., & Jung, T. S. (2013). Innate immunity of finfish: primordial conservation and function of viral RNA sensors in teleosts. *Fish & Shellfish Immunology*, *35*(6), 1689–1702. https://doi.org/10.1016/j.fsi.2013.02.005

Aslam, M. L., Robledo, D., Krasnov, A., Moghadam, H. K., Hillestad, B., Houston, R. D., Baranski, M., Boison, S., & Robinson, N. A. (2020). Quantitative trait loci and genes associated with salmonid alphavirus load in Atlantic salmon: implications for pancreas disease resistance and tolerance. *Scientific Reports*, *10*(1), 10393. https://doi.org/10.1038/s41598-020-67405-8

Bai, H., He, Y., Ding, Y., Chu, Q., Lian, L., Heifetz, E. M., Yang, N., Cheng, H. H., Zhang, H., Chen, J., & Song, J. (2020). Genome-wide characterization of copy number variations in the host genome in genetic resistance to Marek's disease using next generation sequencing. *BMC Genetics*, *21*(1), 77. https://doi.org/10.1186/s12863-020-00884-w

Beemelmanns, A., Zanuzzo, F. S., Xue, X., Sandrelli, R. M., Rise, M. L., & Gamperl, A. K. (2021). The transcriptomic responses of Atlantic salmon (Salmo salar) to high temperature stress alone, and in combination with moderate hypoxia. *BMC Genomics*, *22*(1), 261. https://doi.org/10.1186/s12864-021-07464-x

Bertolotti, A. C., Layer, R. M., Gundappa, M. K., Gallagher, M. D., Pehlivanoglu, E., Nome, T., Robledo, D., Kent, M. P., Røsæg, L. L., Holen, M. M., Mulugeta, T. D., Ashton, T. J., Hindar, K., Sægrov, H., Florø-Larsen, B., Erkinaro, J., Primmer, C. R., Bernatchez, L., Martin, S. A. M., … Macqueen, D. J. (2020). The structural variation landscape in 492 Atlantic salmon genomes. *Nature Communications*, *11*(1), 1–16. https://doi.org/10.1038/s41467-020-18972-x

Bianchi, M. E. (2007). DAMPs, PAMPs and alarmins: all we need to know about danger. *Journal of Leukocyte Biology*, *81*(1), 1–5. https://doi.org/10.1189/jlb.0306164

Bjørgen, H., Barac, F., Fjelldal, P. G., Hansen, T., Hordvik, I., & Koppang, E. O. (2024). Organisation of the Atlantic salmon (Salmo salar) thymus and its content of Ig-expressing cells. *Fish & Shellfish Immunology*, *150*(109652), 109652. https://doi.org/10.1016/j.fsi.2024.109652

Boltana, S., Aguilar, A., Sanhueza, N., Donoso, A., Mercado, L., Imarai, M., & Mackenzie, S. (2018). Behavioral fever drives epigenetic modulation of the immune response in fish. *Frontiers in Immunology*, *9*, 1241. https://doi.org/10.3389/fimmu.2018.01241

Bourret, V., Kent, M. P., Primmer, C. R., Vasemägi, A., Karlsson, S., Hindar, K., McGinnity, P., Verspoor, E., Bernatchez, L., & Lien, S. (2013). SNP-array reveals genome-wide patterns of geographical and potential adaptive divergence across the natural range

of Atlantic salmon (Salmo salar). *Molecular Ecology*, *22*(3), 532–551. https://doi.org/10.1111/mec.12003

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods*, *10*(12), 1213–1218. https://doi.org/10.1038/nmeth.2688

Buenrostro, J. D., Wu, B., Chang, H. Y., & Greenleaf, W. J. (2015). ATAC-seq: A method for assaying chromatin accessibility genome-wide. *Et al [Current Protocols in Molecular Biology]*, *109*(1), 21.29.1-21.29.9. https://doi.org/10.1002/0471142727.mb2129s109

Cairns, B. R. (2009). The logic of chromatin architecture and remodelling at promoters. *Nature*, *461*(7261), 193–198. https://doi.org/10.1038/nature08450

Catanach, A., Crowhurst, R., Deng, C., David, C., Bernatchez, L., & Wellenreuther, M. (2019). The genomic pool of standing structural variation outnumbers single nucleotide polymorphism by threefold in the marine teleost Chrysophrys auratus. *Molecular Ecology*, *28*(6), 1210–1223. https://doi.org/10.1111/mec.15051

Ceballos-Concha, A., Asche, F., & Cárdenas-Retamal, R. (2025). Salmon aquaculture in Chile: Production growth and socioeconomic impacts. *Reviews in Aquaculture*, *17*(1). https://doi.org/10.1111/raq.12993

Cechova, M. (2020). Probably correct: Rescuing repeats with short and long reads. *Genes*, *12*(1), 48. https://doi.org/10.3390/genes12010048

Chang, M., Nie, P., Collet, B., Secombes, C. J., & Zou, J. (2009). Identification of an additional two-cysteine containing type I interferon in rainbow trout Oncorhynchus mykiss provides evidence of a major gene duplication event within this gene family in teleosts. *Immunogenetics*, *61*(4), 315–325. https://doi.org/10.1007/s00251-009-0366-y

Chatterjee, S., & Ahituv, N. (2017). Gene regulatory elements, major drivers of human disease. *Annual Review of Genomics and Human Genetics*, *18*, 45–63. https://doi.org/10.1146/annurev-genom-091416-035537

Cheng, C.-Y., Krishnakumar, V., Chan, A. P., Thibaud-Nissen, F., Schobel, S., & Town, C. D. (2017). Araport11: a complete reannotation of the Arabidopsis thaliana reference genome. *The Plant Journal: For Cell and Molecular Biology*, *89*(4), 789–804. https://doi.org/10.1111/tpj.13415

Chisanga, D., Liao, Y., & Shi, W. (2022). Impact of gene annotation choice on the quantification of RNA-seq data. *BMC Bioinformatics*, *23*(1), 107. https://doi.org/10.1186/s12859-022-04644-8

Clark, T. C., Naseer, S., Gundappa, M. K., Laurent, A., Perquis, A., Collet, B., Macqueen, D. J., Martin, S. A. M., & Boudinot, P. (2023). Conserved and divergent arms of the antiviral response in the duplicated genomes of salmonid fishes. *Genomics*, *115*(4), 110663. https://doi.org/10.1016/j.ygeno.2023.110663

Colgan, T. J., Moran, P. A., Archer, L. C., Wynne, R., Hutton, S. A., McGinnity, P., & Reed, T. E. (2021). Evolution and expression of the immune system of a facultatively anadromous Salmonid. *Frontiers in Immunology*, *12*, 568729. https://doi.org/10.3389/fimmu.2021.568729

Collet, B. (2014). Innate immune responses of salmonid fish to viral infections. *Developmental and Comparative Immunology*, *43*(2), 160–173. https://doi.org/10.1016/j.dci.2013.08.017

Cramer, P. (2019). Organization and regulation of gene transcription. *Nature*, *573*(7772), 45–54. https://doi.org/10.1038/s41586-019-1517-4

Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227*(5258), 561–563. https://doi.org/10.1038/227561a0

Cunningham, F., Allen, J. E., Allen, J., Alvarez-Jarreta, J., Amode, M. R., Armean, I. M., Austine-Orimoloye, O., Azov, A. G., Barnes, I., Bennett, R., Berry, A., Bhai, J., Bignell, A., Billis,

K., Boddu, S., Brooks, L., Charkhchi, M., Cummins, C., Da Rin Fioretto, L., … Flicek, P. (2022). Ensembl 2022. *Nucleic Acids Research*, *50*(D1), D988–D995. https://doi.org/10.1093/nar/gkab1049

Dijkstra, J. M., Grimholt, U., Leong, J., Koop, B. F., & Hashimoto, K. (2013). Comprehensive analysis of MHC class II genes in teleost fish genomes reveals dispensability of the peptide-loading DM system in a large part of vertebrates. *BMC Evolutionary Biology*, *13*(1), 260. https://doi.org/10.1186/1471-2148-13-260

Dowall, M. (2001). Anadromy and homing: two life-history traits with adaptive synergies in salmonid fishes? *Fish and Fisheries (Oxford, England)*, *2*(1), 78–85. https://doi.org/10.1046/j.1467-2979.2001.00036.x

Eslamloo, K., Caballero-Solares, A., Inkpen, S. M., Emam, M., Kumar, S., Bouniot, C., Avendaño-Herrera, R., Jakob, E., & Rise, M. L. (2020). Transcriptomic profiling of the adaptive and innate immune responses of Atlantic salmon to Renibacterium salmoninarum infection. *Frontiers in Immunology*, *11*, 567838. https://doi.org/10.3389/fimmu.2020.567838

Flajnik, M. F. (2018). A cold-blooded view of adaptive immunity. *Nature Reviews. Immunology*, *18*(7), 438–453. https://doi.org/10.1038/s41577-018-0003-9

Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y. L., & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, *151*(4), 1531–1545. https://doi.org/10.1093/genetics/151.4.1531

Fraslin, C., Quillet, E., Rochat, T., Dechamp, N., Bernardet, J.-F., Collet, B., Lallias, D., & Boudinot, P. (2020). Combining multiple approaches and models to dissect the genetic architecture of resistance to infections in fish. *Frontiers in Genetics*, *11*, 677. https://doi.org/10.3389/fgene.2020.00677

Gan, Z., Chen, S. N., Huang, B., Zou, J., & Nie, P. (2020). Fish type I and type II interferons: composition, receptor usage, production and function. *Reviews in Aquaculture*, *12*(2), 773–804. https://doi.org/10.1111/raq.12349

Gao, S., Tan, S., Purcell, S. L., Whyte, S. K., Parrish, K., Zhong, L., Zheng, S., Zhang, Y., Zhu, R., Jahangiri, L., Li, R., Fast, M. D., & Cai, W. (2024). A comparative analysis of alternative splicing patterns in Atlantic salmon (Salmo salar) in response to Moritella viscosa and sea lice (Lepeophtheirus salmonis) infection. *Fish & Shellfish Immunology*, *149*(109606), 109606. https://doi.org/10.1016/j.fsi.2024.109606

Gottesfeld, J. M., & Carey, M. F. (2018). Introduction to the Thematic Minireview Series: Chromatin and transcription. *The Journal of Biological Chemistry*, *293*(36), 13775–13777. https://doi.org/10.1074/jbc.TM118.004544

Grimholt, U. (2016). MHC and evolution in teleosts. *Biology*, *5*(1), 6. https://doi.org/10.3390/biology5010006

Grimholt, U., Fosse, J. H., & Sundaram, A. Y. M. (2020). Selective stimulation of duplicated Atlantic salmon MHC pathway genes by interferon-gamma. *Frontiers in Immunology*, *11*, 571650. https://doi.org/10.3389/fimmu.2020.571650

Grimholt, U., Larsen, S., Nordmo, R., Midtlyng, P., Kjoeglum, S., Storset, A., Saebø, S., & Stet, R. J. M. (2003). MHC polymorphism and disease resistance in Atlantic salmon (Salmo salar); facing pathogens with single expressed major histocompatibility class I and class II loci. *Immunogenetics*, *55*(4), 210–219. https://doi.org/10.1007/s00251-003-0567-8

Gundappa, M. K., To, T.-H., Grønvold, L., Martin, S. A. M., Lien, S., Geist, J., Hazlerigg, D., Sandve, S. R., & Macqueen, D. J. (2022). Genome-wide reconstruction of rediploidization following autopolyploidization across one hundred million years of Salmonid evolution. *Molecular Biology and Evolution*, *39*(1). https://doi.org/10.1093/molbev/msab310

Hai, D. M., Yen, D. T., Liem, P. T., Tam, B. M., Huong, D. T. T., Hang, B. T. B., Hieu, D. Q., Garigliany, M.-M., Coppieters, W., Kestemont, P., Phuong, N. T., & Farnir, F. (2022). A high-quality genome assembly of striped catfish (Pangasianodon hypophthalmus)

based on highly accurate long-read HiFi sequencing data. *Genes*, *13*(5), 923. https://doi.org/10.3390/genes13050923
Hämälä, T., Wafula, E. K., Guiltinan, M. J., Ralph, P. E., dePamphilis, C. W., & Tiffin, P. (2021). Genomic structural variants constrain and facilitate adaptation in natural populations of Theobroma cacao, the chocolate tree. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(35), e2102914118. https://doi.org/10.1073/pnas.2102914118
Hamon, M. A., & Cossart, P. (2008). Histone modifications and chromatin remodeling during bacterial infections. *Cell Host & Microbe*, *4*(2), 100–109. https://doi.org/10.1016/j.chom.2008.07.009
Harstad, H., Lukacs, M. F., Bakke, H. G., & Grimholt, U. (2008). Multiple expressed MHC class II loci in salmonids; details of one non-classical region in Atlantic salmon (Salmo salar). *BMC Genomics*, *9*(1), 193. https://doi.org/10.1186/1471-2164-9-193
Hartley, S. E. (1987). The chromosomes of Salmonid fishes. *Biological Reviews of the Cambridge Philosophical Society*, *62*(3), 197–214. https://doi.org/10.1111/j.1469-185x.1987.tb00663.x
Harvey, T. N., Gillard, G. B., Røsæg, L. L., Grammes, F., Monsen, Ø., Vik, J. O., Hvidsten, T. R., & Sandve, S. R. (2024). The genome regulatory landscape of Atlantic salmon liver through smoltification. *PloS One*, *19*(4), e0302388. https://doi.org/10.1371/journal.pone.0302388
Haugarvoll, E., Bjerkås, I., Nowak, B. F., Hordvik, I., & Koppang, E. O. (2008). Identification and characterization of a novel intraepithelial lymphoid tissue in the gills of Atlantic salmon. *Journal of Anatomy*, *213*(2), 202–209. https://doi.org/10.1111/j.1469-7580.2008.00943.x
Haunshi, S., Burramsetty, A. K., Ramasamy, K., & Chatterjee, R. N. (2018). Polymorphisms in pattern recognition receptor genes of indigenous and White Leghorn breeds of chicken. *Archives Animal Breeding*, *61*(4), 441–449. https://doi.org/10.5194/aab-61-441-2018
Hillestad, B., Makvandi-Nejad, S., Krasnov, A., & Moghadam, H. K. (2020). Identification of genetic loci associated with higher resistance to pancreas disease (PD) in Atlantic salmon (Salmo salar L.). *BMC Genomics*, *21*(1), 388. https://doi.org/10.1186/s12864-020-06788-4
Holen, M. M., Vaaje-Kolstad, G., Kent, M. P., & Sandve, S. R. (2023). Gene family expansion and functional diversification of chitinase and chitin synthase genes in Atlantic salmon (Salmo salar). *G3* . https://doi.org/10.1093/g3journal/jkad069
Houston, R. D., & Macqueen, D. J. (2019). Atlantic salmon (Salmo salar L.) genetics in the 21st century: taking leaps forward in aquaculture and biological understanding. *Animal Genetics*, *50*(1), 3–14. https://doi.org/10.1111/age.12748
Houston, Ross D., Bean, T. P., Macqueen, D. J., Gundappa, M. K., Jin, Y. H., Jenkins, T. L., Selly, S. L. C., Martin, S. A. M., Stevens, J. R., Santos, E. M., Davie, A., & Robledo, D. (2020). Harnessing genomics to fast-track genetic improvement in aquaculture. *Nature Reviews. Genetics*, *21*(7), 389–409. https://doi.org/10.1038/s41576-020-0227-y
Ibn-Salem, J., Muro, E. M., & Andrade-Navarro, M. A. (2017). Co-regulation of paralog genes in the three-dimensional chromatin architecture. *Nucleic Acids Research*, *45*(1), 81–91. https://doi.org/10.1093/nar/gkw813
Ivashkiv, L. B., & Donlin, L. T. (2014). Regulation of type I interferon responses. *Nature Reviews. Immunology*, *14*(1), 36–49. https://doi.org/10.1038/nri3581
Iversen, A., Asche, F., Hermansen, Ø., & Nystøyl, R. (2020). Production cost and competitiveness in major salmon farming countries 2003–2018. *Aquaculture (Amsterdam, Netherlands)*, *522*(735089), 735089. https://doi.org/10.1016/j.aquaculture.2020.735089
Iversen, M., Mulugeta, T., West, A. C., Jørgensen, E. H., Martin, S. A. M., Sandve, S. R., & Hazlerigg, D. (2021). Photoperiod-dependent developmental reprogramming of the

transcriptional response to seawater entry in Atlantic salmon (Salmo salar). *G3* , *11*(4). https://doi.org/10.1093/g3journal/jkab072
Iyer, S. V., Goodwin, S., & McCombie, W. R. (2024). Leveraging the power of long reads for targeted sequencing. *Genome Research*, *34*(11), 1701–1718. https://doi.org/10.1101/gr.279168.124
Jobson, E., & Roberts, R. (2022). Genomic structural variation in tomato and its role in plant immunity. *Molecular Horticulture*, *2*(1), 7. https://doi.org/10.1186/s43897-022-00029-w
Johnston, I. A., Kent, M. P., Boudinot, P., Looseley, M., Bargelloni, L., Faggion, S., Merino, G. A., Ilsley, G. R., Bobe, J., Tsigenopoulos, C. S., Robertson, J., Harrison, P. W., Martinez, P., Robledo, D., Macqueen, D. J., & Lien, S. (2024). Advancing fish breeding in aquaculture through genome functional annotation. *Aquaculture (Amsterdam, Netherlands)*, *583*(740589), 740589. https://doi.org/10.1016/j.aquaculture.2024.740589
Jupe, F., Witek, K., Verweij, W., Sliwka, J., Pritchard, L., Etherington, G. J., Maclean, D., Cock, P. J., Leggett, R. M., Bryan, G. J., Cardle, L., Hein, I., & Jones, J. D. G. (2013). Resistance gene enrichment sequencing (RenSeq) enables reannotation of the NB-LRR gene family from sequenced plant genomes and rapid mapping of resistance loci in segregating populations. *The Plant Journal: For Cell and Molecular Biology*, *76*(3), 530–544. https://doi.org/10.1111/tpj.12307
Kim, D.-H., Joo, M.-S., Woo, S.-J., Son, K.-T., Hong, W.-S., Park, M.-C., Park, J.-C., & Park, S.-O. (2024). The First Report on Saprolegnia parasitica and Neoparamoeba perurans Isolated from Atlantic Salmon (Salmo salar) Reared in Korea. *Microbiology Research*, *15*(2), 1016–1027. https://doi.org/10.3390/microbiolres15020067
Kjøglum, S., Larsen, S., Bakke, H. G., & Grimholt, U. (2006). How specific MHC class I and class II combinations affect disease resistance against infectious salmon anaemia in Atlantic salmon (Salmo salar). *Fish & Shellfish Immunology*, *21*(4), 431–441. https://doi.org/10.1016/j.fsi.2006.02.001
Koch, I. J., Nuetzel, H. M., & Narum, S. R. (2023). Epigenetic effects associated with salmonid supplementation and domestication. *Environmental Biology of Fishes*, *106*(5), 1093–1111. https://doi.org/10.1007/s10641-022-01278-w
Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, *128*(4), 693–705. https://doi.org/10.1016/j.cell.2007.02.005
Krasnov, A., Johansen, L.-H., Karlsen, C., Sveen, L., Ytteborg, E., Timmerhaus, G., Lazado, C. C., & Afanasyev, S. (2021). Transcriptome responses of Atlantic salmon (Salmo salar L.) to viral and bacterial pathogens, inflammation, and stress. *Frontiers in Immunology*, *12*, 705601. https://doi.org/10.3389/fimmu.2021.705601
Krasnov, A., Timmerhaus, G., Schiøtz, B. L., Torgersen, J., Afanasyev, S., Iliev, D., Jørgensen, J., Takle, H., & Jørgensen, S. M. (2011). Genomic survey of early responses to viruses in Atlantic salmon, Salmo salar L. *Molecular Immunology*, *49*(1–2), 163–174. https://doi.org/10.1016/j.molimm.2011.08.007
Ku, C. S., Naidoo, N., Wu, M., & Soong, R. (2011). Studying the epigenome using next generation sequencing. *Journal of Medical Genetics*, *48*(11), 721–730. https://doi.org/10.1136/jmedgenet-2011-100242
Kuehner, J. N., Pearson, E. L., & Moore, C. (2011). Unravelling the means to an end: RNA polymerase II transcription termination. *Nature Reviews. Molecular Cell Biology*, *12*(5), 283–294. https://doi.org/10.1038/nrm3098
Lafferty, K. D., Harvell, C. D., Conrad, J. M., Friedman, C. S., Kent, M. L., Kuris, A. M., Powell, E. N., Rondeau, D., & Saksida, S. M. (2015). Infectious diseases affect marine fisheries and aquaculture economics. *Annual Review of Marine Science*, *7*(1), 471–496. https://doi.org/10.1146/annurev-marine-010814-015646

Lan, X., & Pritchard, J. K. (2016). Coregulation of tandem duplicate genes slows evolution of subfunctionalization in mammals. *Science (New York, N.Y.)*, *352*(6288), 1009–1013. https://doi.org/10.1126/science.aad8411

Lecaudey, L. A., Schliewen, U. K., Osinov, A. G., Taylor, E. B., Bernatchez, L., & Weiss, S. J. (2018). Inferring phylogenetic structure, hybridization and divergence times within Salmoninae (Teleostei: Salmonidae) using RAD-sequencing. *Molecular Phylogenetics and Evolution*, *124*, 82–99. https://doi.org/10.1016/j.ympev.2018.02.022

Lee, J.-M., Galkina, E. I., Levantovsky, R. M., Fossale, E., Anne Anderson, M., Gillis, T., Srinidhi Mysore, J., Coser, K. R., Shioda, T., Zhang, B., Furia, M. D., Derry, J., Kohane, I. S., Seong, I. S., Wheeler, V. C., Gusella, J. F., & MacDonald, M. E. (2013). Dominant effects of the Huntington's disease HTT CAG repeat length are captured in gene-expression data sets by a continuous analysis mathematical modeling strategy. *Human Molecular Genetics*, *22*(16), 3227–3238. https://doi.org/10.1093/hmg/ddt176

Leith, P., Ogier, E., & Haward, M. (2014). Science and social license: Defining environmental sustainability of Atlantic salmon aquaculture in south-eastern Tasmania, Australia. *Social Epistemology*, *28*(3–4), 277–296. https://doi.org/10.1080/02691728.2014.922641

Leviyang, S. (2021). Interferon stimulated binding of ISRE is cell type specific and is predicted by homeostatic chromatin state. *Cytokine: X*, *3*(4), 100056. https://doi.org/10.1016/j.cytox.2021.100056

Li, X.-Y., Thomas, S., Sabo, P. J., Eisen, M. B., Stamatoyannopoulos, J. A., & Biggin, M. D. (2011). The role of chromatin accessibility in directing the widespread, overlapping patterns of Drosophila transcription factor binding. *Genome Biology*, *12*(4), R34. https://doi.org/10.1186/gb-2011-12-4-r34

Li, Y. (2021). Modern epigenetics methods in biological research. *Methods (San Diego, Calif.)*, *187*, 104–113. https://doi.org/10.1016/j.ymeth.2020.06.022

Lien, S., Koop, B. F., Sandve, S. R., Miller, J. R., Kent, M. P., Nome, T., Hvidsten, T. R., Leong, J. S., Minkley, D. R., Zimin, A., Grammes, F., Grove, H., Gjuvsland, A., Walenz, B., Hermansen, R. A., von Schalburg, K., Rondeau, E. B., Di Genova, A., Samy, J. K. A., … Davidson, W. S. (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature*, *533*(7602), 200–205. https://doi.org/10.1038/nature17164

Liu, Z., Zhou, T., & Gao, D. (2022). Genetic and epigenetic regulation of growth, reproduction, disease resistance and stress responses in aquaculture. *Frontiers in Genetics*, *13*, 994471. https://doi.org/10.3389/fgene.2022.994471

Lukacs, M. F., Harstad, H., Grimholt, U., Beetz-Sargent, M., Cooper, G. A., Reid, L., Bakke, H. G., Phillips, R. B., Miller, K. M., Davidson, W. S., & Koop, B. F. (2007). Genomic organization of duplicated major histocompatibility complex class I regions in Atlantic salmon (Salmo salar). *BMC Genomics*, *8*(1), 251. https://doi.org/10.1186/1471-2164-8-251

Lulijwa, R., Alfaro, A. C., Merien, F., Meyer, J., & Young, T. (2019). Advances in salmonid fish immunology: A review of methods and techniques for lymphoid tissue and peripheral blood leucocyte isolation and application. *Fish & Shellfish Immunology*, *95*, 44–80. https://doi.org/10.1016/j.fsi.2019.10.006

Luo, L., Gribskov, M., & Wang, S. (2022). Bibliometric review of ATAC-Seq and its application in gene expression. *Briefings in Bioinformatics*, *23*(3). https://doi.org/10.1093/bib/bbac061

Lutfalla, G., Roest Crollius, H., Stange-Thomann, N., Jaillon, O., Mogensen, K., & Monneron, D. (2003). Comparative genomic analysis reveals independent expansion of a lineage-specific gene family in vertebrates: the class II cytokine receptors and their ligands in mammals and fish. *BMC Genomics*, *4*(1), 29. https://doi.org/10.1186/1471-2164-4-29

Luzuriaga-Neira, A., Subramanian, K., & Alvarez-Ponce, D. (2022). Functional compensation of mouse duplicates by their paralogs expressed in the same tissues. *Genome Biology and Evolution*, *14*(8). https://doi.org/10.1093/gbe/evac126

Macqueen, D. J., & Johnston, I. A. (2014). A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. *Proceedings of the Royal Society B: Biological Sciences*, *281*(1778), 20132881. https://doi.org/10.1098/rspb.2013.2881

Madushani, K. P., Shanaka, K. A. S. N., Madusanka, R. K., & Lee, J. (2021). Molecular characterization and expressional analysis of two poly (ADP-ribose) polymerase (PARP) domain-containing Gig2 isoforms in rockfish (Sebastes schlegelii) and their antiviral activity against viral hemorrhagic septicemia virus. *Fish & Shellfish Immunology*, *118*, 219–227. https://doi.org/10.1016/j.fsi.2021.09.007

Magnadóttir, B. (2006). Innate immunity of fish (overview). *Fish & Shellfish Immunology*, *20*(2), 137–151. https://doi.org/10.1016/j.fsi.2004.09.006

Magnadottir, B. (2010). Immunological control of fish diseases. *Marine Biotechnology (New York, N.Y.)*, *12*(4), 361–379. https://doi.org/10.1007/s10126-010-9279-x

Mahmoud, M., Gobet, N., Cruz-Dávalos, D. I., Mounier, N., Dessimoz, C., & Sedlazeck, F. J. (2019). Structural variant calling: the long and the short of it. *Genome Biology*, *20*(1), 246. https://doi.org/10.1186/s13059-019-1828-7

Manry, J., Laval, G., Patin, E., Fornarino, S., Itan, Y., Fumagalli, M., Sironi, M., Tichit, M., Bouchier, C., Casanova, J.-L., Barreiro, L. B., & Quintana-Murci, L. (2011). Evolutionary genetic dissection of human interferons. *The Journal of Experimental Medicine*, *208*(13), 2747–2759. https://doi.org/10.1084/jem.20111680

Medina-Gali, R., Belló-Pérez, M., Martínez-López, A., Falcó, A., Ortega-Villaizan, M. M., Encinar, J. A., Novoa, B., Coll, J., & Perez, L. (2018). Chromatin immunoprecipitation and high throughput sequencing of SVCV-infected zebrafish reveals novel epigenetic histone methylation patterns involved in antiviral immune response. *Fish & Shellfish Immunology*, *82*, 514–521. https://doi.org/10.1016/j.fsi.2018.08.056

Millán-Zambrano, G., Burton, A., Bannister, A. J., & Schneider, R. (2022). Histone post-translational modifications - cause and consequence of genome function. *Nature Reviews. Genetics*, *23*(9), 563–580. https://doi.org/10.1038/s41576-022-00468-7

Minafra, A. R., Rafii, P., Mossner, S., Bazgir, F., Floss, D. M., Moll, J. M., & Scheller, J. (2023). Synthetic receptor platform to identify loss-of-function single nucleotide variants and designed mutants in the death receptor Fas/CD95. *The Journal of Biological Chemistry*, *299*(8), 104989. https://doi.org/10.1016/j.jbc.2023.104989

Moen, T., Baranski, M., Sonesson, A. K., & Kjøglum, S. (2009). Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar): population-level associations between markers and trait. *BMC Genomics*, *10*, 368. https://doi.org/10.1186/1471-2164-10-368

Moen, T., Torgersen, J., Santi, N., Davidson, W. S., Baranski, M., Ødegård, J., Kjøglum, S., Velle, B., Kent, M., Lubieniecki, K. P., Isdal, E., & Lien, S. (2015). Epithelial cadherin determines resistance to infectious pancreatic necrosis virus in Atlantic salmon. *Genetics*, *200*(4), 1313–1326. https://doi.org/10.1534/genetics.115.175406

Mohamed, A. R., Naval-Sanchez, M., Menzies, M., Evans, B., King, H., Reverter, A., & Kijas, J. W. (2022). Leveraging transcriptome and epigenome landscapes to infer regulatory networks during the onset of sexual maturation. *BMC Genomics*, *23*(1), 413. https://doi.org/10.1186/s12864-022-08514-8

Mokhtar, D., Zaccone, G., Alesci, A., Kuciel, M., Hussein, M., & Sayed, R. (2023). Main components of fish immunity: An overview of the fish immune system. *Fishes*, *8*(2), 93. https://doi.org/10.3390/fishes8020093

Mondal, H., & Thomas, J. (2022). A review on the recent advances and application of vaccines against fish pathogens in aquaculture. *Aquaculture International: Journal of the*

European Aquaculture Society, *30*(4), 1971–2000. https://doi.org/10.1007/s10499-022-00884-w

Monsen, Ø. (2022). *The contribution of repetitive elements to salmonid genome evolution* [Norwegian University of Life Sciences, Ås]. https://hdl.handle.net/11250/3047268

Mordecai, G. J., Miller, K. M., Bass, A. L., Bateman, A. W., Teffer, A. K., Caleta, J. M., Di Cicco, E., Schulze, A. D., Kaukinen, K. H., Li, S., Tabata, A., Jones, B. R., Ming, T. J., & Joy, J. B. (2021). Aquaculture mediates global transmission of a viral pathogen to wild salmon. *Science Advances*, *7*(22), eabe2592. https://doi.org/10.1126/sciadv.abe2592

Morris, R., Kershaw, N. J., & Babon, J. J. (2018). The molecular details of cytokine signaling via the JAK/STAT pathway. *Protein Science: A Publication of the Protein Society*, *27*(12), 1984–2009. https://doi.org/10.1002/pro.3519

Mukiibi, R., Ferraresso, S., Franch, R., Peruzza, L., Dalla Rovere, G., Babbucci, M., Bertotto, D., Toffan, A., Pascoli, F., Faggion, S., Peñaloza, C., Tsigenopoulos, C. S., Houston, R. D., Bargelloni, L., & Robledo, D. (2025). Integrated functional genomic analysis identifies regulatory variants underlying a major QTL for disease resistance in European sea bass. *BMC Biology*, *23*(1), 75. https://doi.org/10.1186/s12915-025-02180-4

Nakanishi, T., Shibasaki, Y., & Matsuura, Y. (2015). T cells in fish. *Biology*, *4*(4), 640–663. https://doi.org/10.3390/biology4040640

Nakato, R., & Sakata, T. (2021). Methods for ChIP-seq analysis: A practical workflow and advanced applications. *Methods (San Diego, Calif.)*, *187*, 44–53. https://doi.org/10.1016/j.ymeth.2020.03.005

Nei, M., Gu, X., & Sitnikova, T. (1997). Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences of the United States of America*, *94*(15), 7799–7806. https://doi.org/10.1073/pnas.94.15.7799

Nei, Masatoshi, & Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families. *Annual Review of Genetics*, *39*(1), 121–152. https://doi.org/10.1146/annurev.genet.39.073003.112240

Noreen, M., Shah, M. A. A., Mall, S. M., Choudhary, S., Hussain, T., Ahmed, I., Jalil, S. F., & Raza, M. I. (2012). TLR4 polymorphisms and disease susceptibility. *Et al [Inflammation Research]*, *61*(3), 177–188. https://doi.org/10.1007/s00011-011-0427-1

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., … Pruitt, K. D. (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733-45. https://doi.org/10.1093/nar/gkv1189

Ortega-Villaizan, M. D. M., Chico, V., & Perez, L. (2022). Fish innate immune response to viral infection-an overview of five major antiviral genes. *Viruses*, *14*(7), 1546. https://doi.org/10.3390/v14071546

Ota, T., & Nei, M. (1994). Divergent evolution and evolution by the birth-and-death process in the immunoglobulin VH gene family. *Molecular Biology and Evolution*, *11*(3), 469–482. https://doi.org/10.1093/oxfordjournals.molbev.a040127

Park, P. J. (2009). ChIP-seq: advantages and challenges of a maturing technology. *Nature Reviews. Genetics*, *10*(10), 669–680. https://doi.org/10.1038/nrg2641

Pestka, S., Krause, C. D., Sarkar, D., Walter, M. R., Shi, Y., & Fisher, P. B. (2004). Interleukin-10 and related cytokines and receptors. *Annual Review of Immunology*, *22*(1), 929–979. https://doi.org/10.1146/annurev.immunol.22.012703.104622

Peter, I. S., & Davidson, E. H. (2017). Assessing regulatory information in developmental gene regulatory networks. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(23), 5862–5869. https://doi.org/10.1073/pnas.1610616114

Peterson, C. L., & Laniel, M.-A. (2004). Histones and histone modifications. *Current Biology: CB*, *14*(14), R546-51. https://doi.org/10.1016/j.cub.2004.07.007
Phan, V., Gao, S., Tran, Q., & Vo, N. S. (2015). How genome complexity can explain the difficulty of aligning reads to genomes. *BMC Bioinformatics*, *16 Suppl 17*(S17), S3. https://doi.org/10.1186/1471-2105-16-S17-S3
Poliseno, L., Lanza, M., & Pandolfi, P. P. (2024). Coding, or non-coding, that is the question. *Cell Research*, *34*(9), 609–629. https://doi.org/10.1038/s41422-024-00975-8
Portin, P., & Wilkins, A. (2017). The evolving definition of the term "gene." *Genetics*, *205*(4), 1353–1364. https://doi.org/10.1534/genetics.116.196956
Poynter, S. J., & DeWitte-Orr, S. J. (2016). Fish interferon-stimulated genes: The antiviral effectors. *Developmental and Comparative Immunology*, *65*, 218–225. https://doi.org/10.1016/j.dci.2016.07.011
Raudstein, M., Kjærner-Semb, E., Barvik, M., Broll, S., Straume, A. H., & Edvardsen, R. B. (2023). In vivo CRISPR/LbCas12a-mediated knock-in and knock-out in Atlantic salmon (Salmo salar L.). *Transgenic Research*, *32*(6), 513–521. https://doi.org/10.1007/s11248-023-00368-4
Redmond, A. K., Zou, J., Secombes, C. J., Macqueen, D. J., & Dooley, H. (2019). Discovery of all three types in cartilaginous fishes enables phylogenetic resolution of the origins and evolution of interferons. *Frontiers in Immunology*, *10*, 1558. https://doi.org/10.3389/fimmu.2019.01558
Reyes-López, F. E., Romeo, J. S., Vallejos-Vidal, E., Reyes-Cerpa, S., Sandino, A. M., Tort, L., Mackenzie, S., & Imarai, M. (2015). Differential immune gene expression profiles in susceptible and resistant full-sibling families of Atlantic salmon (Salmo salar) challenged with infectious pancreatic necrosis virus (IPNV). *Developmental and Comparative Immunology*, *53*(1), 210–221. https://doi.org/10.1016/j.dci.2015.06.017
Reza, M. A. N., Harvey, T. N., Ahmed Abdelrahim Gamil, A., Evensen, Ø., Gillard, G. B., & Sandvik, G. K. (2025). Investigating antiviral pathways in Atlantic salmon cells through interferon receptor knockouts via CRISPR-Cas9. In *bioRxiv*. https://doi.org/10.1101/2025.01.14.632998
Rikardsen, A. H., Righton, D., Strøm, J. F., Thorstad, E. B., Gargan, P., Sheehan, T., Økland, F., Chittenden, C. M., Hedger, R. D., Næsje, T. F., Renkawitz, M., Sturlaugsson, J., Caballero, P., Baktoft, H., Davidsen, J. G., Halttunen, E., Wright, S., Finstad, B., & Aarestrup, K. (2021). Redefining the oceanic distribution of Atlantic salmon. *Scientific Reports*, *11*(1), 12266. https://doi.org/10.1038/s41598-021-91137-y
Robertsen, B. (2006). The interferon system of teleost fish. *Fish & Shellfish Immunology*, *20*(2), 172–191. https://doi.org/10.1016/j.fsi.2005.01.010
Robertsen, B. (2008). Expression of interferon and interferon-induced genes in salmonids in response to virus infection, interferon-inducing compounds and vaccination. *Fish & Shellfish Immunology*, *25*(4), 351–357. https://doi.org/10.1016/j.fsi.2008.02.004
Robertsen, B. (2018). The role of type I interferons in innate and adaptive immunity against viruses in Atlantic salmon. *Developmental and Comparative Immunology*, *80*, 41–52. https://doi.org/10.1016/j.dci.2017.02.005
Robertsen, B., & Greiner-Tollersrud, L. (2024). Atlantic salmon type I interferon genes revisited. *Fish & Shellfish Immunology*, *151*(109694), 109694. https://doi.org/10.1016/j.fsi.2024.109694
Robertson, F. M., Gundappa, M. K., Grammes, F., Hvidsten, T. R., Redmond, A. K., Lien, S., Martin, S. A. M., Holland, P. W. H., Sandve, S. R., & Macqueen, D. J. (2017). Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biology*, *18*(1), 111. https://doi.org/10.1186/s13059-017-1241-z
Robledo, D., Gutiérrez, A. P., Barría, A., Yáñez, J. M., & Houston, R. D. (2018). Gene expression response to sea lice in Atlantic salmon skin: RNA sequencing comparison between

resistant and susceptible animals. *Frontiers in Genetics*, *9*, 287. https://doi.org/10.3389/fgene.2018.00287

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., & Houston, R. (2018). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in Aquaculture*, *10*(3), 670–682. https://doi.org/10.1111/raq.12193

Robledo, D., Taggart, J. B., Ireland, J. H., McAndrew, B. J., Starkey, W. G., Haley, C. S., Hamilton, A., Guy, D. R., Mota-Velasco, J. C., Gheyas, A. A., Tinch, A. E., Verner-Jeffreys, D. W., Paley, R. K., Rimmer, G. S. E., Tew, I. J., Bishop, S. C., Bron, J. E., & Houston, R. D. (2016). Gene expression comparison of resistant and susceptible Atlantic salmon fry challenged with Infectious Pancreatic Necrosis virus reveals a marked contrast in immune response. *BMC Genomics*, *17*(1), 279. https://doi.org/10.1186/s12864-016-2600-y

Rojano, E., Seoane, P., Ranea, J. A. G., & Perkins, J. R. (2019). Regulatory variants: from detection to predicting impact. *Briefings in Bioinformatics*, *20*(5), 1639–1654. https://doi.org/10.1093/bib/bby039

Rombout, J. H. W. M., Huttenhuis, H. B. T., Picchietti, S., & Scapigliati, G. (2005). Phylogeny and ontogeny of fish leucocytes. *Fish & Shellfish Immunology*, *19*(5), 441–455. https://doi.org/10.1016/j.fsi.2005.03.007

Rougemont, Q., & Bernatchez, L. (2018). The demographic history of Atlantic salmon (Salmo salar) across its distribution range reconstructed from approximate Bayesian computations. *Evolution; International Journal of Organic Evolution*, *72*(6), 1261–1277. https://doi.org/10.1111/evo.13486

Sakai, M., Hikima, J.-I., & Kono, T. (2021). Fish cytokines: current research and applications. *Fisheries Science: FS*, *87*(1), 1–9. https://doi.org/10.1007/s12562-020-01476-4

Sandholm, R. M. (2022). *Evolution of alternative splice variation and exon usage following whole genome duplication* [Norwegian University of Life Sciences, Ås]. https://nmbu.brage.unit.no/nmbu-xmlui/handle/11250/3026606

Sandve, S. R., Rohlfs, R. V., & Hvidsten, T. R. (2018). Subfunctionalization versus neofunctionalization after whole-genome duplication. *Nature Genetics*, *50*(7), 908–909. https://doi.org/10.1038/s41588-018-0162-4

Schneider, W. M., Chevillotte, M. D., & Rice, C. M. (2014). Interferon-stimulated genes: a complex web of host defenses. *Annual Review of Immunology*, *32*(1), 513–545. https://doi.org/10.1146/annurev-immunol-032713-120231

Schoggins, J. W., Wilson, S. J., Panis, M., Murphy, M. Y., Jones, C. T., Bieniasz, P., & Rice, C. M. (2011). A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature*, *472*(7344), 481–485. https://doi.org/10.1038/nature09907

Secombes, C. J., Wang, T., Hong, S., Peddie, S., Crampe, M., Laing, K. J., Cunningham, C., & Zou, J. (2001). Cytokines and innate immunity of fish. *Developmental and Comparative Immunology*, *25*(8–9), 713–723. https://doi.org/10.1016/s0145-305x(01)00032-5

Secombes, Chris J., & Zou, J. (2017). Evolution of interferons and interferon receptors. *Frontiers in Immunology*, *8*, 209. https://doi.org/10.3389/fimmu.2017.00209

Smith, N. C., Rise, M. L., & Christian, S. L. (2019). A comparison of the innate and adaptive immune systems in cartilaginous fish, ray-finned fish, and lobe-finned fish. *Frontiers in Immunology*, *10*, 2292. https://doi.org/10.3389/fimmu.2019.02292

Sobhkhez, M., Skjesol, A., Thomassen, E., Tollersrud, L. G., Iliev, D. B., Sun, B., Robertsen, B., & Jørgensen, J. B. (2014). Structural and functional characterization of salmon STAT1, STAT2 and IRF9 homologs sheds light on interferon signaling in teleosts. *FEBS Open Bio*, *4*(1), 858–871. https://doi.org/10.1016/j.fob.2014.09.007

Soliman, A. M., & Barreda, D. R. (2023). The acute inflammatory response of teleost fish. *Developmental and Comparative Immunology*, *146*(104731), 104731. https://doi.org/10.1016/j.dci.2023.104731

Stenløkk, K. S. R. (2023). *Genomic structural variations as drivers of adaptation in salmonid fishes* [Norwegian University of Life Sciences, Ås]. https://hdl.handle.net/11250/3093077

Sun, B., Greiner-Tollersrud, L., Koop, B. F., & Robertsen, B. (2014). Atlantic salmon possesses two clusters of type I interferon receptor genes on different chromosomes, which allows for a larger repertoire of interferon receptors than in zebrafish and mammals. *Developmental and Comparative Immunology*, *47*(2), 275–286. https://doi.org/10.1016/j.dci.2014.08.007

Sun, C., Liu, Y., Hu, Y., Fan, Q., Li, W., Yu, X., Mao, H., & Hu, C. (2013). Gig1 and Gig2 homologs (CiGig1 and CiGig2) from grass carp (Ctenopharyngodon idella) display good antiviral activities in an IFN-independent pathway. *Developmental and Comparative Immunology*, *41*(4), 477–483. https://doi.org/10.1016/j.dci.2013.07.007

Sun, F., Zhang, Y.-B., Jiang, J., Wang, B., Chen, C., Zhang, J., & Gui, J.-F. (2014). Gig1, a novel antiviral effector involved in fish interferon response. *Virology*, *448*, 322–332. https://doi.org/10.1016/j.virol.2013.10.029

Sundaram, A. Y. M., Garseth, Å. H., Maccari, G., & Grimholt, U. (2020). An Illumina approach to MHC typing of Atlantic salmon. *Immunogenetics*, *72*(1–2), 89–100. https://doi.org/10.1007/s00251-019-01143-8

Sveen, L. R., Grammes, F. T., Ytteborg, E., Takle, H., & Jørgensen, S. M. (2017). Genome-wide analysis of Atlantic salmon (Salmo salar) mucin genes and their role as biomarkers. *PloS One*, *12*(12), e0189103. https://doi.org/10.1371/journal.pone.0189103

Takeuchi, O., & Akira, S. (2009). Innate immunity to virus infection. *Immunological Reviews*, *227*(1), 75–86. https://doi.org/10.1111/j.1600-065X.2008.00737.x

Talbert, P. B., & Henikoff, S. (2021). The yin and yang of histone marks in transcription. *Annual Review of Genomics and Human Genetics*, *22*(1), 147–170. https://doi.org/10.1146/annurev-genom-120220-085159

Thorarinsson, R., Wolf, J. C., Inami, M., Phillips, L., Jones, G., Macdonald, A. M., Rodriguez, J. F., Sindre, H., Skjerve, E., Rimstad, E., & Evensen, Ø. (2021). Effect of a novel DNA vaccine against pancreas disease caused by salmonid alphavirus subtype 3 in Atlantic salmon (Salmo salar). *Fish & Shellfish Immunology*, *108*, 116–126. https://doi.org/10.1016/j.fsi.2020.12.002

Uren Webster, T. M., Rodriguez-Barreto, D., Martin, S. A. M., Van Oosterhout, C., Orozco-terWengel, P., Cable, J., Hamilton, A., Garcia De Leaniz, C., & Consuegra, S. (2018). Contrasting effects of acute and chronic stress on the transcriptome, epigenome, and immune response of Atlantic salmon. *Epigenetics: Official Journal of the DNA Methylation Society*, *13*(12), 1191–1207. https://doi.org/10.1080/15592294.2018.1554520

Vaattovaara, A., Leppälä, J., Salojärvi, J., & Wrzaczek, M. (2019). High-throughput sequencing data and the impact of plant gene annotation quality. *Journal of Experimental Botany*, *70*(4), 1069–1076. https://doi.org/10.1093/jxb/ery434

Valenzuela-Miranda, D., Boltaña, S., Cabrejos, M. E., Yáñez, J. M., & Gallardo-Escárate, C. (2015). High-throughput transcriptome analysis of ISAV-infected Atlantic salmon Salmo salar unravels divergent immune responses associated to head-kidney, liver and gills tissues. *Fish & Shellfish Immunology*, *45*(2), 367–377. https://doi.org/10.1016/j.fsi.2015.04.003

van de Geijn, B., McVicker, G., Gilad, Y., & Pritchard, J. K. (2015). WASP: allele-specific software for robust molecular quantitative trait locus discovery. *Nature Methods*, *12*(11), 1061–1063. https://doi.org/10.1038/nmeth.3582

Verrier, E. R., Dorson, M., Mauger, S., Torhy, C., Ciobotaru, C., Hervet, C., Dechamp, N., Genet, C., Boudinot, P., & Quillet, E. (2013). Resistance to a rhabdovirus (VHSV) in rainbow trout: identification of a major QTL related to innate mechanisms. *PloS One*, *8*(2), e55302. https://doi.org/10.1371/journal.pone.0055302

Verrier, E. R., Langevin, C., Benmansour, A., & Boudinot, P. (2011). Early antiviral response and virus-induced genes in fish. *Developmental and Comparative Immunology*, *35*(12), 1204–1214. https://doi.org/10.1016/j.dci.2011.03.012

Vinkler, M., Fiddaman, S. R., Těšický, M., O'Connor, E. A., Savage, A. E., Lenz, T. L., Smith, A. L., Kaufman, J., Bolnick, D. I., Davies, C. S., Dedić, N., Flies, A. S., Samblás, M. M. G., Henschen, A. E., Novák, K., Palomar, G., Raven, N., Samaké, K., Slade, J., … Westerdahl, H. (2023). Understanding the evolution of immune genes in jawed vertebrates. *Journal of Evolutionary Biology*, *36*(6), 847–873. https://doi.org/10.1111/jeb.14181

Wang, J., Yang, M., Ali, O., Dragland, J. S., Bjørås, M., & Farkas, L. (2024). Predicting regulatory mutations and their target genes by new computational integrative analysis: A study of follicular lymphoma. *Computers in Biology and Medicine*, *178*(108787), 108787. https://doi.org/10.1016/j.compbiomed.2024.108787

Wang, T., & Secombes, C. J. (2013). The cytokine networks of adaptive immunity in fish. *Fish & Shellfish Immunology*, *35*(6), 1703–1718. https://doi.org/10.1016/j.fsi.2013.08.030

Wang, X., Xia, H., Liu, S., Cao, L., & You, F. (2021). Epigenetic regulation in antiviral innate immunity. *European Journal of Immunology*, *51*(7), 1641–1651. https://doi.org/10.1002/eji.202048975

Wang, Y., Xu, X., Zhang, A., Yang, S., & Li, H. (2024). Role of alternative splicing in fish immunity. *Fish & Shellfish Immunology*, *149*(109601), 109601. https://doi.org/10.1016/j.fsi.2024.109601

Wang, Z., Xu, J., Feng, J., Wu, K., Chen, K., Jia, Z., Zhu, X., Huang, W., Zhao, X., Liu, Q., Wang, B., Chen, X., Wang, J., & Zou, J. (2022). Structural and functional analyses of type I IFNa shed light into its interaction with multiple receptors in fish. *Frontiers in Immunology*, *13*, 862764. https://doi.org/10.3389/fimmu.2022.862764

West, A. C., Mizoro, Y., Wood, S. H., Ince, L. M., Iversen, M., Jørgensen, E. H., Nome, T., Sandve, S. R., Martin, S. A. M., Loudon, A. S. I., & Hazlerigg, D. G. (2021). Immunologic profiling of the Atlantic salmon gill by single nuclei transcriptomics. *Frontiers in Immunology*, *12*, 669889. https://doi.org/10.3389/fimmu.2021.669889

Wilson, A. B. (2017). MHC and adaptive immunity in teleost fishes. *Immunogenetics*, *69*(8–9), 521–528. https://doi.org/10.1007/s00251-017-1009-3

Wu, Z., Jiang, Z., Li, T., Xie, C., Zhao, L., Yang, J., Ouyang, S., Liu, Y., Li, T., & Xie, Z. (2021). Structural variants in the Chinese population and their impact on phenotypes, diseases and population adaptation. *Nature Communications*, *12*(1), 6501. https://doi.org/10.1038/s41467-021-26856-x

Zhang, Q., & Cao, X. (2019). Epigenetic regulation of the innate immune response to infection. *Nature Reviews. Immunology*, *19*(7), 417–432. https://doi.org/10.1038/s41577-019-0151-6

Zhang, R., Calixto, C. P. G., Marquez, Y., Venhuizen, P., Tzioutziou, N. A., Guo, W., Spensley, M., Entizne, J. C., Lewandowska, D., Ten Have, S., Frei Dit Frey, N., Hirt, H., James, A. B., Nimmo, H. G., Barta, A., Kalyna, M., & Brown, J. W. S. (2017). A high quality Arabidopsis transcriptome for accurate transcript-level analysis of alternative splicing. *Nucleic Acids Research*, *45*(9), 5061–5073. https://doi.org/10.1093/nar/gkx267

Zhang, T., Cooper, S., & Brockdorff, N. (2015). The interplay of histone modifications - writers that read. *EMBO Reports*, *16*(11), 1467–1481. https://doi.org/10.15252/embr.201540945

Zhang, Y.-B., Liu, T.-K., Jiang, J., Shi, J., Liu, Y., Li, S., & Gui, J.-F. (2013). Identification of a novel Gig2 gene family specific to non-amniote vertebrates. *PloS One*, *8*(4), e60588. https://doi.org/10.1371/journal.pone.0060588

Zou, C., Sun, K., Mackaluso, J. D., Seddon, A. E., Jin, R., Thomashow, M. F., & Shiu, S.-H. (2011). Cis-regulatory code of stress-responsive transcription in Arabidopsis thaliana. *Proceedings of the National Academy of Sciences of the United States of America*, *108*(36), 14992–14997. https://doi.org/10.1073/pnas.1103202108

Zou, J., Gorgoglione, B., Taylor, N. G. H., Summathed, T., Lee, P.-T., Panigrahi, A., Genet, C., Chen, Y.-M., Chen, T.-Y., Ul Hassan, M., Mughal, S. M., Boudinot, P., & Secombes, C. J. (2014). Salmonids have an extraordinary complex type I IFN system: characterization of the IFN locus in rainbow trout oncorhynchus mykiss reveals two novel IFN subgroups. *The Journal of Immunology*, *193*(5), 2273–2286. https://doi.org/10.4049/jimmunol.1301796

Zou, J., & Secombes, C. J. (2016). The function of fish cytokines. *Biology*, *5*(2), 23. https://doi.org/10.3390/biology5020023

6 Papers

Paper 1

Dissecting the genetic basis of response to salmonid alphavirus in Atlantic salmon

Domniki Manousi [1], Dorota Monika Jaskula [1], Fabian Grammes [2], Tim Martin Knutsen [2], Shahmir Naseer[3], Samuel AM Martin [3], Thomas Moen[2], Marie Saitou [1] and Sigbjørn Lien[1]

[1]Centre for Integrative Genetics, Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Oluf Thesens vei 6, 1433 Ås, Norway
[2]AquaGen AS, Havnegata 9, 7010 Trondheim, Norway
[3]Scottish Fish Immunology Research Centre, School of Biological Sciences, University of Aberdeen, Aberdeen AB24 2TZ, United Kingdom

Corresponding authors: Domniki Manousi, domniki.manousi@nmbu.no

Abstract

Background

The development of effective disease management strategies is crucial for the assurance of welfare and sustainability of the aquaculture industries. Pancreas disease (PD) is a major challenge faced by Atlantic salmon aquaculture with viral outbreaks resulting in substantial production losses and raising significant welfare concerns for farmed salmon populations. Previous research has identified several quantitative trait loci (QTL) associated with PD resistance accounting for a substantial additive genetic component. However, pinpointing the underlying causal variation remains challenging, partly due to the location of the QTL within duplicated regions of the Atlantic salmon genome that share high sequence similarity. The present study leverages the latest advancements in Atlantic salmon genomics in order to uncover the genetic landscape underlying PD resistance and identify genomic variation with putative functional impact on disease response.

Results

Association mapping and haplotype analysis of fish challenged with salmonid alphavirus (SAV3), either through peritoneal injection or infectious cohabitation, confirmed the presence of a major QTL region on chromosome Ssa03. Additionally, another QTL on

Ssa07 was detected, linked to infection-specific response. Transcriptomics analysis of the genes overlapping the Ssa03 QTL region revealed significant expression differences among three tandemly duplicated *gig1-like* genes, whereas allele-specific expression analysis detected several SNPs with putative functional impact on the particular genes. Use of long-read sequencing and construction of disease-associated haplotypes identified more complex variation in the region and offering a detailed exploration of the genetic architecture underlying PD resistance. Finally, integration of the regulatory landscape of Atlantic salmon during viral infection response improved genomic resolution, providing novel insight into the potential causal variation underlying pancreas disease in Atlantic salmon.

Conclusions

This study provides a detailed investigation of the genetic architecture underlying PD resistance in farmed Atlantic salmon. Using advanced genomic resources, three copies of the gig1-like gene were identified as likely causal candidates for a major QTL associated with PD resistance. Additionally, genomic variations with potential functional impacts on gig1-like expression were uncovered. These findings hold promise for application in developing effective disease management strategies in Atlantic salmon aquaculture.

Keywords: Atlantic salmon, pancreas disease, salmonid alphavirus, structural variation, whole genome duplication

Background

In recent years, aquaculture has emerged as one of the fastest growing sectors of food production, contributing significantly to the global food supply chain [1]. With the rapid growth in aquaculture, several challenges have emerged regarding production sustainability and animal welfare, with one of the most significant problems being disease resistance.

Pancreas disease (PD), caused by salmonid alphavirus (SAV), poses a significant challenge for Atlantic salmon (*Salmo salar*) aquaculture. Although the SAV family consists of several subtypes (SAV1-SAV7), only SAV2 and SAV3 subtypes are found in Norway, with the SAV3 subtype being more aggressive in terms of mortality and clinical symptoms [2]. Infection with PD can occur during both the freshwater and seawater phases of salmon production and results primarily in inflammation, extensive necrosis and loss of exocrine pancreatic tissue. In addition to clinical symptoms, PD manifestations include loss of appetite, yellow mucoid gut contents, lethargy, abnormal swimming activity, skin erosion and skeletal myopathies [3, 4] while at the same time, immuno-compromised individuals become susceptible to secondary bacterial infections [5–7]. PD viral outbreaks result in mortality rates of up to 28% [8] and although recovery is possible, surviving individuals remain severely stunted, causing further damage to production [4]. PD virus is transmitted via water contact; viral particles from mucous and fecal excretions of infected or decomposing fish in the enclosure are released into the water, infecting healthy cohabitants [9]. Current efforts to mitigate the impact of PD include prophylactic vaccinations and selective breeding [10, 11]. Although application of such measures has managed to reduce the incidence and severity of PD outbreaks [11, 12], they remain only partially effective.

Due to the negative impact of PD on Atlantic salmon aquaculture, numerous studies have focused on uncovering the genomic basis of resistance to SAV3 virus. Genome scans using SNP-arrays have attributed a substantial genetic component to the disease with heritability ranging between 0.2-0.6 [13–15], while several quantitative trait loci (QTL) associated with PD resistance have been detected on chromosomes 3, 6, 7 and 23 (Ssa03, Ssa06, Ssa07 and Ssa23, respectively). Across studies, the QTL with highest significance for PD response is found on Ssa03, explaining between 15 – 41% of the total genetic variation [13–15]. However, this QTL covers a large region (approximately 27 Mb), making identification of candidate genes and putative causal variation particularly challenging.

A complicating factor for fine-mapping and identification of underlying causal variation is the physical position of the Ssa03 QTL; the particular QTL-region is located at the end of the q-arm of Ssa03, sharing high sequence similarity with a duplicated region on the p-arm of Ssa06 (Lien et al., 2016). The observed similarity is attributed to delayed

rediploidization following the salmonid-specific whole-genome duplication event (Ss4R) that occurred approximately 89-125 million years ago (Gundappa et al., 2022). The high sequence similarity between the two chromosomal regions resulted in a substantially fragmented assembly and poor resolution of the duplicated sequences in the first Atlantic salmon reference genome (ICSASG_v2; GCA_000233375.4) (Lien et al., 2016), compromising downstream analyses.

Recently, some of these challenges have been addressed by the construction of a long-read based reference genome for Atlantic salmon (Ssal_v3.1; GCA_905237065.2), which has greatly improved contiguity and genomic resolution of complex genomic regions. In addition, functional information of the Atlantic salmon genome is currently available [16], providing invaluable insight over the regulation of genes during disease infection [17]. The present study leveraged the latest advances in Atlantic salmon genomics to fine map the genomic loci associated with PD resistance and elucidate the underlying genomic variation landscape, providing novel insights for developing sustainable and effective disease management strategies.

Materials and Methods

Resource populations

The present study utilized several populations originating from both commercial lines and from the breeding population of AquaGen AS. In particular, 293 fish from the AquaGen AS breeding population, hatched between 1998 and 2014, were used to provide whole-genome sequence level genotype information through short-read sequencing technologies. In addition to these samples, seven more individuals from the breeding population of AquaGen AS were used to provide whole-genome sequence level genomic information through long-read sequencing technologies.
Furthermore, four distinct populations (5,628 fish) from the AquaGen AS commercial line were used in this study, participating in independent *Salmonid alphavirus* subtype 3 (SAV3) viral challenge trials. The latter four populations, hatched during 2018, 2019, 2020 and 2022, were used in three distinct SAV3 survival challenges (year-classes 2018, 2019 and 2020) and one SAV3 viral cohabitation challenge trial (year-class 2022), respectively. Informed consent of participation for all privately owned individuals was provided by AquaGen AS.

Besides the populations described, the present study additionally leveraged transcriptomics and regulatory information from samples that participated in previously published research [15, 17, 18]. Details regarding the origin of these individuals can be found in their respective studies whereas informed consent for participation regarding these samples was either not required (publicly available data) or provided by their respective owners. A visual representation of the genomic material used in the present study is provided in **Figure 1**.

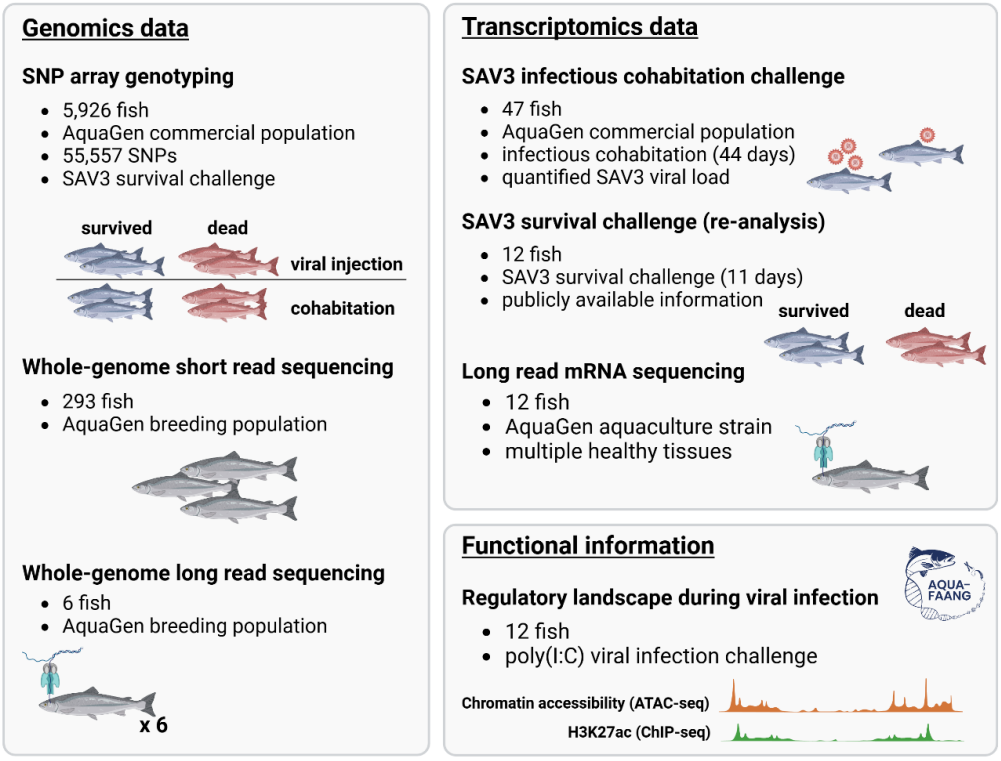


Figure 1. *Visual representation of the material used in the present study. In brief, samples from several populations were used, originating from breeding and commercial lines of AquaGen AS. Origin of each population used in the study is marked in the figure while sampling and experimental details of the same groups are described in the respective methods sections. In addition to samples retrieved from commercial populations, transcriptomics (long read mRNA sequencing and gene expression data from a SAV3 survival challenge) as well as functional information were retrieved from previously published research. Further details regarding the origin of these data are provided in the respective methods section whereas informed consent for the use of these data was either not required (publicly available repositories) or provided by their rightful owners.*

SAV3 survival challenge samples

Genotype and phenotype data from a cohort of 5,628 Atlantic salmon parr originating from three distinct AquaGen commercial populations (AquaGen AS), hatched between the years 2018, 2019 and 2020 were used in this study. Participating individuals were PIT-tagged and adipose fin clips were collected from anaesthetized fish for DNA extraction and SNP genotyping. SNP genotyping was performed using either of two proprietary AquaGen genotyping arrays (55,557 SNPs, Affymetrix axiom arrays Ssa70kv1 and Ssa70kv2, **Table 1**). Singular exception constituted the year class of 2020

for which genotype data were available for only a part of the challenge population (**Table 1**).

Each of the three year-groups participated independently in a controlled PD challenge trial at the VESO Vikan facilities (Namsos, Norway). Standard operating procedures were used for conduction of each SAV3 challenge, and all challenges were approved beforehand by the Norwegian National Animal Research Authority (FOTS) (approval no. 15233, 19332 and 23006 for conduction of the 2018, 2019 and 2020 challenge tests, respectively). For each challenge test, grouped samples of Atlantic salmon parr (3-4g growth stage) were subjected to SAV3 infection either via peritoneal injection (IP group) or through cohabitation with injected fish (CH group). For the IP group, fish were first anaesthetized with 3 ml of 20 % benzocaine in 10 L of water and then injected with 100 µl of a SAV3 isolate (Jnr 2618, passaged 3 times in CHSE-214 cell). Each challenge test was performed in a 0.6 m tank containing 125 L of 12 ºC water, with a 24:0 light - darkness regime and feeding according to appetite, whereas the ratio of CH to IP individuals per tank ranged between 1.0 - 1.4. Challenge tests spanned a duration of 7 to 9 weeks with tank inspections occurring two times a day. During inspection, dead and moribund fish were removed from the tank and individual survival as well as infection group of each fish were recorded accordingly. Fish found in a moribund state during the challenges as well as challenge survivors were first anesthetized and then euthanized using 6 ml of 20 % benzocaine per 10 L of water. All infectious challenges were conducted in accordance with the legislation regarding experiments and procedures for live animals in Norway (Animal welfare Act of June 19th, 2009 and the Regulation on Animal Experimentation of January 15th, 1996).

Whole-genome sequencing data and variant calling

A whole-genome level genotype dataset was generated using a collection of 293 whole genome resequenced (WGS) Atlantic salmon fish originating from the AquaGen breeding population (AquaGen AS). The particular population comprised of 94 ancestors hatched between the years 1998 and 2001 as well as a group of 199 fish hatched in 2014. The latter samples included direct ancestors (parents) of fish participating in the PD challenge trials of this study (year-class 2018). To produce genetic information from this population, fish were anaesthetized, and fin clips were obtained from each individual for DNA extraction and sequencing. DNA extraction followed standard protocols and whole-genome sequencing was performed using BGISEQ and Illumina sequencing technologies.

Produced DNA sequencing reads from all samples were aligned against the latest Atlantic salmon genome reference Ssal_v3.1 (GenBank accession GCA_905237065.2) using BWA-mem2 and duplicated reads were marked with Samblaster 0.1.26 [19, 20]. Following alignment, Samtools and the GATK4 tools ‘BaseRecalibratorSpark’ and ‘ApplyBQSRSpark’ were used in a custom - built pipeline in order to ensure high

consistency of variant calling across samples sequenced using different sequencing platforms [21, 22]. Finally, genetic variants were called for each sample using the 'DeepVariantWGS' mode of the variant caller DeepVariant v1.1.2 and individually detected variants were merged across all fish using GL-NEXUS v1.4.1 [23]. Filtering of the GLNEXUS output followed previously described benchmark recommendations [24], omitting filtering for allele balance. Instead, the excess of heterozygosity was considered in this study, discarding variants with allele departures lower than 1e-8. Quality filtering of SNP variant calls from GL-NEXUS retained in total 16,339,880 bi-allelic SNPs.

Filtering and imputation of whole genome sequencing variants in the SNP array genotyped fish

A reference population was created from the 293 WGS fish. Following previously established quality control protocols [25], SNP filtering was performed using PLINK2 [26] and the following threshold criteria: minor allele frequency (maf) < 0.01 and SNP calling rate < 0.90, retaining in total 6,115,911 SNPs. Prior to performing SNP imputation on the array-genotyped challenge population, a cross–validation analysis was implemented in the WGS dataset in order to filter SNPs with high genotype inference accuracy. To achieve this, the WGS dataset was randomly divided into five subsets and the genotypes of each subset were sequentially masked, retaining only SNPs commonly found between the WGS samples and the SNP array-genotyped fish of the SAV3 viral challenge. Each masked dataset was then haplotyped and imputed back into WGS genotype density using Beagle 5.4 and the remaining four datasets as the reference [27, 28]. Imputed SNP genotypes were finally compared against the masked "true" genotypes and imputation accuracy - estimated as the r^2 correlation between masked and inferred genotype - was assessed for each imputed SNP, discarding variants with r^2 accuracy below 0.80. Filtering of SNPs with high imputation reliability in the WGS population retained a total of 891,641 SNPs.

For the SAV3 viral challenge, 5,628 fish were genotyped for 55,557 SNP using either of two Affymetrix SNP arrays (**Table 1**). SNP quality filtering with PLINK2 [26] removed variants with minor allele frequency < 0.02 and missing genotype call rate < 5%, retaining in total 53,295 SNPs. The reference and alternative alleles of filtered SNPs were then matched across those of the respective SNPs in the WGS population dataset and allelic consistency was corrected using the software conform-gt (version 24May16.cee.jar). The genotypes of allele - conforming SNPs on the SAV3 challenge fish were finally haplotyped using Beagle 5.4 and the genotypes of high imputation accuracy SNPs were inferred from the reference WGS dataset using Beagle 5.4 with customized haplotype and effective population options (overlap=5, window=39, and ne=10000, respectively). To remove potential artefacts from the imputation process, the imputed

dataset was additionally filtered against minor allele frequencies lower than 2% using PLINK2.

Estimation of heritability and variance component analysis

SNP-based heritability for pancreas disease resistance, described as survival from a SAV3 infectious challenge (dead or alive), and survivability (survived days post infection, dpi) were estimated using GCTA [26, 29]. A genomic relationship matrix (GRM) was used, constructed using the imputed genotypes and the **GCTA-LDMS** method implemented in the same software. The particular method was used in order to account for linkage disequilibrium (LD) bias observed due to genotype inference and increased SNP density (whole-genome level) [30]. Heritability estimations followed the linear mixed model:

$$Y=X\beta+Zu+\epsilon \quad (1)$$

where **Y** is a vector of 'n' records on PD resistance and/or survivability, **β** is a matrix of coefficients for the fixed effects as well as the intercept; **Z** is the corresponding incidence matrix and **u** is the additive genetic effect utilizing the GRM and is distributed as ~ $N(0, G\sigma a^2)$, where G corresponds to the GRM and σa^2 is the additive variance. Finally, **e** is a vector of residual effects.

Data were analysed separately for each year-class as well as collectively across all challenge trials for PD resistance (death or survival) and infectious group (CH or IP). Depending on the type of collective analysis, infection type (IP and CH) and/or year-class were fitted as fixed effects. In addition, a series of bivariate analyses were run to estimate the genetic correlation between PD resistance and survivability for the collective datasets as well as for the independent year-classes and infection groups (**Additional file 1: Supplementary table S1**). To perform the bivariate analyses a modification of model (1) was used where the y is a matrix of n records on PD resistance and survivability.

Genome-wide association study

A series of univariate genome-wide association analyses (GWA) were conducted to identify significant associations between PD resistance and viral infection approach (IP or CH) using the imputed SAV3 challenge fish. Association analysis of the challenge population used a linear mixed model approach, as this is implemented in the software BOLT-LMM [31]. The model considered individual year-class and viral infection group as covariates, as well as the first 5 genetic variance components (PCs) estimated from 53,295 array-genotyped SNPs using PLINK2 [26]. Association significance for each analysis was corrected for multiple testing-bias using the Bonferroni correction criterion implemented in RStudio [32]. In addition, the genome-wide significance cut-off threshold was estimated using the R package Rainbowr [33] and the uncorrected p-

values. To calculate this cut-off threshold, the Bonferroni correction criterion was considered with a significance threshold of 0.05.

Finally, following the GWAS analysis, the approach described in Shim et al. was used to estimate the proportion of phenotypic variance explained by the SNP(s) with highest association for PD resistance [34]. Estimation used the information provided from the association analysis output in the following relationship:

$$\frac{2\hat{\beta}^2 MAF(1-MAF)}{2\hat{\beta}^2 MAF(1-MAF) + \left(se(\hat{\beta})\right)^2 2NMAF(1-MAF)} \quad (2)$$

where MAF notates the minor allele frequency for a given SNP, N is the sample size and se($\hat{\beta}$) is the standard error of the effect size for that SNP.

Haplotype block and linkage disequilibrium analysis

Haplotype-resolved (phased) genotypes of the imputed challenge population were used to define blocks of SNPs sharing substantial linkage disequilibrium (LD). Haplotype block analysis was carried out in PLINK v1.9 using the method of Gabriel et al. [35, 36]. To avoid possible disruption of LD patterns, the full length of PD associated chromosomes was analyzed. Haplotype analysis constructed 1,225 and 794 haplotype blocks for chromosomes Ssa03 and Ssa07, respectively. Within each of the two chromosomes, the haplotype block containing the highest associated SNP for PD resistance was retained.

In order to further refine the PD associated haplotypes, the LD patterns of SNPs inside the retained haplotype blocks were visualized in Haploview [35, 37]. For each block, the Pairwise LD relationships between each SNP and the QTL top-SNP were evaluated, filtering markers sharing near perfect LD ($r^2 > 0.95$). Filtering was repeated until the retained SNPs of each block formed haplotype combinations with unique major and minor allele instances. Refined haplotype blocks were then converted into genotypes using the R package GHap [38, 39] and association significance was tested using Gemma [40]. For the haplotype-based association analysis, the same linear mixed model as before was used (model 1), together with the respective parameterization.

Transcriptome analysis of a SAV3 infectious cohabitation challenge

Transcriptomics data were obtained from an AquaGen AS commercial population, hatched in 2022, that participated in a SAV3 infectious cohabitation challenge. For the infectious cohabitation challenge, 618 healthy and PIT-tagged parr (average weight 17.3 g) with known genotype information for 63,849 SNPs (AquaGen proprietary Affymetrix 70kv3 SNP genotyping array) were placed in a common garden

experiment together with SAV3 infected fish from the same aquaculture strain. Infected individuals were first anaesthetized and then virally injected with SAV3 (Jnr 2618, passaged 3 times in CHSE-214 cells, 100 µl inoculation volume). After 44 days of viral cohabitation all fish were anaesthetized and euthanized using 6 ml of 20 % benzocaine per 10 L of water, and heart tissue of infected cohabitants was sampled and stored in RNAlater (Thermofisher). Similarly to the SAV3 survival challenge tests, conduction of the viral cohabitation was approved beforehand by the Norwegian Animal Research Authority (approval no. 29261) while the challenge trial followed standard operating procedures of VESO Vikan (Namsos, Norway).

Total RNA extraction as well as qPCR for SAV3 viral load was conducted by BlueAnalytics (Bergen, Norway). Selection of samples for mRNA sequencing was based on the individual genotype for the Ssa03 QTL top-SNP (Ssa03_95086730) and infection severity – classified as the quantified viral load title (difference between qPCR viral titre and qPCR expression of housekeeping gene *elf1a*). Extracted mRNA material from 47 selected samples (18 reference homozygotes, 10 heterozygotes and 19 alternative homozygotes) with viral load values (delta Ct= Ct SAV – Ct EF1α) ranging between 1-8 was sequenced, producing approximately 150 million pair-end reads per individual. Raw RNA sequencing reads were then quality filtered using fastp v0.23.2 [41] to remove sequencing adapters and low quality reads and aligned against the Atlantic salmon transcriptome (Ssal_v3.1, accession number GCA_905237065.2) using STAR v2.7.10 [42]. Gene expression of aligned reads was quantified using FeatureCounts v2.0.3 [43], whereas sequencing reads that did not map successfully to the salmonid genome where aligned against the SAV3 viral genome using Kraken2 v2.1.2 [44] to verify correspondence between viral concentration and qPCR-based viral load (p-value < 2.2e-16, spearman's rank test).

In order to investigate differences in expression profile between mildly and severely infected samples, quantified expression was used to perform a differential gene expression (DEG) analysis using the package DESeq2 in RStudio [45, 46]. To reduce the dimensionality of the viral load dataset (titre scaled between 1-8), quantified viral load was classified as either high, medium, or low (values < 4 = high viral load, values > 6 = low viral load, in-between values=medium viral load). To observe higher contrasts in gene expression profile, the DEG analysis focused exclusively on comparison between samples grouped into high and low viral load classes, each of which consisted of 12 samples. Genes with significant expression differences were filtered based on False discovery rate (FDR) adjusted p-value (p-value < 0.05) and visualization of expression patterns was performed using the package 'ggplot' in Rstudio [47].

Transcriptome re-analysis of a SAV3 viral injection challenge

Publicly available mRNA sequencing data from a SAV3 viral challenge [27] were obtained from the European Nucleotide Archive repository (Study accession: PRJNA590166). In this challenge, 12 samples were peritoneally (IP) injected with SAV3 virus and participated in a survival challenge for four weeks. During the experiment, any participant fish found in moribund state was removed and survival was recorded as either dead or alive after 11 days.

To produce comparable results between experimental setups, quality filtering and quantification of the survival challenge transcriptomics data followed the same pipeline as previously described, omitting the step of viral sequence quantification. Similarly to the viral cohabitation samples, a DEG analysis was performed using DESeq2 to explore differential gene expression between IP fish that survived or died after 11 days of the viral challenge and genes with (FDR) adjusted p-value < 0.05 were retained as significantly differentially expressed.

Allele specific expression (ASE) analysis using infectious cohabitation samples

An allele specific expression (ASE) analysis was performed in the 10 SAV3 infected cohabitants that carried the heterozygote genotype for the Ssa03 QTL top-SNP (ssa03_95086730) in order to identify SNP variants with allele specific impact on gene expression. For the ASE analysis, all biallelic SNPs overlapping the Ssa03 QTL region (Ssa03: 95,054,413 – 95,127,543) were obtained from the quality filtered WGS dataset - irrespective of imputation accuracy score. Retained SNPs were haplotyped using Beagle v5.4 [27] and the LD relationship between each of these SNPs and the QTL top-SNP was estimated using VCFtools v0.1.16 [48]. SNPs with LD score higher than 0.6 were then re-coded as heterozygous and used to identify mRNA sequencing reads mapping uniquely within the Ssa03 QTL region. Unique read mapping was performed using STAR v2.7.10 aligner and the option *--waspOutputMode* [42, 49]. Furthermore, in order to account for increased sequence similarity due to the salmonid-specific genome duplication, the alignment process was adjusted using the option *–outSAMmapqUnique* [50]. Allele specific expression was estimated for each SNP based on the uniquely mapping read alignment using the GATK4 tool ASEReadCounter [51]. Finally, SNP-wise significance between gene expression and the alleles of each SNP was estimated using the exact binomial test option '*binom.test'* in R [51, 52].

Detection of complex variants using long-read sequencing technologies

To analyze complex variation in the salmonid genome, whole-genome sequencing information was obtained from six fish originating from an AquaGen breeding population (AquaGen AS). The fish were anesthetized using a 20% benzocaine overdose bath and then bludgeoned, and blood was sampled from the bludgeoning point for DNA extraction. Conduction of the sampling took place at an AquaGen AS

licensed facility (Steigen, Norway), following the legislations regarding animal welfare and operation of aquaculture facilities in Norway (Animal welfare Act of June 19th, 2009, and regulation on the Operation of Aquaculture Facilities of June 17th, 2008).

Extraction of DNA for these samples followed standard protocols and genomic material was sequenced using Oxford Nanopore (ONT) sequencing technologies on the PromethION sequencing platform. In total, DNA sequencing reads were produced for each sample, providing a sequencing coverage of approximately 20x for each sequenced individual. In addition to the obtained fish, the present study also included publicly available sequencing information from the individual used to construct the Ssal_v3.1 genome assembly (sample accession SAMEA8062739). Sequencing quality and therefore coverage of the particular individual was significantly higher (approximately 70x).

Sequencing reads for the seven samples were quality filtered using Filtlong v0.2.1 to remove the worst 10% of reads as well as reads with length shorter than 4000 bp. Filtered output was then aligned against the Atlantic salmon genome reference Ssal_v3.1 using minimap2 [52, 53] and reads overlapping up to 100 Kb upstream and downstream of the Ssa03 QTL region (Ssa03: 94,976,092-95,221,176) were extracted using Samtools v1.11 [53]. Sequencing read-based haplotyping was performed on the aligned sequence of each sample using the software PEPPER to link adjacent variants through Recurrent Neural Networks and Convolutional Neural Network. Haplotyped reads were then tagged to a respective haplotype using the software Margin. Both haplotyping tools are implemented within the PEPPER-Margin-DeepVariant r0.8 workframe [54]. Haplotype-tagged reads were classified as either of two haplotypes for reads with high confidence discrimination, whereas reads with low confidence on haplotype discrimination were tagged as ambiguous. Based on the assigned tags, haplotype resolved as well as "ambiguous" reads were filtered for each of the two haplotypes of each sample to ensure sufficient depth for homozygous individuals as well as for homozygous regions of heterozygous individuals. Finally, assemblies of each haplotype were produced using the option *--nano-raw* in Flye v2.9 [55]. Each of the assembled haplotypes consisted of singular contigs with read coverage ranging between 7-17. Variant calling and genotyping of the 14 assembled haplotypes was performed using the pangenome graph-based tool Minigraph-Cactus v2.6.9 [56]. Implementation of the software followed user guidelines with the exception of the *'graphmap-split'* computational step. To distinguish disease associated haplotypes in the long-read sequenced fish, the allele segregation of SNP variants matching the Ssa03 QTL top-SNP as well as the set of 10 SNPs sharing LD higher than 0.95 in the SAV3 challenge data was assessed. Finally, variants with distinct allele segregation patterns between disease-associated haplotypes that carried either the reference or alternative top-SNP alleles were retained (**Additional file 1: Supplementary Table S4**).

Long-read sequencing transcriptomics data and manual annotation of the Ssa03 QTL genes

Whole genome transcriptomics data from ONT long read sequencing technologies were obtained in order to study the structure of PD associated genes [50, 51]. The dataset included sequencing samples from multiple tissues (brain, liver, head kidney, muscle, gill, intestine, and gonad) of 12 healthy individuals that participated in the AQUA-FAANG project (https://www.aqua-faang.eu/). DNA extraction and sequencing details for the respective samples have been previously described [51, 52]. In brief, quality filtered sequencing mRNA reads of all sample tissues were aligned against the latest Atlantic salmon genome reference (Ssal_v3.1, accession number GCA_905237065.2) using minimap2 [52, 53] and gene expression was quantified using the Flair bioinformatics pipeline [57].

Aligned transcriptomic information from the long read sequencing data was inspected manually using the Integrative Genomics Viewer platform [58] and the distribution of mRNA sequencing reads was compared against short-read transcriptomics data of samples infected with high viral load in the previously described SAV3 infectious cohabitation challenge. In addition, two publicly available versions of the Atlantic salmon transcriptome (Ssal_v3.1, accession numbers: GCA_905237065.2 and GCF_905237065.1) were also consulted in order to assess the structure of genes associated with pancreas disease (**Figure 7**). The nucleotide sequence corresponding to the short- and long mRNA read alignment was extracted and coding sequence for each likely exon region was predicted using the open reading frame (ORF) tool, ORF Finder that is implemented in the NCBI platform [59] (**Additional file 1: Supplementary table S3**). Prediction was performed using default settings (minimum ORF > 75 bp and only matching an ATG start codon). Finally, the multiple amino acid sequence alignment of predicted exons was used to perform protein domain prediction using the InterPro database with default settings [60].

Investigation of variants with putative regulatory impact

In order to identify genomic variation with putative regulatory impact on the Ssa03 QTL region, the regulatory landscape of Atlantic salmon during viral infection stimulation was investigated. Regulatory information regarding histone modification (ChIP-seq) and chromatin accessibility (ATAC-seq) were obtained from an *in vivo* experiment of artificial Interferon (IFN) stimulation of salmon parr head kidney (HK) tissue using the viral mimic poly(I:C). The provided data were part of the Aqua-FAANG ImmunoMap project (AQUA-FAANG consortium, https://www.aqua-faang.eu/) and a detailed explanation regarding DNA extraction and production of ChIP-seq and ATAC-seq signal peak data is provided elsewhere [60, 61]. Briefly, H3K27ac and H3K27me3 sequencing data from ChIP-seq and chromatin accessibility data from ATAC-seq were obtained from six samples treated to 100 µg of poly(I:C) and six samples treated to

control (PBS media) peritoneal injections. Sequencing reads were quality filtered and processed using the nf-core pipelines ChIP- and ATAC-seq, respectively to produce quantified signal peaks [61, 62]. Detection of peaks with significant differences in signal abundance between IFN stimulated and control samples was performed using DESeq2 in Rstudio and intervals of significant difference in signal abundance were filtered based on FDR adjusted p-value < 0.05 [62, 63].

Phylogenetic analyses of PD associated *gig1-like* genes

The manually annotated amino acid sequence for the three *gig1-like* gene copies was used to perform a protein BLAST search on Ensembl (release version 112) and NCBI public databases in order to identify *gig1-like* gene copies on chromosome Ssa06 in Atlantic salmon [63, 64]. Genes located within ohnolog regions of the Atlantic salmon genome were filtered based on synteny analysis information obtained through the Salmobase platform (https://salmobase.org). Synteny filtering retained the gene ENSSSAG00000003892 (Ssa06:5,558,971-5,559,951). The amino acid sequence for that gene was retrieved and aligned against the three Ssa03 gene copies using MAFFT v7.511 [65].

In order to investigate the life history of the four Atlantic salmon *gig1-like* genes, an additional protein BLAST was performed to detect homologous genes with high similarity in the Northern pike (*Esox lucius*), a close ancestral species that did not undergo the salmonid-specific genome duplication. The coding sequence of the four *gig1-like* salmonid genes together with the Northern pike gene showing highest amino acid sequence similarity (ENSELUG00000010316, 11:36,088,635-36,091,358) was aligned using MAFFT v7.511 [65]. Multiple sequence alignment used default options and the automated model selection strategy in order to utilize the best fitting model for the provided dataset. Furthermore, a neighbor joining tree was constructed based on the gap-free and moderately conserved sites of the coding sequence alignment and bootstrap values were calculated using 500 replicates. The produced phylogenetic tree was visualized using the iTOL platform [66].

Results

Pancreas disease SAV3 challenge trial

Our analysis utilized genotype and disease survival data collected from three independent challenge trials of Atlantic salmon fish infection with the Salmonid alphavirus SAV3 (**Table 1** and **Figure 2**, Materials and methods). The three challenges were carried out between 2018 and 2020, classifying the participating fish respectively. For each of the three trials, fish were infected with Salmonid alphavirus (SAV3) either via peritoneal viral injection (IP infection group: 1,112, 1,023 and 1,060 fish of year class 2018, 2019 and 2020 respectively) or through infectious cohabitation with injected

individuals (CH infection group: 1,137 and 1,296 fish of year class 2018 and 2019, respectively), with the latter method mimicking SAV3 transmission during natural an aquaculture outbreak scenario. Mortality rates reported during each trial showed a uniform and sharp increase at approximately 15 days post infection (dpi) for IP infected fish and at approximately 25 dpi for the CH group (**Figure 2**). Similarly, mortalities across the three challenges reached a uniform plateau at approximately 30 and 45 dpi for IP and CH infected fish, respectively (**Figure 2**). By termination of the three trials, the mortality rate of IP infected fish was moderate to high (31-76%), whereas mortality of infected cohabitants (CH) was substantially lower (22-30%) (**Table 1 and Figure 2**). Estimated heritability of PD resistance estimated across all challenges was moderately high (0.28 +/- 0.08), whereas infection-based heritability was higher for IP infected fish (0.36 +/- 0.02) and considerably lower for the CH infection group (0.25 +/- 0.03). Lastly, disease survival and survivability (dpi) were highly correlated across challenges (~99%, **Additional file 1: Supplementary table S1**) and therefore the following association analyses focused exclusively on the phenotype of PD resistance. Heritability and genetic correlation estimates of PD resistance and survivability for each year class and infection group are provided in **Supplementary table S1** (**Additional file 1**).

Table 1. *Composition of the PD study population. In total 5,628 fish were used in the present study that originated from three distinct populations hatched between the years 2018, 2019 and 2020. Individuals were genotyped using proprietary AquaGen genotyping arrays prior to the infection challenge. Infection of participating individuals occurred either through peritoneal injection (IP group) or via cohabitation of healthy fish together with injected individuals (CH group). Singular exception constitutes the 2020 year-class where only IP infected individuals were available for analysis.*

Year class	*IP group*	*CH group*	*Genotyping array*	*Challenge termination (days post infection)*	*Mortality rate (CH/IP group)*	*Total*
2018	1,112	1,137	Ssa70kv1	48	22% / 31%	2,249
2019	1,023	1,296	Ssa70kv2	61	30% / 76%	2,319
2020	1,060	-	Ssa70kv2	47	- / 44%	1,060
Total	3,195	2,433				5,628

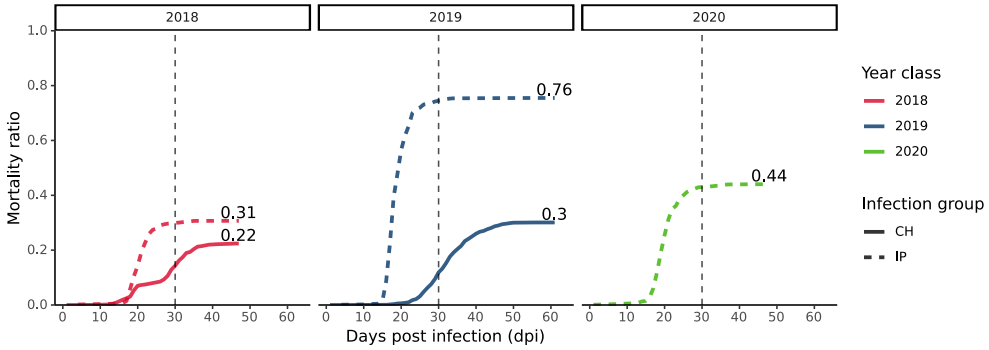


Figure 2. *Mortality curves of three independent survival challenges of fish infected with Salmonid Alphavirus (SAV3), classified by the year they were carried out. In total 5,628 individuals were infected either by peritoneal SAV3 injection (IP group: 1,112, 1,023 and 1,060 fish of year class 2018, 2019 and 2020 respectively) or by infectious cohabitation with virally injected individuals (CH group: 1,137 and 1,296 fish of year class 2018 and 2019, respectively). Colors indicate the challenge year class while dashed and solid lines represent the IP and CH infection groups, respectively. Numbers above each line indicate the mortality rate at termination of each trial, 48 dpi, 61 dpi and 47 dpi for the challenges of year class 2018, 2019 and 2020, respectively.*

Two QTL regions govern response to PD

To improve the mapping resolution of QTL related to PD resistance in Atlantic salmon, we improved the genome scan by imputing missing genotypes in the PD challenge population and conducted a genome-wide association study (GWAS). More specifically, the application of genotype imputation inferred genotypes for 885,287 missing SNPs, ultimately increasing the genotype density of the PD challenge population to 938,582 SNPs.

Association testing and correction for multiple testing bias ($p_{Discovery} < 10^{-7.23}$) detected significant results for PD survival on chromosomes Ssa03 and Ssa07 (**Figure 3**). The most significant association peak was located within a broad region (33Mb) on Ssa03 (Ssa03: 64,745,587 – 97,690,777 Mb), while a second peak appeared within an 11Mb region on Ssa07 (Ssa07: 43,093,416 – 54,519,108). Within each QTL region, the highest association signals were detected for the SNPs ssa03_95086730 and ssa07_49303476, explaining 4% and 2% the overall phenotypic variance, respectively. Interestingly, independent analyses of PD survival for each of the two viral infection groups (IP or CH fish) both detected the QTL on Ssa03. In contrast, the QTL on Ssa07 was significantly associated only in the IP infection group (**Figure 3**).

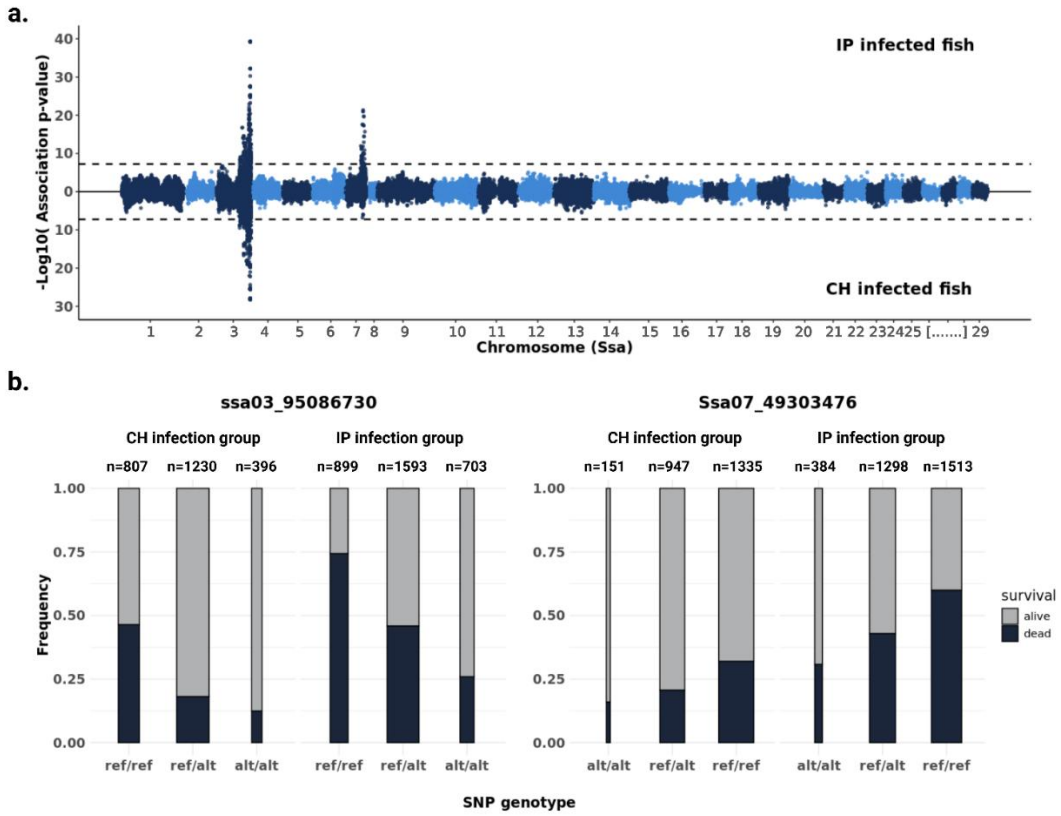


Figure 3. *(**a.**) Genome wide association analysis for resistance against PD using SAV3 challenge populations of virally injected (IP) fish and infected cohabitants (CH). Association significance is adjusted for multiple testing bias using the Bonferroni correction criterion while dashed lines indicate the genome wide significance threshold. (**b.**) Genotype frequencies of the two SNPs with highest association to PD resistance for in the QTL regions on Ssa03 (left) and Ssa07 (right). Width of bars is adjusted to the number of fish with the respective Reference (ref/ref) or alternative (alt/alt) homozygote and heterozygote (ref/alt) genotype for the Ssa03 and Ssa07 QTL top-SNPs, respectively*

Building on the GWAS results for PD survival, we further investigated the QTL regions on Ssa03 and Ssa07 and performed linkage disequilibrium (LD) analyses to identify narrow genomic regions containing haplotypes with SNP alleles most strongly associated with PD resistance. Analyses for the Ssa03 QTL region revealed a narrow LD block in position Ssa03:95,054,413 – 95,127,543 consisting of 58 SNPs. Of these, 10 SNPs were found in almost perfect LD with the most significantly associated variant from the GWAS, narrowing the most likely candidate QTL region to 45 Kb (Ssa03:95,076,092-95,121,176). Using the Ensembl gene annotation (release v112), we detected six genes overlapping the particular region (**Table 2**). Since only one of the six genes was

annotated with a gene name, a cross-search against publicly available databases was conducted. This identified the genes as 'NACHT, LRR and PYD domains-containing protein 3-like' (*nlrp3)*, three copies of a gene previously characterized as *gig1* [18] and *abcc6b.1*, respectively. No matching hit was available for the sixth identified gene (ENSSSAG00000108304), thereafter referred to as *sspd* (**Figure 4 and Table 2**).

Similar investigation of the Ssa07 QTL region revealed a 23.6Kb haplotype block (Ssa07:49,279,835 – 49,303,476) consisting of 18 SNPs. Selection of SNPs in almost perfect LD with the highest associated variant further narrowed the candidate region to 23.1 Kb (Ssa07:49,280,373 – 49,303,476), overlapping the two genes *scaf11 and slc38a2* (ENSSSAG00000015784 and ENSSSAG00000015905, respectively, **Table 2** and **Additional file 2: Supplementary Figure 1**).

Analysis of allele combinations within the haplotype blocks revealed in total 16 distinct haplotypes in the SAV3 challenge trial (**Additional file 1: Supplementary Table S2**). Application of a cutoff threshold (frequency > 0.2) retained two haplotypes for each QTL block. Association testing using the selected haplotypes and the same PD survival population as before yielded highly significant results (GWAS p-value < e-17). Despite the significant genotype and haplotype association results for both QTLs on Ssa03 and Ssa07, lack of significant association from independent analysis of the IP and CH infection groups revealed that only the Ssa03 QTL was significantly associated with PD resistance irrespective of infectious route. Based on these findings, subsequent analyses focused primarily on the narrowed QTL region on Ssa03.

Table 2. *Genes overlapping the regions associated with PD response. Haplotype analyses revealed two narrow regions with significant association to PD resistance on chromosomes Ssa03 and Ssa07. The two regions were overlapping six and two genes, respectively. The physical coordinates and name of each gene is shown in the table, together with the respective gene accession number based on the Ssal_v3.1 Ensembl and NCBI public annotations*

Chromosome	Location	Name	Gene accession (Ensembl)	Gene accession (NCBI)
Ssa03	95,060,336 - 95,079,624	*NACHT, LRR and PYD domains-containing protein 3-like*	ENSSSAG00000071002	LOC106601922
Ssa03	95,085,053 - 95,090,404	*Gig1-like copy 1*	ENSSSAG00000119632	LOC106596079
Ssa03	95,100,560 - 95,107,946	*Gig1-like copy 2*	ENSSSAG00000094181	LOC106602070
Ssa03	95,115,341 -	*Gig1-like*	ENSSSAG00000121001	LOC106602018

	95,117,818	*copy 3*		
Ssa03	95,112,232 - 95,187,058	*abcc6b.1*	ENSSSAG00000044528	LOC106601939
Ssa03	95,117,752 - 95,146,420	*sspd*	ENSSSAG00000108304	
Ssa07	49,266,424 - 49,292,752	*SCAF11*	ENSSSAG00000015784	LOC106609516
Ssa07	49,301,132 - 49,313,096	*slc38a2*	ENSSSAG00000015905	LOC106609517

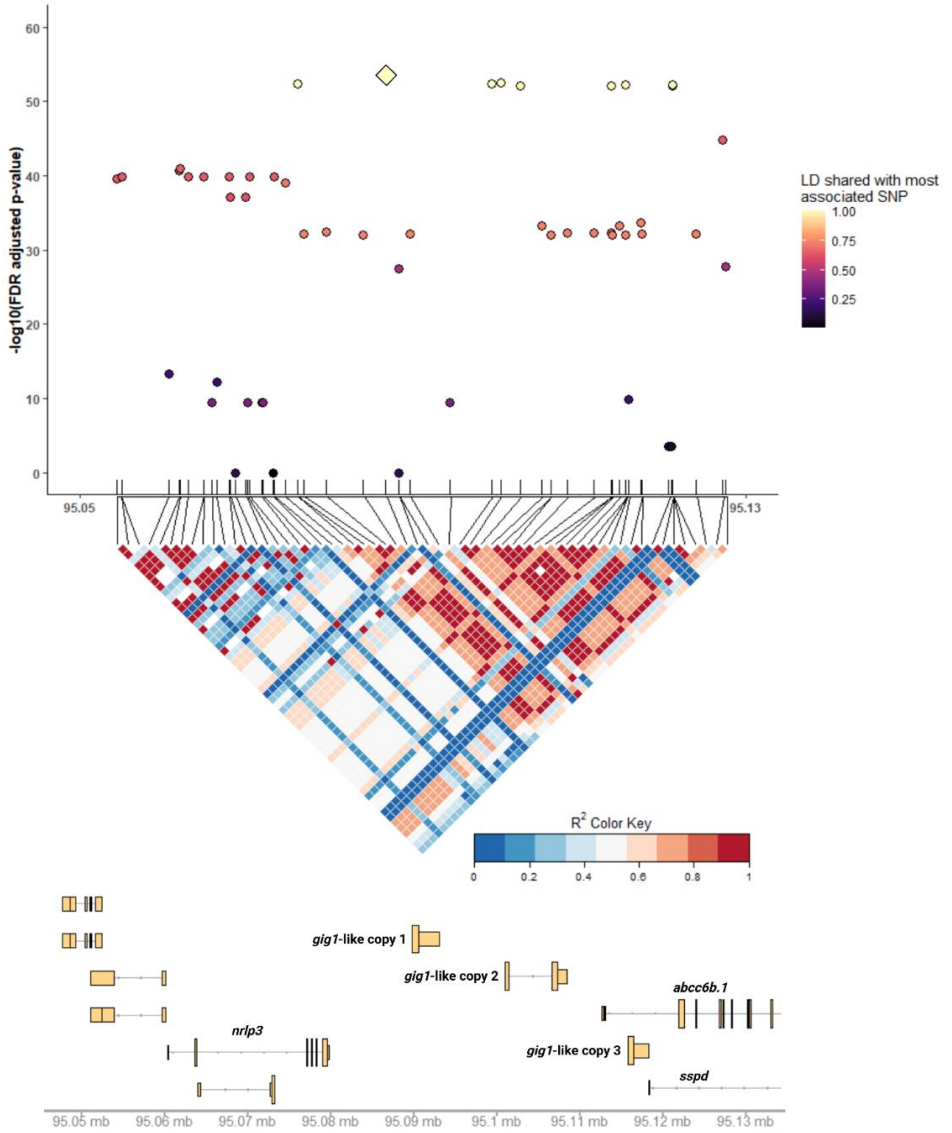


***Figure* 4.** *Linkage disequilibrium and haplotype block analyses within the Ssa03 QTL region. Each colored point represents an individual SNP within the Ssa03 QTL region for PD (x-axis) with statistical association to SAV3 resistance (y-axis), while the large diamond shape indicates the SNP with highest association, hereby referred to as the top-SNP. The color for each point represents the linkage disequilibrium (LD) shared between each SNP and the top-SNP of the Ssa03 QTL, while the LD plot below highlights a narrow genomic region of high LD. Underneath, the gene landscape overlapping the QTL region is shown, based on the Ensembl functional annotation (release 112). In particular, the narrow QTL region overlapped 6 genes,* namely *nlrp3*, three copies of *gig1-like*, *abcc6.b* and *sspd*

Transcriptome analyses and functional annotation highlight the importance of the gig genes within Ssa03 QTL region

To identify genes responsive to PD virus, we analysed gene expression data from two different experiments involving SAV3 viral infection. First, we studied expression patterns in PD infected fish with high or low viral concentration titre (viral load) following an infectious SAV3 cohabitation (CH) challenge. Differential gene expression analysis identified 14,712 genes with significant differences in expression profile between fish with high and low viral load (FDR adjusted p-value < 0.05). Of these, 7,225 genes were significantly upregulated in fish with low viral titre, whereas 7,487 genes were downregulated. Secondly, we reanalysed gene expression data from a previous study [17] that used SAV3 virally injected (IP) fish with survival records (dead or alive) following an four-week PD infectious challenge. This analysis identified 18,760 genes with significant expression differences between survivors and moribund fish in the four-week infectious trial. Of these, 10,024 genes were up-regulated in IP survivors, while 8,736 genes were down-regulated. Within the Ssa03 QTL region, both experiments revealed significant genome-wide expression differences for three *gig1-like* genes in response to SAV3 infection, indicating that these genes are likely candidates underlying the Ssa03 QTL (**Figures 4B and 4C**).

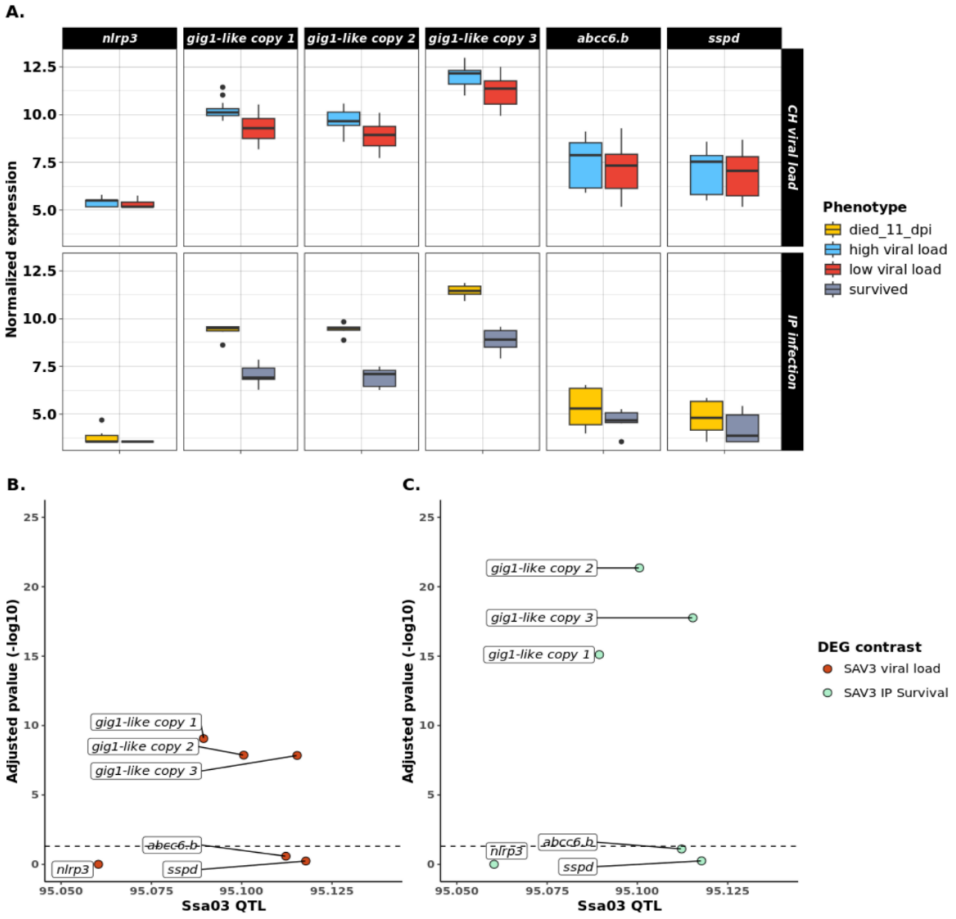


Figure 5. A. *Differences in expression pattern of genes overlapping the Ssa03 QTL region across two SAV3 infectious challenge studies. Top bins show gene expression of samples infected with high or low viral load during an of viral cohabitation (CH) challenge. Bottom bins show gene expression of SAV3 virally injected samples that survived or died after 11 days of a four-week challenge trial. X-axis represents the expression of each gene with colors indicating the respective phenotypes for each of the two studies. Y-axis shows the expression level of genes represented as the log2-normalized value of mapped reads for each gene.* ***B.*** *Differential gene expression analysis focusing on low/high viral load in CH infectious challenge individuals within Ssa03 QTL haplotype block. The physical position of genes is noted on x-axis whereas the y-axis represents the statistical significance of differential gene expression, expressed as the -log10 of the FDR adjusted p-value, with higher values indicating greater significance. Dashed lines represent the 0.05 FDR adjusted p-value cutoff point.* ***C.*** *Differential gene expression analysis focusing on death/survival in IP survival challenge individuals within Ssa03 QTL haplotype block. The physical position of genes is noted on x-axis whereas the y-axis represents the statistical significance of differential gene expression, expressed as the -log10 of the FDR adjusted p-value, with higher values indicating greater significance. Dashed lines represent the 0.05 FDR adjusted p-value cutoff point*

Comparison of the likely candidate genes for PD response between publicly available functional annotations (NCBI and Ensembl genome databases) revealed large structural differences between *gig1-like* gene copies (**Figure 6**). In particular, we identified differences in the number of exons for the first and third *gig1-like* gene copies between publicly available annotations as well as differences in the length of the annotated exons for the first and second *gig1-like* genes. In order to address and correct the structural discrepancies of *gig1-like* genes between the NCBI and Ensembl databases for the Ssal_v3.1 assembly, we obtained long-read transcriptome data from the head kidney of 12 healthy fish [58] and combined it with short-read transcriptome data from the SAV3 infectious cohabitation challenge to create a consensus annotation. This approach revealed gene predictions with high structural similarity across the three *gig1-like* genes (**Figure 6**) where each gene copy consisted of one small exon (98–136 bp) and one larger exon (2,230–3,310 bp), separated by a long intronic region (**Additional file 1: Supplementary Table S3**). Open reading frame (ORF) prediction for each *gig1-like* gene detected open reading frames on the second exon of all *gig1-like* genes, coding for proteins of 240, 222 and 222 amino acids for the first, second and third copies, respectively (**Additional file 1: Supplementary Table S3**).

Following the creation of a consensus annotation of the *gig1-like* genes, we aimed to elucidate the underlying regulatory landscape and identify likely regulators surrounding the *gig1-like* genes. This investigation utilized regulatory information generated using transposase-accessible chromatin assay (ATAC-Seq) as well as chromatin immunoprecipitation assay (ChIP) data from an experiment of IFN stimulation of Atlantic salmon head kidney tissue samples [59]. Mimicking viral infection, samples were treated with 100 μg of poly(I:C) and regulatory signals were compared against those of samples treated with control (PBS media) conditions (**Figure 6**). Investigation of regulatory genomic regions within the Ssa03 QTL region revealed both ATAC and H3K27ac peaks overlapping the first exons of the first and third *gig1-like* gene copies, indicating strong promoter activation in these two genes in response to IFN stimulation. In contrast, no such activation was observed for the second *gig1-like* gene copy (**Figure 6**).

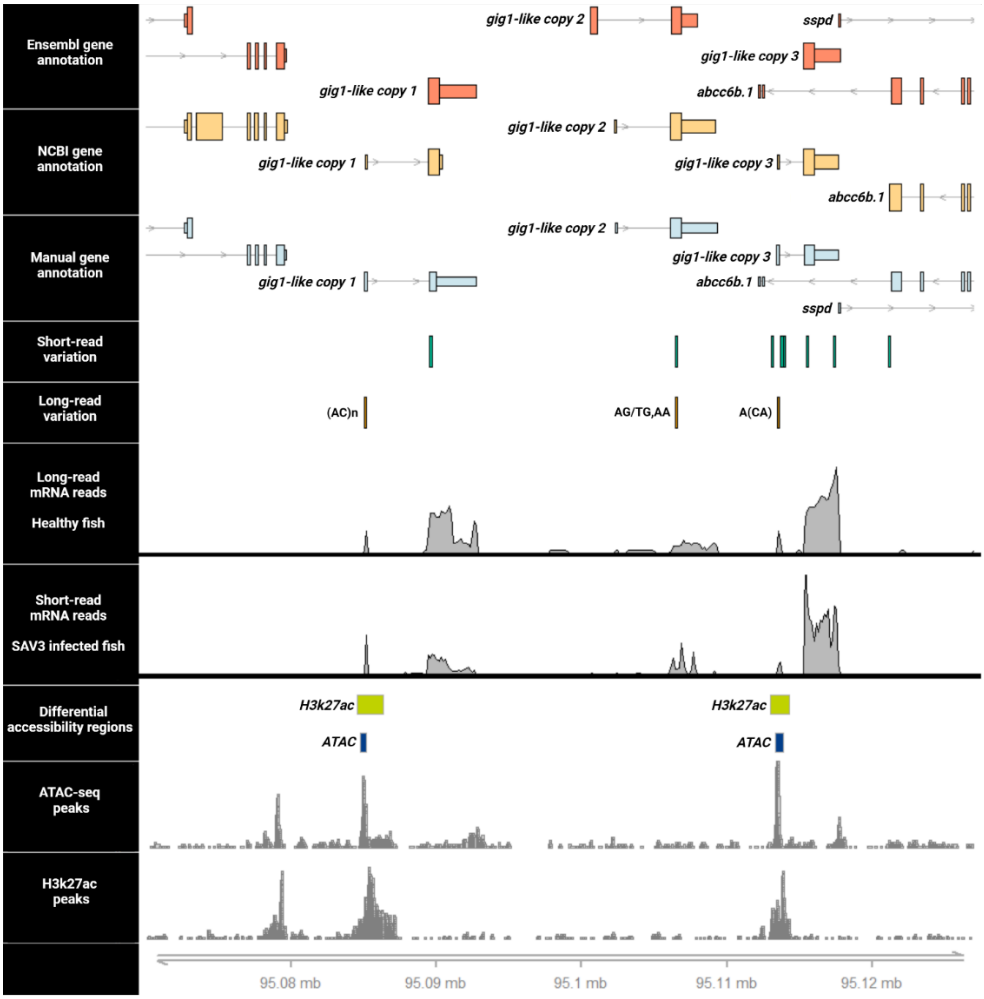


***Figure 6**. The putative functional variation landscape of the major QTL for PD resistance. From top to bottom, bins represent the gene landscape according to the Ensembl (release 112) and NCBI publicly available gene annotations for Atlantic salmon. Below, the manually curated gene landscape is shown, constructed based on long read and short read sequencing transcriptomics data of healthy and SAV3 infected individuals, respectively. Underneath, genetic variants segregating together with the Ssa03 QTL SNP were detected through short and long read sequencing technologies. Finally, the regulatory landscape of the Ssa03 QTL during interferon (IFN) stimulation was investigated to reveal transcriptionally active regions (ATAC-seq) as well as regions of histone modification (ChIP-seq). Within the first gig1-like copy, an (AC)n microsatellite intersects the coding region of the first exon as well as regions of accessible chromatin and histone acetylation marks, whereas another short variant overlaps the coding region of the second exon, within the 3' UTR region. One short variant overlaps the coding region of the second gig1-like copy together with a biallelic substitution (AG->TG, AA). Finally, for the third gig1-like gene copy, five variants located in regions of histone acetylation harboring the gene's*

promoter region and five additional variants overlap the coding sequence of the second exon, respectively. Of the variants proximal to the gene copy's promoter, two SNP variants and another, shorter, microsatellite (ACA<-A) additionally overlap with regions of open chromatin, while only the small deletion intersects with coding sequence.

To identify variation overlapping the functional and regulatory landscape within the Ssa03 QTL region we utilized SNPs and short indels called from the aforementioned set of 293 whole-genome short-read sequencing ancestors of the PD challenge individuals. Using these variants, we conducted a linkage disequilibrium (LD) analysis to identify those sharing substantial LD with the Ssa03 top-SNP for PD ($r^2 > 0.6$). Out of 115 detected variants only 23 shared substantial linkage disequilibrium with the top-SNP (**Additional file 1: Supplementary table S4**), while four of these overlapped accessible chromatin as well as histone acetylation marks near the promoter region of the third *gig1-like* copy (**Figure 6**). Additionally, one, one, and five of the SNPs in linkage phase with the top-SNP were located within the coding regions of the first, second, and third *gig1-like* gene copies, respectively. To identify SNPs potentially affecting *gig1-like* gene expression, we conducted an allele-specific expression (ASE) analysis. This analysis utilized seven SNPs and gene expression data from 10 fish with a heterozygous genotype for the Ssa03 QTL top-SNP. These samples were derived from the SAV3 infectious cohabitation (CH) challenge. The mRNA sequencing reads of heterozygous samples were uniquely mapped to each of the three *gig1-like* genes based on overlapping genetic variation and gene expression was quantified for each SNP allele (see Materials and Methods). The ASE analysis revealed transcriptional activity for all *gig1-like* copies in addition to higher gene expression for SNP alleles in linkage phase with the reference allele of the top-SNP for the QTL on Ssa03, linking higher *gig1-like* gene expression to lower PD resistance (**Figure 7**). Notably, significant differences between gene expression level and allele patterns were observed for SNPs within the first and third *gig1-like* gene copies (exact binomial test p-value< 0.01), while no significant gene expression differences were detected for the alleles of the SNP within the second copy. Investigation of the mutational consequences for SNPs with significant allele-specific influence on gene expression revealed three missense mutations for SNPs ssa03_95115562, ssa03_95115565 and ssa03_95115580. overlapping the third *gig1-like* gene. The three mutations resulted in amino acid alterations of three codons, specifically aAg/aTg (K to M amino acid), aGg/aAg (R to K amino acid) and gGt/gAt (G to D amino acid).

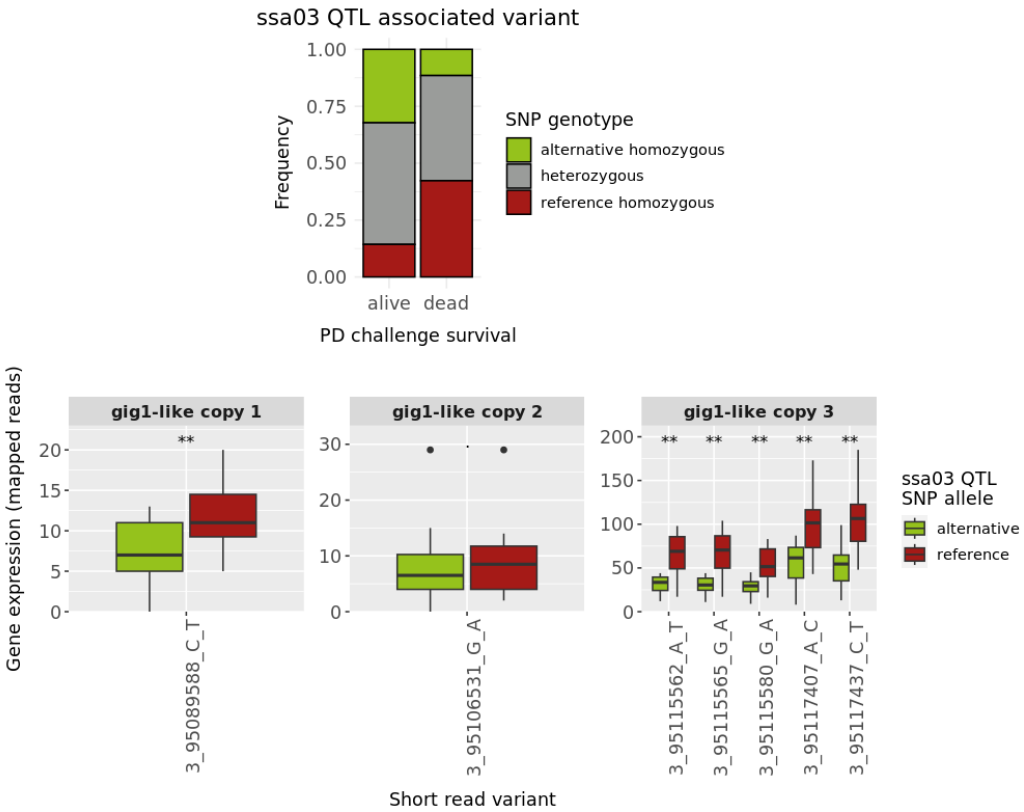


Figure 7. *Allele specific expression (ASE) analysis using transcriptomics data from 10 samples from a SAV3 infectious cohabitation challenge. Samples were selected based on their genotype (heterozygous) for the highest disease-associated SNP (top-SNP) within the Ssa03 QTL region. ASE analysis detected 7 biallelic variants whose physical position overlapped transcribed sequences of the three gig1-like gene copies and alleles shared high linkage disequilibrium across 293 ancestors of the population used in the PD GWAS study (LD > 0.6). In total, 1, 1 and 5 SNPs overlapped the first, second and third copy of the gig1-like gene, respectively. Asterisks across the bottom panels represent variants whose allele phase resulted in significant gene expression differences based on an exact binomial test (p-values < 0.01).*

Generating a detailed description of the genomic variation within the Ssa03 QTL region

While short-read sequencing offered valuable insight into the genomic landscape of the Ssa03 QTL in our study, the read lengths of this technology (approx. 120 bp) prohibit the detection and genotyping of large and more complex variation. In order to comprehensively identify all genomic variation underlying the Ssa03 QTL for PD response, including long as well as complex variants overlapping functional and regulatory features, we utilized whole-genome long-read nanopore sequencing data from seven salmon fish that originated from the same aquaculture strain as the GWAS

population in this study (AquaGen AS). Variant detection across the seven long-read individuals identified 446 variants within the QTL region (Ssa03: 95,076,090-95,121,178), consisting of 127 SNPs as well as 319 variants of variable complexity including multiallelic variants, microsatellites, tandemly repeated sequence and structural variants (SVs). To assess allele segregation patterns of these variants, disease-associated haplotypes were constructed for the seven fish, classified as either reference (8 haplotypes) or alternative (6 haplotypes) based on the alleles of the 11 SNPs sharing near perfect LD within the Ssa03 haplotype block (**Figure 4**). Then, using the SNP with highest association to PD resistance from our GWAS study (top-SNP, ssa03_95086730), we assessed the allele segregation of all variants across the 14 haplotypes, retaining variation in near perfect linkage phase with the reference and alternative top-SNP alleles. Allele segregation analysis identified in total 92 variants including 49 SNPs, 10 insertions, 7 deletions, and 9 microsatellites (**Additional file 1: Supplementary table S4**). Finally, comparison of the physical position for filtered variants with the functional and regulatory landscape of Ssa03 QTL revealed 25, 8 and 10 variants intersecting the three *gig1-like* gene copies (**Figure 7 and Additional file 1: Supplementary table S4**). Notably, a highly variable (AC)n microsatellite intersected the first exon of the first *gig1-like* gene copy in addition to regions of accessible chromatin and histone acetylation marks whereas another shorter microsatellite sequence (ACA/A) overlapped the respective coding and regulatory regions of the third *gig1-like* gene (**Figure 6**).

Gig1-like gene phylogenetic analysis

To elucidate the origin of the three *gig1-like* genes on Ssa03, we utilized the protein sequence of the manually annotated *gig1-like* genes together with the protein sequence of a *gig1-like* gene copy located on the ohnolog region of Ssa06 (ENSSSAG00000003892, Ssa06:5,558,971-5,559,951). The amino acid (aa) sequence alignment across the four *gig1-like* genes revealed relatively high similarity among the Ssa03 gene copies, whereas the sequence of the Ssa06 gene was notably shorter, diverging largely between the 191st – 240th aa (**Figure 8a**). Focusing on the amino acid sequence alignment of the Ssa03 *gig1-like* genes, two deletions of similar size were identified in the second and third copies compared to the first *gig1-like* gene, in particular one 10 aa and one 8 aa deletion at positions 188 and 214, respectively. To determine whether the observed amino acid differences conferred functional divergence among the Ssa03 gene copies, a functional domain prediction analysis was performed using InterPro [60]. Domain prediction identified the PANTHER domain '*SI:CH211-198C19.1-RELATED*' across all genes (PTHR38706, **Additional file 1: Supplementary Table S3**), likely indicating the conserved coding region for the *gig1-like* protein. However, no additional domains were identified within the regions showing the largest structural divergence between *gig1-like* gene copies (189 – 240 aa, **Figure 8a**).

Finally, a phylogenetic analysis of the four *gig1-like* genes' coding sequence was carried out in order to explore the evolutionary history of the Ssa03 copies. In addition to the four salmonid genes, the gene ortholog of *Esox lucius,* a close ancestor species that did not undergo the salmonid-specific whole genome duplication, was included as an outgroup (ENSELUG00000010316, Materials and methods). Phylogenetic analysis indicated that the *gig1-like* gene located on Ssa06 and the first copy on Ssa03 shared higher similarities, whereas the second and third copies on Ssa03 likely arose through segmental duplications of this variant. (**Figure 8b**).

a.

```
ssa03_gig1-like_copy1       MVRTLNNMEELRASRFGRPSPRHGLKLLFWFANDYIVFDDDNQMVADYDPKDGDFGFHHF 60
ssa03_gig1-like_copy2       MVRTLNNMEELRASRFGRPSPRHGLKLLFWFANDYIVFDDDNQMVADHDPKEGDFGFHHF 60
ssa03_gig1-like_copy3       MVRTLNNMEELRASRFGRPSPRHGLKLLFWFANDYIGFDDDNQMFADHDPKEGDFGFHRF 60
ssa06_gig1-like             MVRTLNNMAELRGSRFGRPWPRHGLKLLFWFAKDYIVFDNDNQMVADHDPKGGGFGFHHF 60
                            ******** ***.****** ***********:*** **:****.**:*** *.****:*

ssa03_gig1-like_copy1       RNRLECENNVCKKLLPDGGYPFYEVGNLHLTASDSMPKYVRKYNTGNIDTSNMDRLIISM 120
ssa03_gig1-like_copy2       QNRRECENNVCKRLLPYDGYPFYEVGNLHLTASDSMPEYVSEDYAGQIDNSNMDRLIISM 120
ssa03_gig1-like_copy3       LNRRECENNVCKRLLPDGGCPFYEVGNLHLTASDSMPKYVRKYNTGNIDTSNMDRLIISM 120
ssa06_gig1-like             CNRRECENNVCMRLLPDDGYPFYEVGNLHLTASDSMPKYVRKYNTGNIDTSNMDRLIISM 120
                             ** ******* :*** .* *****************:** :  :*:**.**********

ssa03_gig1-like_copy1       RPDMTVDKVYVTQHEDLRSFDPVNTYCISRGLLMIICGHSFADMSFRNFLEQAGYSTHEP 180
ssa03_gig1-like_copy2       RPDMTVDKVYVTQHEDLRSFDPVNTYRISRGLLMSICGHPSADMSLSNFLEQAGCPTHEP 180
ssa03_gig1-like_copy3       RPDMTVDKVYVTQHEDLRSFDPVNTYRISRGLLMIICGHSSADMSLSNFLEQAGCPTHEE 180
ssa06_gig1-like             RPDRTVDKVYVTQHEDLRSFDPVNTYCISRGLLMIICGHPSADMSLSNFLEQAGYSTHEP 180
                            *** ********************** ******* ****. ****: ******* .***

ssa03_gig1-like_copy1       MGYIYQSSSPAQIQFPSPAPRESRDTRMNIAVGSRVPPRAAPRVPPRAAPGFWENYCTIL 240
ssa03_gig1-like_copy2       NLITHQGI----------APRESRDTKIDIEGGS--------RVPPRAAPGFWESYCTIL 240
ssa03_gig1-like_copy3       NVMMYQGI----------AARESRDTKIDIEGGS--------RVPPRAAPGFWESYCTIL 240
ssa06_gig1-like             MRYMYQSSSP--------LANEH----------------------------WLSWLA-- 240
                                :*.           ..*                              * .: :
```

b.

Figure 8. (a) *Multiple amino acid of the three manually annotated gig1-like genes on chromosome 3 (Ssa03) and the gig1-like copy located in the ohnologous region on Ssa06. Alignment was carried out using MAFFT v7.511 and default settings. Symbols underneath each peptide represent peptide conservation with asterisk and colon symbols indicating high and moderate positional conservation, respectively.* ***(b)*** *Phylogenetic analysis of coding sequence for the Atlantic salmon Ssa03 and Ssa06 gig1-like genes. In order to reveal evolutionary relationships across gene paralogs, the gene ortholog of the ancestor species Esox lucius was included as an outgroup (Elu11). The coding sequence of the five genes was aligned using MAFFT v7.511 and a neighbor-joining (NJ) tree was constructed using the gap-free and moderately conserved alignment sites. Numbers on the tree branches indicate the bootstrap support values estimated using 500 replicates*

Discussion

Pancreas disease (PD) represents a significant challenge for Atlantic salmon aquaculture as viral outbreaks can have serious implications on animal welfare, production costs, and the reputation of the industry. Understanding the genetic basis of PD resistance is important for developing sustainable and efficient disease management strategies that improve fish welfare, minimize losses, and ensure the long-term sustainability of production. While previous studies have identified several genomic loci associated with PD response, limitations in genomic resources have hindered the fine mapping and thorough investigation of putatively causal variation.

In the present study, we used 938,582 SNPs in order to conduct a high-powered GWAS in a population of fish infected with the SAV3 virus either through peritoneal injection or infectious cohabitation (IP and CH, respectively). Increasing the SNP density of the association study population through genotype imputation allowed us to narrow the Ssa03 and Ssa07 QTL regions from several megabases to 45 and 23 Kb, respectively (**Figures 3 and Additional file 2: Supplementary Figure 1**).

Interestingly, differences in viral transmission (IP vs. CH infection) influenced the association signals of the QTL on Ssa07, limiting the locus' significance to only IP infected fish (**Figure 3**). Peritoneal (IP) infection constitutes an artificial method of viral delivery that stimulates the host's immune response bypassing any physical immunity barriers [67, 68]. In contrast, during natural PD outbreak scenarios, viral transmission occurs through contact of the fish with water-suspended viral particles released from infected or decomposing cohabitants [4, 69]. In this context, the observed differences between the IP and CH infection groups suggest that varying infection methods may induce distinct immune responses. Supporting this, previous studies examining the impact of different infection models on the response of rainbow trout (*Oncorhynchus mykiss)* to *Yersinia ruckeri* have revealed significant differences in inflammatory (IL-1b1 and IL-8) and anti-microbial (Cath-1 and Cath-2) responses between IP injected fish and infected cohabitants [70]. The use of suboptimal experimental models - and the comparison of results across studies employing different infection methods - can introduce bias into association studies, potentially hindering the detection of disease-resistance loci and, by extension, the discovery of causal variation [71]. Collectively, these findings highlight the importance of selecting representative infection models to accurately characterize and interpret immune responses in aquaculture settings, where infectious disease outbreaks can significantly impact both production efficiency and animal welfare.

Our study focused on the PD associated region on Ssa03, narrowing the position of this QTL to 45kb (Ssa03: 95,054,413 – 95,127,543). Based on the Ensembl gene annotation, the region encompasses the genes *nlrp3, abcc6.b, sspd,* as well as

three tandemly duplicated copies of the *gig1-like* gene, suggesting that these genes play important roles in PD resistance. Although several of these genes have been previously reported as likely candidates for PD resistance [70], transcriptomics analysis of fish with quantified SAV3 viral concentration, as well as re-analysis of data from a SAV3 IP survival challenge [71], showed that only the three *gig1-like* genes responded significantly to PD infection within the QTL region (**Figures 4B and 4C**). Consistency in the gene expression findings across studies sheds light on the functional landscape of PD resistance, highlighting *gig1-like* gene copies as the most probable causal candidates.

Grass carp reovirus (*GCRV*)-induced gene 1 (*gig1*) is a fish-specific gene family belonging to interferon (IFN) stimulated genes, that plays an important role in antiviral response [72]. In Atlantic salmon, multiple *gig1* genes have been identified [73], whereas expression of these genes has previously been associated with response to aquaculture diseases, including PD [14, 74, 75]. Although *gig1* has been shown to suppress GCRV viral replication [76, 77], overexpression of *gig1-like* genes in Atlantic salmon is often linked to poor immunity outcomes for virally infected fish [10, 74]. Consistently, our findings showed significantly higher *gig1-like* gene expression in fish with higher SAV3 viral load and greater disease susceptibility (**Figure 5A**). Although *gig1-like* genes are induced through an IFN-mediated immune response [72, 73, 77], the observed combination of increased disease susceptibility together with higher gene expression in Atlantic salmon suggests that *gig1-like* genes may serve as indicators of viral replication rather than viral suppression. This points to a potential link between an elevated IFN response and heightened disease susceptibility [75, 78]. IFN stimulation is a key feature of antiviral immunity, however, the negative impact of prolonged IFN response [79] has been previously discussed in the context of aquaculture disease susceptibility [80, 81]. Based on this hypothesis, changes in the transcription of *gig1-like* genes could affect gig1 effectiveness, leading to a prolonged and ultimately harmful IFN response. Furthermore, the presence of multiple gene copies within the narrow QTL region could introduce additional complexity in *gig1-like* antiviral responses, potentially by altering the regulation of these genes coupled with variations in the transcribed protein [82] .

To explore functional differences between the three *gig1-like* gene copies, we used both short and long read sequencing data and confirmed the genes' structure and protein products (**Figures 5, 7 and Table S2**). Protein sequence alignments of the *gig1-like* genes on Ssa03, together with the gene paralog located on the duplicated region of Ssa06 [83], revealed substantial differences between the *gig1-like* genes across the duplicated regions (**Figure 8**). Although the same protein domain was predicted across the four *gig1-like* paralogs, absence of a QTL locus for PD on Ssa06 as well as absence of transcriptomic response suggests that the particular gene copy does not constitute a strong candidate for PD resistance.

Finally, to better understand how the *gig1-like* genes respond to viral infection, we performed an allele specific expression (ASE) analysis using short-read sequencing variation detected in 293 ancestors of the GWAS fish and focused on SNPs located in transcribed regions of the three *gig1-like* gene copies (**Figure 7**). Of these, three SNPs represented missense mutations, leading to amino-acid changes within the predicted functional domain of the third *gig1-like* gene copy (peptides 72,73 and 78, **Figure 8a**). Location of the SNPs within the gene's predicted functional domain increase their impact and functional importance, particularly in respect to amino acid changes inside highly conserved peptides (**Figure 8a**). However, since *gig1-like* genes are not thoroughly studied yet, the functional consequences of the detected amino acid changes remain elusive.

Moreover, we focused on the regulatory regions within the QTL region for PD and investigated variations with potential functional and regulatory impact on the three *gig1-like* gene copies. Using long-read sequencing information from seven aquaculture fish to construct disease associated haplotypes revealed large and complex variation segregating together with the top-SNP for PD resistance in the QTL region (**Additional file 1: Supplementary table S4**). Interestingly, an AC(n) microsatellite located on the first *gig1-like* copy, and another shorter microsatellite ACA/A on the third *gig1-like* copy overlapped transcription start sites with regulatory activity during viral infection response (**Figure 6**). Comparison of the alleles for these two variants across disease associated haplotypes revealed a link between allele length and disease susceptibility (**Additional file 1: Supplementary table S4**). Several studies have described the impact of repeat sequence expansion on gene expression [84, 85]. In our study, the overlap between the microsatellite and the promoter region of the first *gig1-like* gene copy, along with additional overlap with regulatory elements, supports the hypothesis that expansions of this particular variant could significantly impact the regulation and expression of the gene. Furthermore, detection of a similar, yet shorter (AC) microsatellite on the third *gig1-like* gene copy could indicate the presence of similar variation occurring across the particular genes. However, the limited number of long-read sequenced individuals in our study restricted further analysis of these variants on *gig1-like* gene expression, underscoring the need for follow up studies to fully understand the mechanisms underlying PD resistance.

As highlighted above, further investigation is needed to validate the functional impact of the three *gig1-like* gene copies and the putative causal variants underlying PD resistance. In our analysis, the high sequence similarity within the PD associated region on Ssa03 posed challenges for short-read sequencing technologies, particularly for detecting complex variants and accurately assigning transcriptomic reads among *gig1-like* gene copies. In this context, targeted long-read sequencing of fish with phenotypic records on PD resistance could provide stronger evidence linking complex genetic variants, such as the polymorphic (AC)n microsatellites, to PD response [86]. Similarly,

applying long-read sequencing methods to transcriptomic analyses could yield a more accurate representation of the functional landscape underlying disease resistance [86]. In addition, gene editing approaches, including knock-in or knock-out experiments, as well as gene overexpression experiments *in vitro*, could be used to validate the functional impact of repeat length expansions in regulation of *gig1-like* gene expression [87]. Furthermore, sequential inactivation of the *gig1-like* copies in SAV3 infected cells could uncover potential regulatory interactions among the three tandem duplicates and shed light on their role in mediating disease response [82]. Investigating the relationship between the *gig1-like* gene expression and PD resistance, along with the effects of the putatively causal variants, could significantly advance our understanding of the molecular mechanisms underlying PD response. Incorporating this knowledge into breeding programs may strengthen disease management strategies, ultimately enhancing animal welfare and promoting the sustainability of aquaculture.

Conclusions

In conclusion the present study provides a comprehensive analysis of the genomic and functional landscape underlying response to PD, advancing our understanding of the genetic basis of PD resistance. Our study confirmed a major QTL in chromosome Ssa03 and a secondary region on chromosome Ssa07 with the latter locus addressing infection-specific interactions with the host. Transcriptomics analysis of viral load and disease susceptibility within the major QTL for PD resistance on Ssa03 revealed three in-tandem duplicated copies of the *gig1-like* gene as the likely causal gene candidates. Increased expression of these genes in fish with poor disease outcomes suggests that the *gig1-like* copies serve as indicators of SAV3 viral replication levels in infected fish, emphasizing their potential role in aquaculture. Using the latest advances in Atlantic salmon genomics we elucidated the genomic and functional landscape underlying these genes, uncovering variation with putative impact on coding as well as regulatory regions of *gig1-like* genes. Notably, two polymorphic microsatellites overlapped the transcription start of the first and third *gig1-like* copies, respectively, that overlapped transcriptionally active regulatory regions during anti-viral response. Use of long-read sequencing technologies revealed a link between microsatellite allele length and disease susceptibility, underscoring the importance of using advanced genomic technologies to unravel the complex interplay between genetic variation and response to traits with great importance to aquaculture sustainability.

Declarations

Ethics approval and consent to participate

All Infectious challenge trials described in the study were approved beforehand by the Norwegian National Animal Research Authority (FOTS). All aspects of the challenges followed the guidelines of VESO Vikan Standard Operating Procedures for conduction of fish experiments. Informed consent of participation for all privately owned fish used in this study was obtained by their respective owners.

Consent for publication

Not applicable

Availability of data and materials

The raw genotype, phenotype and whole-genome sequencing data used in this study is property of a commercial enterprise and therefore not publicly available. Reasonable request for access to the raw material should be directed to the corresponding author. Raw material for the analysis of the Atlantic salmon regulatory landscape is available on the European Nucleotide Archive under accession numbers PRJEB50077 and PRJEB56698 for ATAC-seq and ChIP-seq information, respectively. Raw short-read transcriptomics data related to the SAV3 infectious cohabitation challenge is available on the European Nucleotide archive under accession number PRJEB85594. Raw transcriptomics data from the re-analysis of a SAV3 infectious survival challenge are available on the European Nucleotide Archive under accession number PRJNA590166. Additional data supporting the findings of the current study are available upon request to the corresponding author.

Competing interests

The authors claim no competing interests.

Funding

The study was supported by The Research Council of Norway grant no. 325874 and the Norwegian University of Life Sciences. The funding agencies played no role in the study conceptualization, design, analysis, decision to publish or preparation of the manuscript. Finally, open access funding was provided by Norwegian University of Life Sciences.

Authors' contributions

Sigbjørn Lien and Marie Saitou conceived and designed the study. Thomas Moen provided genotype, phenotype and transcriptomics data related to SAV3 challenges, short-read whole genome resequencing data from a commercial population as well as RNA material and genotype information of samples participating in a SAV3 infectious

cohabitation challenge. Fabian Grammes and Tim M. Knutsen performed bioinformatics analyses and assisted in interpretation of results. Shahmir Naseer and Samuel AM Martin provided all information related to the Atlantic salmon regulatory landscape during antiviral response, produced through ATAC-seq and ChIP-seq technologies. Domniki Manousi performed the bioinformatics and statistical analyses related to the short read whole-genome sequencing, SAV3 challenge genotype and transcriptomics data and composed the first draft of the manuscript. Dorota M. Jaskula performed bioinformatics analyses related to the long-read sequencing data. All authors contributed to the interpretation of results and improvement of the resulting publication. The authors have read and approved of the final manuscript.

Acknowledgements

The authors acknowledge the Orion High Performance Computing Center (OHPCC) at the Norwegian University of Life Sciences (NMBU) for providing the required computational resources that have contributed to the research results reported in this manuscript. In addition, the authors acknowledge the use of BioRender for the creation of Figure 1 and the visual improvement of Figures 3,4,6 and 8 in this manuscript.

Supplementary Materials

Supplementary materials for this manuscript are available at:

material description

supplementary figures

supplementary tables

List of abbreviations

SAV: Salmonid alphavirus
SAV[1-3]: Subtype 1-7 of the Salmonid Alphavirus
PD: Pancreas disease
IP: Peritoneal injection
CH: Cohabitation
dpi: Days post-infection
QTL: Quantitative trait locus
SNP: Single nucleotide polymorphism
SV: structural variant
GWAS: Genome-wide association study
LD: Linkage disequilibrium
Mb: Mega base-pair
Kb: Kilo base-pair
TPM: Transcripts per million mapped reads
WGS: whole genome sequence
GRM: Genomic relationship matrix
PC: Principal component
qPCR: quantitative Polymerase chain reaction
ONT: Oxford Nanopore Technologies
FDR: False discovery rate
DEG: Differential gene expression
ASE: allele-specific expression
ORF: Open reading frame
IFN: Interferon
HK: head kidney
GCRV: Grass carp reovirus
nlrp3: NACHT, LRR and PYD domains-containing protein 3-like
gig1: Grass carp reovirus-induced gene 1
abcc6b.1: ATP-binding cassette, sub-family C, member 6b, tandem duplicate 1
sspd: salmo salar pancreas disease
slc38a2: solute carrier family 38 member 2
scaff11: SR-related CTD-associated factor 11
mRNA: messenger ribonucleic acid
DNA: Deoxyribonucleic Acid
aa: amino acid

References

1. FAO. The State of World Fisheries and Aquaculture 2020: Sustainability in action. Food and Agriculture Organization of the United Nations; 2020.

2. Jansen MD, Jensen BB, Brun E. Clinical manifestations of pancreas disease outbreaks in Norwegian marine salmon farming - variations due to salmonid alphavirus subtype. J Fish Dis. 2015;38:343–53.

3. McLoughlin MF, Nelson RN, McCormick JI, Rowley HM, Bryson DB. Clinical and histopathological features of naturally occurring pancreas disease in farmed Atlantic salmon, Salmo salar L. J Fish Dis. 2002;25:33–43.

4. Jansen MD, Bang Jensen B, McLoughlin MF, Rodger HD, Taksdal T, Sindre H, et al. The epidemiology of pancreas disease in salmonid aquaculture: a summary of the current state of knowledge. J Fish Dis. 2017;40:141–55.

5. Pettersen JM, Brynildsrud OB, Huseby RB, Rich KM, Aunsmo A, Bang BJ, et al. The epidemiological and economic effects from systematic depopulation of Norwegian marine salmon farms infected with pancreas disease virus. Prev Vet Med. 2016;132:113–24.

6. Gao S, Liu X, Han B, Wang N, Lv X, Guan X, et al. Salmonid alphavirus non-structural protein 2 is a key protein that activates the NF-κB signaling pathway to mediate inflammatory responses. Fish Shellfish Immunol. 2022;129:182–90.

7. Reid KM, Patel S, Robinson AJ, Bu L, Jarungsriapisit J, Moore LJ, et al. Salmonid alphavirus infection causes skin dysbiosis in Atlantic salmon (Salmo salar L.) post-smolts. PLoS One. 2017;12:e0172856.

8. Pettersen JM, Rich KM, Jensen BB, Aunsmo A. The economic benefits of disease triggered early harvest: A case study of pancreas disease in farmed Atlantic salmon from Norway. Prev Vet Med. 2015;121:314–24.

9. Skjold P, Sommerset I, Frost P, Villoing S. Vaccination against pancreas disease in Atlantic salmon, Salmo salar L., reduces shedding of salmonid alphavirus. Vet Res. 2016;47:78.

10. Robinson NA, Krasnov A, Burgerhout E, Johnsen H, Moghadam HK, Hillestad B, et al. Response of the Salmon Heart Transcriptome to Pancreas Disease: Differences Between High- and Low-Ranking Families for Resistance. Sci Rep. 2020;10:868.

11. Thorarinsson R, Wolf JC, Inami M, Phillips L, Jones G, Macdonald AM, et al. Effect of a novel DNA vaccine against pancreas disease caused by salmonid alphavirus subtype 3 in Atlantic salmon (Salmo salar). Fish Shellfish Immunol. 2021;108:116–26.

12. Houston RD. Future directions in breeding for disease resistance in aquaculture species. R Bras Zootec. 2017;46:545–51.

13. Gonen S, Baranski M, Thorland I, Norris A, Grove H, Arnesen P, et al. Mapping and validation of a major QTL affecting resistance to pancreas disease (salmonid alphavirus) in Atlantic salmon (Salmo salar). Heredity (Edinb). 2015;115:405–14.

14. Aslam ML, Robledo D, Krasnov A, Moghadam HK, Hillestad B, Houston RD, et al. Quantitative trait loci and genes associated with salmonid alphavirus load in Atlantic salmon: implications for pancreas disease resistance and tolerance. Sci Rep. 2020;10:10393.

15. Hillestad B, Makvandi-Nejad S, Krasnov A, Moghadam HK. Identification of genetic loci associated with higher resistance to pancreas disease (PD) in Atlantic salmon (Salmo salar L.). BMC Genomics. 2020;21:388.

16. Johnston IA, Kent MP, Boudinot P, Looseley M, Bargelloni L, Faggion S, et al. Advancing fish breeding in aquaculture through genome functional annotation. Aquaculture. 2024;583:740589.

17. Clark TC, Naseer S, Gundappa MK, Laurent A, Perquis A, Collet B, et al. Conserved and divergent arms of the antiviral response in the duplicated genomes of salmonid fishes. Genomics. 2023;115:110663.

18. Sandholm RM. Evolution of alternative splice variation and exon usage following whole genome duplication. Norwegian University of Life Sciences, Ås; 2022.

19. Faust GG, Hall IM. SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics. 2014;30:2503–5.

20. Vasimuddin M, Misra S, Li H, Aluru S. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. In: 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). 2019. p. 314–24.

21. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10.

22. Van der Auwera GA, O'Connor BD. Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. "O'Reilly Media, Inc."; 2020.

23. Yun T, Li H, Chang P-C, Lin MF, Carroll A, McLean CY. Accurate, scalable cohort variant calls using DeepVariant and GLnexus. Bioinformatics. 2021. https://doi.org/10.1093/bioinformatics/btaa1081.

24. Pedersen BS, Brown JM, Dashnow H, Wallace AD, Velinder M, Tristani-Firouzi M, et al. Effective variant filtering and expected candidate variant yield in studies of rare human disease. NPJ Genom Med. 2021;6:60.

25. Yoshida GM, Yáñez JM. Increased accuracy of genomic predictions for growth under chronic thermal stress in rainbow trout by prioritizing variants from GWAS using imputed sequence data. Evol Appl. 2022;15:537–52.

26. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015;4:7.

27. Browning BL, Tian X, Zhou Y, Browning SR. Fast two-stage phasing of large-scale sequence data. Am J Hum Genet. 2021;108:1880–90.

28. Browning BL, Zhou Y, Browning SR. A One-Penny Imputed Genome from Next-Generation Reference Panels. Am J Hum Genet. 2018;103:338–48.

29. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.

30. Yang J, Bakshi A, Zhu Z, Hemani G, Vinkhuyzen AAE, Lee SH, et al. Genetic variance estimation with imputed variants finds negligible missing heritability for human height and body mass index. Nat Genet. 2015;47:1114–20.

31. Loh P-R, Tucker G, Bulik-Sullivan BK, Vilhjálmsson BJ, Finucane HK, Salem RM, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet. 2015;47:284–90.

32. Studio R. Integrated development environment for R. Boston, MA: R Studio Inc. 2018.

33. Hamazaki K, Iwata H. RAINBOW: Haplotype-based genome-wide association study using a novel SNP-set method. PLoS Comput Biol. 2020;16:e1007663.

34. Shim H, Chasman DI, Smith JD, Mora S, Ridker PM, Nickerson DA, et al. A multivariate genome-wide association analysis of 10 LDL subfractions, and their response to statin treatment, in 1868 Caucasians. PLoS One. 2015;10:e0120758.

35. Gabriel SB, Schaffner SF, Nguyen H, Moore JM, Roy J, Blumenstiel B, et al. The structure of haplotype blocks in the human genome. Science. 2002;296:2225–9.

36. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007;81:559–75.

37. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005;21:263–5.

38. Utsunomiya YT, Milanesi M, Utsunomiya ATH, Ajmone-Marsan P, Garcia JF. GHap: an R package for genome-wide haplotyping. Bioinformatics. 2016;32:2861–2.

39. Utsunomiya YT, Milanesi M, Barbato M, Utsunomiya ATH, Sölkner J, Ajmone-Marsan P, et al. Unsupervised detection of ancestry tracks with the GHaprpackage. Methods Ecol Evol. 2020;11:1448–54.

40. Zhou X. A UNIFIED FRAMEWORK FOR VARIANCE COMPONENT ESTIMATION WITH SUMMARY STATISTICS IN GENOME-WIDE ASSOCIATION STUDIES. Ann Appl Stat. 2017;11:2027–51.

41. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics. 2018;34:i884–90.

42. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013;29:15–21.

43. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2013;30:923–30.

44. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. Genome Biol. 2019;20:257.

45. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15:550.

46. R Core Team. R: A language and environment for statistical computing. 2021.

47. Irizarry RA. Ggplot2. In: Introduction to Data Science. Boca Raton: Chapman and Hall/CRC; 2024. p. 107–25.

48. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27:2156–8.

49. van de Geijn B, McVicker G, Gilad Y, Pritchard JK. WASP: allele-specific software for robust molecular quantitative trait locus discovery. Nat Methods. 2015;12:1061–3.

50. Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, et al. The Atlantic salmon genome provides insights into rediploidization. Nature. 2016;533:200–5.

51. Castel SE, Levy-Moonshine A, Mohammadi P, Banks E, Lappalainen T. Tools and best practices for data processing in allelic expression analysis. Genome Biol. 2015;16:195.

52. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.

53. Li H. New strategies to improve minimap2 alignment accuracy. Bioinformatics. 2021;37:4572–4.

54. Shafin K, Pesout T, Chang P-C, Nattestad M, Kolesnikov A, Goel S, et al. Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. Nat Methods. 2021;18:1322–32.

55. Kolmogorov M, Bickhart DM, Behsaz B, Gurevich A, Rayko M, Shin SB, et al. metaFlye: scalable long-read metagenome assembly using repeat graphs. Nat Methods. 2020;17:1103–10.

56. Hickey G, Monlong J, Ebler J, Novak AM, Eizenga JM, Gao Y, et al. Pangenome graph construction from genome alignments with Minigraph-Cactus. Nat Biotechnol. 2024;42:663–73.

57. Tang AD, Soulette CM, van Baren MJ, Hart K, Hrabeta-Robinson E, Wu CJ, et al. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. Nat Commun. 2020;11:1–12.

58. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6.

59. Wheeler DL, Church DM, Federhen S, Lash AE, Madden TL, Pontius JU, et al. Database resources of the National Center for Biotechnology. Nucleic Acids Res. 2003;31:28–33.

60. Paysan-Lafosse T, Blum M, Chuguransky S, Grego T, Pinto BL, Salazar GA, et al. InterPro in 2022. Nucleic Acids Res. 2023;51:D418–27.

61. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. Nat Biotechnol. 2017;35:316–9.

62. Ewels PA, Peltzer A, Fillinger S, Patel H, Alneberg J, Wilm A, et al. The nf-core framework for community-curated bioinformatics pipelines. Nat Biotechnol. 2020;38:276–8.

63. Shiryev SA, Papadopoulos JS, Schäffer AA, Agarwala R. Improved BLAST searches using longer words for protein seeding. Bioinformatics. 2007;23:2949–51.

64. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009;10:421.

65. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol. 2013;30:772–80.

66. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021;49:W293–6.

67. Magnadóttir B. Innate immunity of fish (overview). Fish Shellfish Immunol. 2006;20:137–51.

68. Smith NC, Rise ML, Christian SL. A comparison of the innate and adaptive immune systems in cartilaginous fish, ray-finned fish, and lobe-finned fish. Front Immunol. 2019;10:2292.

69. Stene A, Hellebø A, Viljugrein H, Solevåg SE, Devold M, Aspehaug V. Liquid fat, a potential abiotic vector for horizontal transmission of salmonid alphavirus? J Fish Dis. 2016;39:531–7.

70. M Monte M, Urquhart K, Secombes CJ, Collet B. Individual monitoring of immune responses in rainbow trout after cohabitation and intraperitoneal injection challenge with Yersinia ruckeri. Fish Shellfish Immunol. 2016;55:469–78.

71. Aschard H, Vilhjálmsson BJ, Joshi AD, Price AL, Kraft P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. Am J Hum Genet. 2015;96:329–39.

72. Zhang YB, Gui JF. Identification of two novel interferon-stimulated genes from cultured CAB cells induced by UV-inactivated grass carp hemorrhage virus. Dis Aquat Organ. 2004;60:1–9.

73. Krasnov A, Timmerhaus G, Schiøtz BL, Torgersen J, Afanasyev S, Iliev D, et al. Genomic survey of early responses to viruses in Atlantic salmon, Salmo salar L. Mol Immunol. 2011;49:163–74.

74. Timmerhaus G, Krasnov A, Takle H, Afanasyev S, Nilsen P, Rode M, et al. Comparison of Atlantic salmon individuals with different outcomes of cardiomyopathy syndrome (CMS). BMC Genomics. 2012;13:205.

75. Krasnov A, Johansen L-H, Karlsen C, Sveen L, Ytteborg E, Timmerhaus G, et al. Transcriptome responses of Atlantic salmon (Salmo salar L.) to viral and bacterial pathogens, inflammation, and stress. Front Immunol. 2021;12:705601.

76. Sun C, Liu Y, Hu Y, Fan Q, Li W, Yu X, et al. Gig1 and Gig2 homologs (CiGig1 and CiGig2) from grass carp (Ctenopharyngodon idella) display good antiviral activities in an IFN-independent pathway. Dev Comp Immunol. 2013;41:477–83.

77. Sun F, Zhang Y-B, Jiang J, Wang B, Chen C, Zhang J, et al. Gig1, a novel antiviral effector involved in fish interferon response. Virology. 2014;448:322–32.

78. Ortega-Villaizan MDM, Chico V, Perez L. Fish innate immune response to viral infection-an overview of five major antiviral genes. Viruses. 2022;14:1546.

79. Olejnik J, Hume AJ, Mühlberger E. Toll-like receptor 4 in acute viral infection: Too much of a good thing. PLoS Pathog. 2018;14:e1007390.

80. Robledo D, Taggart JB, Ireland JH, McAndrew BJ, Starkey WG, Haley CS, et al. Gene expression comparison of resistant and susceptible Atlantic salmon fry challenged with Infectious Pancreatic Necrosis virus reveals a marked contrast in immune response. BMC Genomics. 2016;17:279.

81. Reyes-López FE, Romeo JS, Vallejos-Vidal E, Reyes-Cerpa S, Sandino AM, Tort L, et al. Differential immune gene expression profiles in susceptible and resistant full-sibling families of Atlantic salmon (Salmo salar) challenged with infectious pancreatic necrosis virus (IPNV). Dev Comp Immunol. 2015;53:210–21.

82. Loker R, Mann RS. Divergent expression of paralogous genes by modification of shared enhancer activity through a promoter-proximal silencer. Curr Biol. 2022;32:3545-3555.e4.

83. Robertson FM, Gundappa MK, Grammes F, Hvidsten TR, Redmond AK, Lien S, et al. Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. Genome Biol. 2017;18:111.

84. Cheng M, Ren J, Shen F, Huang Y, Fan Z, Price M, et al. Genome-wide investigation of microsatellite polymorphism in coding region of the giant panda (Ailuropoda melanoleuca) genome: a resource for study of phenotype diversity and abnormal traits. Mammal Research. 2019;64:353–63.

85. Wu Z, Gong H, Zhou Z, Jiang T, Lin Z, Li J, et al. Mapping short tandem repeats for liver gene expression traits helps prioritize potential causal variants for complex traits in pigs. J Anim Sci Biotechnol. 2022;13:8.

86. Viļuma A, Mikko S, Hahn D, Skow L, Andersson G, Bergström TF. Genomic structure of the horse major histocompatibility complex class II region resolved using PacBio long-read sequencing technology. Sci Rep. 2017;7:45518.

87. Gymrek M, Willems T, Guilmatre A, Zeng H, Markus B, Georgiev S, et al. Abundant contribution of short tandem repeats to gene expression variation in humans. Nat Genet. 2016;48:22–9.

Paper 2

Evolution of *gig* immune genes in teleosts shaped by gene gain and loss: regulatory and functional insights from Atlantic salmon

Manousi, D.[1], Naseer, S.[2], Martin, S.A.M.[2], Sandve,S.R.[1] and Saitou,M.[1]

[1]Centre for Integrative Genetics, Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Oluf Thesens vei 6, 1433 Ås, Norway

[2]Scottish Fish Immunology Research Centre, School of Biological Sciences, University of Aberdeen, Aberdeen AB24 2TZ, United Kingdom

Abstract

Interferon-stimulated genes (ISGs) are key players in vertebrate antiviral immunity, yet their evolution and functional divergence remain unclear. Among teleost ISGs, *gig1* and *gig2* contribute to antiviral defense, but their regulatory mechanisms and evolutionary origins are not fully understood as they are absent in mammalian genomes. This study investigated *gig* gene evolution across teleosts, with a focus on Atlantic salmon (*Salmo salar*). Phylogenetic analysis revealed that *gig1* is teleost-specific, while *gig2* extends to lower vertebrates. Whole-genome duplication drove lineage-specific expansions, particularly of *gig2* in salmonids. Structural and transcriptomic analyses showed that *gig1* and *gig2* differ in domain composition, repeat content, and regulation. Our findings suggest the complex interplay of duplication history, structural divergence, and transcriptional regulation in shaping immune gene repertoires in teleosts, with implications for understanding host-pathogen interactions and aquaculture disease responses.

Key Words: immune genes, interferon-stimulated genes, whole-genome duplication, antiviral response, gene regulation, aquaculture

Introduction

Vertebrates have acquired complex immune systems to defend against a variety of pathogens (Iwama and Moran 2023). The interferon (IFN) signaling pathway plays a key role in vertebrate antiviral response as a molecular frontline in host defense (Dalskov *et al.* 2023). Upon viral infection, IFN activates a cascade of interferon-stimulated genes (ISGs); genes encoding proteins that suppress viral replication and enhance immune signaling and inflammatory responses (Walker *et al.* 2021). While insights into the functional diversification of ISGs have expanded, much remains to be explored outside the mammalian lineage (Purushotham *et al.* 2025). Teleost fish, in particular, present an interesting group to study immune gene diversity and evolution, owing to their aquatic habitats, exposure to diverse pathogens, and history of multiple whole-genome duplications (Dornburg *et al.* 2021).

Teleost fish occupy environments ranging from freshwater rivers to deep-sea habitats, representing the most taxonomically and ecologically diverse group of vertebrates (Robertsen 2006; Sun *et al.* 2009; Mayer and Pšenička 2024). Their aquatic lifestyle imposes unique immunological challenges as they are continuously exposed to high pathogen loads via permeable surfaces such as gills, skin, and gut epithelium (Gomez *et al.* 2013; Wang *et al.* 2025). Teleosts possess a distinct immune system architecture, including specialized mucosal lymphoid tissues and the use of the pronephros for hematopoiesis, that differs markedly from that of terrestrial vertebrates. (Salinas *et al.* 2011; Levraud *et al.* 2019; Stosik *et al.* 2023). These ecological and physiological features are thought to have driven the diversification of ISG repertoires in teleosts. Notably, teleosts underwent an additional, teleost-specific whole-genome duplication (3R) (Qi *et al.* 2024), with further duplications occurring in lineages such as salmonids and cyprinids (4R), potentially contributing to the expansion and functional specialization of immune gene families (Glasauer and Neuhauss 2014; Parey *et al.* 2022). These duplications provide opportunities for functional diversification of immune gene families, allowing for adaptation to diverse pathogen pressures (Rivera and Swanson 2022; Tasnim *et al.* 2024).

Among the many interferon-stimulated gene (ISG) families in vertebrates, the *grass carp reovirus-induced gene* (gig) family stands out for multiple reasons. First, it is absent in mammals but present in fish and lower vertebrates, representing a lineage-restricted ISG family that is potentially shaped by aquatic pathogen pressures (Zhang 2003; Zhang

and Gui 2004); (Zhang *et al.* 2013); (An *et al.* 2025). Second, gig genes show clear transcriptional responses to viral infection, across multiple species, including zebrafish, grass carp, and salmonids (Krasnov *et al.* 2011; Sun *et al.* 2014; An *et al.* 2025; Manousi *et al.* 2025). Consistently, in grass carp, overexpression of *gig1* and *gig2* enhanced resistance to grass carp reovirus (GCRV) (Sun *et al.* 2013). While some differences in the molecular responses of gig genes have been reported (Rao and Su 2015), their functional mechanisms and evolutionary relationships, especially between gig1 and gig2 genes, remain largely unknown. Third, gig genes are expressed in response to economically important diseases in aquaculture species, such as cardiomyopathy syndrome (CMS) and pancreas disease (PD) in salmonids (Timmerhaus *et al.* 2012; Robinson *et al.* 2020; Krasnov *et al.* 2021). As a recent example, three tandemly duplicated gig1 genes have emerged as candidate genes in a genome-wide association study for pancreas disease resistance in Atlantic salmon (Manousi *et al.* 2025). Collectively, these features suggest that gig genes are of particular interest for their potential roles in aquaculture and disease resistance in aquatic vertebrates (Krasnov *et al.* 2011).

In this study, we present a comprehensive investigation of *gig* gene evolution and functional diversification in teleosts, with a particular focus on Atlantic salmon. Atlantic salmon (*Salmo salar*) provides an interesting system to investigate the evolution and regulation of the *gig* gene families; a lineage-specific whole-genome duplication (Ss4R) occurred in the ancestor of all salmonids, leading to possible functional divergence and specialization. Atlantic salmon is also one of the fish species with the most developed genomic resources, including publicly available transcriptomic and epigenetics datasets (https://www.aqua-faang.eu/) (Johnston *et al.* 2024). We therefore integrate phylogenetic analysis and functional genomics data from in-vivo and in-vitro viral challenges to elucidate how *gig1* and *gig2* have evolved and contributed to immune gene innovation in non-mammalian vertebrate lineages.

Materials and Methods

Species selection and gene family phylogenetic tree construction

To investigate the evolutionary history and diversification of *gig* genes across teleosts, we selected 11 representative species and constructed a gene family phylogenetic tree.

All genome assemblies and gene annotations were retrieved from the Ensembl database (version 112), ensuring consistency in gene models and facilitating reliable cross-species comparisons of *gig* homologs. Species were chosen to represent major teleost clades and to include lineages that have undergone whole-genome duplication (WGD) events, which are relevant to the expansion and diversification of immune-related gene families. The selected taxa, Atlantic salmon, Northern pike, zebrafish, common carp, Coho salmon, Chinook salmon, rainbow trout, brown trout, Japanese medaka, three-spined stickleback, and spotted gar (**Table 1**), span a range of phylogenetic positions (e.g., Otophysi, Euteleostei, Holostei) and encompass both WGD and non-WGD lineages.

To identify *gig* homologs, coding DNA sequences of *gig1* and *gig2* from crucian carp were used as queries against the selected genomes. The corresponding gene names and descriptions were retrieved via BioMart (Ensembl v112) and NCBI gene annotations when applicable. For each gene, the longest protein-coding transcript was retained, and its amino acid sequence was used for downstream analysis.

Amino acid sequences were aligned using GUIDANCE2 (Sela *et al.* 2015), in order to assess the reliability of sequence alignments and filter out poorly aligned regions. Alignments were refined by excluding positions with consistency scores below 0.93 (site alignments present in > 93% of GUIDANCE replicates), ensuring robust phylogenetic inference. Reliability filtering of sequence alignments removed one Atlantic salmon *gig2* ortholog, ENSSSAG00000082930, which could be a pseudogene or partial duplication. The remainder alignments were used to construct maximum likelihood phylogenies in IQ-TREE version 2.3.3, implementing the *Model-finder* option to find and use the best-fit substitution model for each *gig* phylogenetic tree (Nguyen *et al.* 2015; Kalyaanamoorthy *et al.* 2017). Support values for each tree were estimated using 1,000 replicates and the ultrafast bootstrap method implemented in IQ-TREE (Hoang *et al.* 2018). The resulting trees, together with a species tree constructed in timetree.org, were visualized in the iTOL platform (Letunic and Bork 2021). Phylogenetic clades in the constructed gene trees were then defined based on the visualized phylogenies.

Gig functional domain prediction

To characterize conserved protein domains and infer functional divergence between *gig1* and *gig2* genes, the predicted amino acid sequences of the longest isoform per

gene were analyzed using InterProScan v98.0 (Paysan-Lafosse *et al.* 2023). This tool integrates domain predictions from multiple databases, including Pfam, SMART, and Gene3D. Domains found in fewer than 5% of all *gig* genes within a given clade, typically appearing only 2–3 times, were excluded to reduce the influence of rare features. By doing so, we aimed to improve signal specificity in detecting conserved functional domains across teleost lineages.

Domains identified in *gig1* and *gig2* genes were then grouped based on functional similarity to assess potential divergence in molecular function (**Supplementary Table 1**). To visualize functional domains in Atlantic salmon gig genes, similar functional features that were identified across multiple databases were collapsed together.

Detection of repeat content within coding regions of *gig* genes

To identify internal sequence repetitions within *gig* paralogs, the nucleotide sequences were analyzed using a self-comparison approach. Pairwise comparisons of the coding sequence for each gig gene were performed using the 'dotPlot' function in the R package *seqinr* (Charif and Lobry 2007; R Core Team 2021) to visualize self-aligned sequence and detect repetitive motifs. A sliding window approach was implemented to detect local similarities within each sequence, using a window size of 10 nucleotides and a step size of 1 nucleotide. A dot was plotted at each position where 9 or more nucleotides matched between two overlapping windows. This allowed for the visualization of internal repeats and low-complexity regions that may indicate recent duplications or structural features relevant to *gig* gene evolution.

Evolutionary characterization of Atlantic salmon *gig* genes

To investigate the genomic context and evolutionary origins of *gig* genes in Atlantic salmon, we examined their chromosomal positions in relation to known patterns of rediploidization following the salmonid-specific whole genome duplication event (Ss4R). Genomic coordinates for *gig1* and *gig2* were retrieved from Salmobase (https://salmobase.org/) and compared against chromosomal regions previously classified as undergoing either Lineage-specific Ohnolog Resolution (LORe) or Ancestral Ohnolog Resolution (AORe) (Lien *et al.* 2016; Gundappa *et al.* 2022), reflecting two phases of post-duplication divergence.

In addition to genome-wide duplication, we also assessed the potential contribution of tandem duplication events to *gig* gene family expansion. Tandemly duplicated genes

were defined as those located on the same chromosome and separated by less than 1 Mbp. This classification allowed us to distinguish between large-scale (WGD-derived) and local duplication mechanisms in shaping the current genomic architecture of the *gig* family in Atlantic salmon.

Manual annotation of *gig* paralogs

Many *gig* genes across species showed substantial sequence divergence. This was reflected in the large number of amino acid residues removed during alignment filtering. We hypothesized that this pattern reflects genuine biological variation, especially a signature of rapid evolution in immune-related genes, rather than a result of annotation errors. To test this hypothesis, we examined the transcript structures of *gig* genes using publicly available RNA-seq datasets. While this phenomenon was not limited to Atlantic salmon, we focused on this species because of the rich availability of both short- and long-read transcriptomic data.

Short-read RNA-seq data from poly(I:C) stimulation and SAV3 infection challenges were obtained from the AQUAFAANG public repository (accession number PRJEB50076,(Clark *et al.* 2023)) and the European Nucleotide Archive (accession number PRJEB85594, (Manousi *et al.* 2025)), respectively. RNA-seq reads were aligned to the Atlantic salmon genome (Ssal_v3.1, Ensembl v112) using STAR 2.7.10a (Dobin *et al.* 2013) and the alignment coverage of samples with the highest *gig* gene expression was visualized using the Integrative Genome Viewer (IGV) (Robinson and Zemo jtel 2017).

In addition to short-read RNA-seq data, long-read transcriptomic data from 12 healthy individuals, previously sequenced using Oxford Nanopore Technologies (ONT) were additionally incorporated to refine gene annotations and assess the limitations of short-read mapping in complex genomic regions. Details regarding sampling and alignment procedures for the ONT-sequenced individuals against the Ssal_v3.1 genome are available elsewhere (Sandholm 2022). Aligned transcriptomic data from both platforms were compared to the existing gene annotation (Ssal_v3.1) in IGV, enabling refinement and validation of exon-intron boundaries, identification of transcription start sites (TSS), and confirmation of alternative isoforms.

To further evaluate coding potential, sequences corresponding to candidate exons (including 100 bp upstream and downstream flanking regions) were extracted and analyzed using the NCBI ORFfinder. Where necessary, predicted protein products were

submitted to InterProScan (Paysan-Lafosse *et al.* 2023) to confirm the presence or absence of conserved functional domains.

Discovery of interferon-specific regulatory regions using chromatin accessibility and histone modification data

Interferon-stimulated response elements (ISREs, (Collet and Secombes 2001)) were examined within 2 kb upstream and 200 bp downstream of the transcription start site of Atlantic salmon *gig* genes to assess their regulatory potential in interferon-mediated immune responses (Holen *et al.* 2023).

To evaluate the functional relevance of these predicted regulatory elements, publicly available chromatin accessibility (ATAC-seq) and histone modification (H3K27ac, H3K27me) data were incorporated. These datasets were derived from an interferon-stimulation experiment in Atlantic salmon head kidney tissue, where poly(I:C)-treated samples were compared to untreated controls (Clark *et al.* 2023). In that dataset, DNA was extracted from head kidney tissue of six individuals per condition, and sequencing was performed using ATAC-seq and ChIP-seq technologies to assess differential chromatin accessibility and histone modifications. For this study, the nf-core pipelines *atacseq* and *chipseq* were used, providing standardized, modular workflows for quality control, read alignment, and peak calling (Ewels et al. 2020).

Further refining the regulatory element identification, an Irreproducible Discovery Rate (IDR) analysis was performed using the software ChIP-R (Newell *et al.* 2020) in order to ensure the reproducibility of regulatory signals across samples in each condition and reduce false positives. Finally, the differential accessibility of peaks between conditions was assessed using DESeq2 (Love *et al.* 2014) and regulatory elements were considered statistically significant if they exhibited an FDR-adjusted p-value < 0.05. These analyses enabled the identification of ISRE motifs and chromatin modifications that potentially regulate *gig* gene expression in response to interferon stimulation.

Transcriptomics data and gene expression analysis of Atlantic salmon *gig* genes

To explore the immune-related functional divergence of *gig1* and *gig2*, we analyzed transcriptomic responses to both interferon stimulation and viral infection using three publicly available RNA-seq datasets. Two datasets were derived from interferon-stimulation experiments (accession number PRJEB50076, (Clark *et al.* 2023)). In these

studies, Atlantic salmon head kidney tissue and primary head kidney cells were treated with poly(I:C), while control groups received PBS medium. The third dataset originated from a viral challenge experiment involving Salmonid alphavirus (SAV3), in which heart tissue was sampled from fish that survived or died following 11 days of infection (Hillestad *et al.* 2020). For all datasets, total RNA was extracted and sequenced using Illumina-based RNA-seq platforms. Quality filtering, alignment of sequencing reads to the salmonid transcriptome (Ssal_v3.1, GCA_905237065.2) and quantification of gene expression were performed using the nf-core *rnaseq* pipeline *(Patel et al. 2024)*.
Differential gene expression analysis was performed independently for each dataset using DESeq2 (Love *et al.* 2014), with a threshold of FDR-adjusted p-value < 0.05 to identify regulated genes. This integrative approach enabled a comparative assessment of *gig* gene expression under both direct interferon stimulation and virus-induced immune activation, providing insight into their condition-specific regulatory dynamics.

A Large Language Model usage

We have used a Large Language Model (LLM) in our work, specifically ChatGPT (GPT-4) developed by OpenAI. The model was accessed through the OpenAI API, and our interactions primarily involved two types of queries: seeking assistance with coding-related issues and refining English expressions as non-native speakers. Our queries were structured in natural language, with iterative refinements applied when necessary to improve clarity and accuracy. As authors, we acknowledge our responsibility for the accuracy of the content generated, ensuring that it is free from plagiarism, and properly attributing all sources, including material produced by the LLM.

Results

Expansion and divergence of *gig1* and *gig2* gene families across teleosts species

Comparative genomic analysis across ancestral and teleost species revealed distinct evolutionary trajectories for *gig1* and *gig2* gene families. While *gig2* genes were identified in all surveyed non-amniote vertebrates, in agreement with previous reports (Zhang et al. 2013), *gig1* genes were exclusively identified in teleosts (**Figure 1**). Notably, we did not identify any *gig1* homologs in the spotted gar genome using our criteria. However, a recent study (An *et al.* 2025) reported a candidate *gig1* gene in spotted gar

based on conserved domain architecture. Upon our inspection, the gene in question (ENSLOCG00000006921) did show relative similarities in domain composition, but in Ensembl, it clusters separately in its own orthogroup (ENSGT00940000174456). Given this ambiguity, and to ensure consistency in phylogenetic comparison, we excluded this gene from the *gig1* set used in our study. Lastly, comparisons between species that experienced lineage-specific WGD events (e.g., common carp, salmonids) and their closest non-duplicated relatives (e.g., zebrafish, northern pike) revealed that whole-genome duplication (WGD) played a key role in the expansion of the two gig families, but with clear lineage-specific differences; *gig1* genes underwent expansion in common carp (*Cyprinus carpio*) (Xu et al. 2019), whereas *gig2* genes expanded in salmonids.

To further investigate the evolutionary history of *gig* genes across the selected species, we constructed phylogenetic trees using the protein translation of the longest coding transcripts (**Figure 2 and Supplementary Figures 1 and 2**). Multiple sequence alignments were refined using GUIDANCE2 (Sela et al. 2015) to eliminate potential annotation errors and excessively divergent residues (Materials and Methods). Sequence reliability filtering revealed that 31% of *gig1* alignments were retained after filtering low-conservation sites, compared to 49% for *gig2*, indicating that *gig1* genes have evolved at a faster rate. Additionally, one Atlantic salmon *gig2* gene (ENSSSAG00000082930) was removed during the filtering process due to extreme sequence divergence, possibly linked to pseudogenization. The phylogenetic analysis indicated two major clades of *gig1* genes (*gig1A* and *gig1B*) and three (*gig2A*, *gig2B*, and *gig2C*) clades in *gig2* genes (**Figures 2a and 2b and Supplementary Figures 1 and 2**). Further investigation of the phylogenetic structure also supported that the *gig1B* clade underwent expansion in common carp, likely driven by both WGD and tandem duplications; the common carp genes clustered in *gig1B* clade are located on two scaffolds in the carp genome assembly (scaffolds CAJNDQ010000023.1 and CAJNDQ010000028.1, **Supplementary Figure 1**).

In contrast, *gig2* genes exhibited more complex, lineage-specific divergence. While *gig2A* and *gig2B* expanded in common carp, *gig2C* showed a pronounced increase among salmonids. Within salmonids, *gig2* continued to diversify, with brown trout (*Salmo trutta*) showing expansion of ancestral-like *gig2A* genes, whereas rainbow trout (*Oncorhynchus mykiss*) exhibited an increase in *gig2C* **(Supplementary Figure 2)**. Notably, brown trout exhibited an expansion of ancestral-like *gig2A* genes, whereas

rainbow trout showed an increase in *gig2C* copies, a pattern that may reflect species-specific adaptations to environmental or pathogenic pressures. However, the functional implications of these evolutionary trajectories, including how regulatory and structural divergence has shaped the immunological roles of *gig* gene paralogs, remain to be clarified. In the following sections, we focus on the Atlantic salmon and explore genomic context, duplication history, and transcriptional regulation to uncover how gene duplication has contributed to the diversification of *gig* gene function across teleost species.

Genome duplication and tandem repeats shaped the *gig* gene emergence and structural changes in Atlantic salmon

Atlantic salmon harbors the highest number of *gig* genes among the surveyed species (**Figure 1**), a pattern likely shaped by the salmonid-specific whole-genome duplication (Ss4R) that occurred approximately 125 million years ago (Gundappa *et al.* 2022). To investigate the evolutionary origin and structural divergence of these genes, we examined chromosomal organization, duplication patterns, and predicted protein features. In particular, we obtained synteny information from the publicly available database *Salmobase* (https://salmobase.org/genomes/AtlanticSalmon/Ssal_v3.1) and assessed the physical location of gig paralogs considering the impact of the Ss4R genome duplication in the Atlantic salmon genome. Based on the rediploidization history following the salmonid-specific whole genome duplication event (Gundappa *et al.* 2022), we distinguished gig paralogs that originated from early (AORe) and recent (LORe) rediploidization events. Rediploidization timing has a large impact on sequence similarity, with earlier occurrences allowing for higher sequence divergence. We therefore used the AORe and LORe regions and aimed to determine whether rediploidization timing influenced functional specialization (Krasnov *et al.* 2021).

Synteny analysis using data from Salmobase (Ssal_v3.1) revealed that three *gig1* and one *gig2* genes are located in AORe regions, while the remaining 13 *gig1* and 15 *gig2* genes reside in LORe regions (**Figure 3**). Two *gig2* genes were located on unplaced scaffolds (CAJNNT020001543.1 and CAJNNT020001561.1) and were therefore excluded from further analyses. Our results suggest that the majority of the 32 Atlantic salmon *gig* genes expanded following salmonid lineage diversification and distributed across 10 chromosomes. A high proportion of them was tandemly duplicated, particularly within LORe regions on chromosomes ssa03, ssa04 and ssa08 (**Figure 3**), but no tandem

duplicates were identified within AORe regions. The different *gig* gene distribution pattern between AORe and LORe regions raises the possibility that distinct evolutionary forces may be acting in these ancestral duplicate regions.

Protein domain prediction using InterPro (version 102.0, (Paysan-Lafosse *et al.* 2023)) identified 17 conserved domains across all Atlantic salmon *gig* genes, which were grouped into eight clusters based on functional similarity (**Figure 4 and Supplementary Table 1**). Notably, *gig1* genes contained exclusively *signal peptide*-related as well as '*SI:CH211-198C19.1-RELATED'* domains, with the latter domain being detected across all *gig1* genes, suggesting a high degree of structural conservation within this family. Similarly, *gig2* genes obtained exclusively the *Gig2-(like)* protein, ADP-ribosylation-related but also *coil* domains, with the latter ones overlapping '*consensus disorder*' predicted domains (**Figures 4a and 4b**).

This family-specific domain composition supports the functional divergence between *gig1* and *gig2* lineages. Furthermore, there was variation in the length of the *'SI:CH211-198C19.1-RELATED'* domain between *gig1A* genes and *gig1B* genes (**Figure 4a**), pointing to potential functional divergence within the *gig1* clade. Besides gig family-specific domains, membrane-related and *consensus disorder* domains were found across some genes of both families (**Supplementary Table 1**). Among *gig1* genes, six carried membrane-related domains and four carried signal peptide domains; these were mostly located on ssa06, ssa04, ssa14, and ssa22. Meanwhile, only two *gig1B* genes, *gig1B_ssa04f* and *gig1B_ssa04g*, which are tandem duplicates, retained such domains (**Figure 4a**).

Given the role of repeat content in functional diversification (Rivera and Swanson 2022; Pajic *et al.* 2022), we further examined the presence of repeat structures within *gig* genes (**Figure 5 and Supplementary Figures 3-6**). In accordance with earlier reports (Zhang *et al.* 2013), comparison of the self-alignments of *gig1* and *gig2* genes identified several different tandem repeats and fusions. Specifically, repeat content was detected in genes located within large tandem duplication clusters; *gig1* genes located on chromosome ssa04, and *gig2* genes located across chromosomes ssa02, ssa08, and ssa12. Interestingly, *gig1* genes with elongated domains lacked repeats, whereas *gig2* genes containing heavily repeated sequence exhibited elongated and duplicated ADP-ribosylation domains (**Figure 5, Supplementary Figures 1 and 2**). In addition, the genes with high repeat content *gig2B_ssa02d* and *gig2B_ssa12a*, which were WGD-derived

duplicates, overlapped predicted *consensus disorder* and *coil* domains (**Figures 4b and 5**), however there were not substantial differences in the length of the respective *Gig2 protein*-related domain. Notably, the particular genes contained an exceptionally high abundance of glutamate (E) repeats in their amino acid sequences (**Figure 5**). However, besides this extensive repeat expansion, the overall length of the core Gig2-related domain remained relatively stable across *gig2* paralogs. This suggests that while the central functional domain may be conserved, structural variation has occurred primarily in the N-terminal end of the protein. Although the functional consequences of these changes remain to be determined, they raise the possibility of differential regulatory or immunological roles among paralogs. Taken together, these findings indicate that *gig* gene expansion in Atlantic salmon has been shaped not only by large-scale duplication events but also by within-gene tandem repeat expansions and structural rearrangements, such as internal repeat proliferation. To assess whether such gene-specific structural features correspond to functional divergence, we next examined the transcriptional responses of *gig* paralogs to immune stimulation.

Expression divergence of *gig* paralogs reveals functional diversification under immune stimulation

To explore the functional roles and regulatory divergence of *gig* genes in Atlantic salmon, we analysed their expression responses to immune stimulation using three independent RNA-seq datasets. These included in vitro and in vivo poly(I:C) stimulation of head kidney (HK) cells and tissues, as well as heart tissue transcriptomes from a salmonid alphavirus (SAV3) infection challenge (Hillestad *et al.* 2020). Our goal was to distinguish between interferon-induced gene regulation and more complex host responses to viral infection.

Gig1 genes displayed heterogeneous transcriptional responses, with expression patterns differing between the gig1A and gig1B clades (**Figure 6**). Several *gig1B* genes, particularly those located on chromosomes ssa03 and ssa06, consistently responded to both *in vitro* and *in vivo* immune stimulation, including SAV3 infection. In contrast, only one out of four *gig1A* genes, *gig1A_ssa14*, showed notable upregulation. This gene also exhibited relatively high baseline expression, suggesting it may play a more prominent immune role within the *gig1A* clade. Other *gig1A* members showed minimal or no transcriptional induction under any condition tested. On ssa04, multiple gig1B paralogs were detected; however, only *gig1B_ssa04b* and *gig1B_ssa04g* responded to

stimulation, with the latter showing a more pronounced response in head kidney tissue. Overall, the expression profiles of *gig1* genes did not correlate strongly with phylogenetic clade structure, indicating that post-duplication regulatory divergence has played a key role in shaping their transcriptional pattern.

In contrast, gig2 genes exhibited relatively more conserved expression patterns, particularly among WGD-derived paralogs. Members of the gig2A clade, most of which are located in LORe regions of ssa04, were generally upregulated in response to viral infection. An exception was *gig2A_ssa08a*, which showed low baseline expression and minimal induction. Similarly, gig2B paralogs (*gig2B_ssa12a* and *gig2B_ssa02d*) showed interferon-induced upregulation in head kidney cells, suggesting coordinated regulation in specific immune contexts.

Expression responses among gig2C paralogs were more variable. While *gig2C_ssa02a* and *gig2C_ssa02b* exhibited little expression under any condition, *gig2C_ssa12c* and *gig2C_ssa24* showed detectable responses to infection, indicating that regulatory divergence may be emerging within this clade. However, given the limited number of gig2C genes and their overall low expression levels, further data would be required to draw definitive conclusions about functional divergence in this group. Taken together, these results suggest that gene sequence similarity does not always predict transcriptional regulation. While some closely related paralogs maintain similar expression patterns, others exhibit greater regulatory plasticity, potentially reflecting functional specialization or even differential adaptation to pathogenic stimuli. These findings underscore the complex interplay between duplication history and transcriptional regulation in shaping immune gene function.

Cis-regulatory motif analysis highlights context-dependent regulation of *gig* genes

To further elucidate the transcriptional regulation of *gig* genes in Atlantic salmon, we performed a scan for transcription factor (TF) binding motifs. Given the established classification of *gig* genes as Interferon (IFN)-stimulated genes (Zhang and Gui 2004; Sun *et al.* 2014), we searched for interferon-stimulated response elements (ISREs) - canonical sequences that bind IFN-regulated TFs during antiviral responses- in close proximity to *gig* genes. Using ISRE motifs with known impact in salmonids (Collet and Secombes 2001), and a physical proximity threshold of 2000 bp upstream and 200 bp

downstream the TSS of each gene, we identified 25 ISRE motifs in 13 *gig1* genes and 24 ISRE motifs in 13 *gig2* genes (**Table S3**).

To assess the regulatory relevance of these motifs, we then identified genomic regions that showed chromatin remodeling in response to poly(I:C)-induced interferon activation (Materials and Methods). We then detected ISREs that overlapped with differentially regulated chromatin accessibility peaks upon poly(I:C)-induced interferon activation. This allowed us to group gig genes into three categories: (i) genes obtaining nearby ISRE motifs associated with virally induced chromatin accessibility changes, (ii) genes obtaining nearby ISRE motifs without association to chromatin accessibility changes, and (iii) genes lacking nearby ISRE motifs (**Figure 6**). Comparison of expression profiles revealed a general trend: genes in class (i) tended to show higher expression under immune stimulation (**Figure 7**) however, this difference was only significant for gig1 genes (Mann–Whitney U test, gig1: $p = 0.012$) gig2: $p = 0.37$).

Interestingly, exceptions to this pattern were observed in tandemly duplicated genes. For example, *gig1B_ssa03a* harbored a transcriptionally active ISRE and showed upregulation, while its tandem duplicates, *gig1B_ssa03b* and *gig1B_ssa03c*, which also were upregulated lacked predicted ISREs (**Figure 7**). These findings suggest that cis-regulatory elements may exert effects across multiple adjacent paralogs or that alternative, uncharacterized motifs contribute to gig gene expression. Such context-specific transcriptional activity underscores the complexity of regulatory evolution in duplicated genomic regions.

Together, these results demonstrate that the transcriptional regulation of gig genes reflects both conserved IFN-driven control and divergent, context-dependent mechanisms.

Discussion

Evolutionary origin and diversification of gig genes

The *gig* gene families constitute a group of IFN-stimulated genes (ISGs) that are absent in mammals but conserved and expanded in fish and other aquatic vertebrates (Zhang and Gui 2004; Krasnov *et al.* 2011); (Zhang *et al.* 2013; An *et al.* 2025). These genes exhibit transcriptional activation during antiviral responses, indicating their important

roles in vertebrate immune evolution, particularly within the aquatic environment where pathogens and environmental stressors differ substantially from those in terrestrial habitats (Sun *et al.* 2013; Gao *et al.* 2017; Krasnov *et al.* 2021; Madushani *et al.* 2021). In Atlantic salmon, a species of additional importance for the aquaculture industry, gig genes have shown involvement in antiviral responses (Krasnov *et al.* 2011). Recent studies suggest that tandemly duplicated *gig1* genes, in particular, may play key roles in resistance to serious pathogens in farmed fish (Manousi *et al.* 2025). These findings underline the necessity to further explore the evolutionary and functional divergence of *gig* genes, not only to deepen our understanding of immune evolution in teleosts but also to inform disease management strategies and sustainability efforts within aquaculture.

Our comparative analyses revealed that while both *gig1* and *gig2* genes are upregulated in response to viral challenges, they have evolved along distinct evolutionary trajectories. Previous studies detected the earliest emergence of gig2 in Cyclostomata (*P. marinus*), whereas the earliest emergence of gig1 genes was reported in Holostei (*L. oculatus*) (Zhang *et al.* 2013; An *et al.* 2025). Phylogenetic analysis across teleost species with additional whole-genome duplications, including Atlantic salmon and common carp, revealed the complex interplay between gene duplication and divergence in shaping immune gene repertoires. Interestingly, the divergent evolutionary trajectories of *gig1* genes between salmonids and cyprinids may reflect lineage-specific responses to ecological niches and pathogen pressures. While salmonids face the challenges of anadromous life cycles and varied pathogen exposures across freshwater and marine environments (Vollset *et al.* 2021), common carp inhabit more stable freshwater habitats with distinct immune challenges (Banet *et al.* 2022). The observed expansion, retention, and divergence of *gig* genes therefore highlight the importance of host ecology and evolutionary history in shaping immune gene families across teleosts.

In addition, the phylogenetic and sequence conservation analyses in our study identified higher sequence divergence of *gig1* genes, as well as substantial length variation of the *gig1* protein sequence and likely protein-coding domain between phylogenetic clades. These results suggest possibly broader functional divergence of gig1 genes compared to the gig2 family. Still, both gene families exhibited extensive sequence divergence and high turnover rates, with frequent gene gains and losses shaping their evolutionary histories; These findings are consistent with earlier studies (Zhang *et al.* 2013; Sun *et al.* 2014; An *et al.* 2025) as well as the rapid birth-death evolutionary turnover rate

identified in other IFN-related genes (Redmond *et al.* 2019). The observed sequence diversity and difficulty in reconstruction of clear phylogenetic relationships within both *gig* families in the present study is in line with the idea that gig genes are evolving fast under strong selective pressures driven by host-pathogen interactions (Shultz and Sackton 2019).

In salmonids, lineage-specific expansions of *gig* genes, particularly in the *gig2* family, reflect the unique evolutionary history of this group and the complex immune challenges they face across freshwater and marine environments (Vollset *et al.* 2021). Our phylogenetic study showed a wide divergence of *gig2* genes and their subgroup expansions within salmonids, that may be partly shaped by ecological and pathogenic pressures imposed on these lineages.

It is, though, unlikely that such expansions have occurred over the short timescales of aquaculture practices alone; instead, these expansions represent long-term evolutionary processes, such as genetic drift acting on redundant copies and differential retention of gene duplicates after whole-genome duplications. In line with this hypothesis, restriction of transcriptional activity and absence of antiviral immunity-related cis-regulatory elements occurred in some *gig* paralogs, further supporting possible pseudogenization or specialization of certain gene copies. Nevertheless, understanding the resulting diversity of *gig* paralogs may have practical relevance for aquaculture, particularly in light of the high viral burdens faced by salmonid populations, including salmon anemia virus (ISAV) and salmonid alphavirus (SAV), which affect fish health in aquaculture (Krasnov *et al.* 2021). Further functional studies are therefore needed to clarify whether this evolutionary legacy has functional consequences that could be leveraged to develop disease management strategies for aquaculture.

Structural evolution of *gig1* and *gig2* genes in Atlantic Salmon

In our focused analysis on Atlantic salmon, functional domain analysis identified exclusive protein-coding and ADP-ribosylation domains within each family, consistent with previous studies (Zhang and Gui 2004; Sun *et al.* 2013). Interestingly, while earlier work suggested that *gig1* proteins primarily reside in the cytoplasm (Sun *et al.* 2014), our analysis revealed that a subset of Atlantic salmon *gig1* genes carry membrane-associated domains, indicating potential divergence in subcellular localization of gig paralogs.

However, only a few of these membrane-associated *gig1* paralogs exhibited significant transcriptional responses to viral infection, raising the possibility that structural modifications alone do not dictate functional activity. Additionally, repeat elements were frequently found across *gig1* and *gig2* paralogs. Large repeats in certain *gig1* genes were associated with apparent gaps in functional domain prediction, although manual annotation confirmed the integrity of the coding sequences. In *gig2*, repeat expansions coincided with duplicated functional domains, particularly in *gig2_ssa08b* and *gig2_ssa12b*, while shorter tandem repeats contributed to glutamate (E)-rich disorder regions in *gig2_ssa02d* and *gig2_ssa12a*. Repeat-driven genomic rearrangements are known to influence protein stability and plasticity, facilitating functional diversification (Rivera and Swanson 2022). For immunity in particular, the presence of repeat content in the amino acid sequence can increase plasticity, promoting functional diversification over short evolutionary timescales, which is an important factor in the arms race between hosts and pathogens (Teekas *et al.* 2022). In our dataset, repeat-containing *gig* genes often exhibited low expression and tissue-specific responses, suggesting that such structural modifications may contribute to either adaptive specialization to additional roles or pseudogenization of certain paralogs following gene family expansion.

Transcriptional Regulation and Functional Divergence of *gig* genes in Atlantic salmon

The presence of IFN-stimulated response elements (ISREs) in most *gig* genes indicates their associated role in immune regulation. However, only a subset of *gig1* and *gig2* paralogs exhibited regulatory activity during viral infection, concordant with the idea that transcriptional control, rather than simple promoter presence, modulates functional relevance (Ray-Jones and Spivakov 2021). Consistent with their classification as IFN-stimulated genes, *gig* genes showed strong responses to both poly(I:C)-induced (IFN-mediated) immune activation and SAV3 infection in heart and head kidney cells and tissues. While most paralogs exhibited similar expression across different infection conditions, a small subset showed distinct regulatory responses, suggesting functional divergence among retained copies. Interestingly, tandemly duplicated *gig* genes did not consistently share expression patterns, implying that regulatory elements may be exchanged between physically adjacent paralogs. This pattern aligns with previous findings in antiviral genes, where tandem duplication has led to transcriptional interference or regulatory divergence over evolutionary time (Krasnov *et al.* 2021). At

the same time, correspondence of gig expression between IFN-stimulated cells and tissue samples indicates the possibility of studying the functional divergence of gig genes using in vitro models, thus reducing the need for conduction of infectious challenges and consequent sacrifice of live fish.

Despite their established role as IFN-responsive genes, earlier studies in grass carp suggested that *gig* genes might also be activated through an IFN-independent pathway (Sun *et al.* 2013). Our findings in Atlantic salmon suggest that while *gig* genes largely function within an IFN-dependent framework, regulatory divergence may have led to paralog-specific tuning of immune responsiveness. Notably, some *gig* paralogs exhibited consistently low expression, raising the possibility of functional redundancy or pseudogenization following multiple rounds of whole-genome and segmental duplication, supporting the idea of a combination of selection and drift shaping the evolution of the *gig* gene family. Such patterns align with recent studies on antifreeze proteins (AFPIs) in fish, which were shown to have originated from non-coding sequences of a seemingly unrelated *gig2* gene (Graham *et al.* 2022; Rives *et al.* 2024). This provides a marked example of how functionally redundant paralogs can be retained as raw material for evolutionary innovation, reinforcing the idea that immune genes frequently undergo repurposing under selective pressures. Still, further analysis is required, potentially through application of targeted gene editing approaches in order to identify the regulatory mechanisms and functional impact of *gig* paralogs.

Conclusion

In conclusion, our study highlights the complex evolutionary dynamics of gig genes in teleost fish, particularly in Atlantic salmon. The divergence of gig1 and gig2 lineages, shaped by whole-genome duplications, tandem duplications, and repeat-driven structural variations, reflects the adaptive potentials of immune genes in aquatic environments. Despite shared discovery in response to viral challenges, these genes have diversified in sequence, regulatory motifs, and expression patterns, suggesting functional specialization. These findings not only deepen our understanding of immune gene evolution in fish but also provide a foundation for exploring how host-pathogen interactions and environmental pressures shape immune gene repertoires in vertebrates. Such insights may inform disease management strategies in aquaculture by leveraging the evolutionary plasticity of immune genes.

Tables

Table 1. Genome assemblies utilized in this study

Scientific name	Common name (Ensembl)	Accession	N50	Citation
Oncorhynchus tshawytscha	Chinook salmon	GCA_018296145.1	2884253	(Christensen *et al.* 2018)
Oncorhynchus mykiss	Rainbow trout	GCA_013265735.3	39165350	(Gao *et al.* 2021)
Gasterosteus aculeatus	Stickleback	GCA_016920845.1	20445003	(Nath *et al.* 2021)
Lepisosteus oculatus	Spotted gar	GCA_000242695.1	6928108	(Braasch *et al.* 2016)
Cyprinus carpio	Common carp german mirror	GCA_905221575.1	28473900	-
Danio rerio	Zebrafish	GCA_000002035.4	4737936	Genome reference consortium

Oncorhynchus kisutch	Coho salmon	GCA_002021735.2	68265700	-
Salmo salar	Atlantic salmon	GCA_905237065.2	28058890	-
Salmo trutta	Brown trout	GCA_901001165.1	52209666	(Hansen *et al.* 2021)
Esox lucius	Northern pike	GCA_011004845.1	37473588	-
Oryzias latipes	Japanese medaka HNI	GCA_002234675.1	31218526	(Ichikawa *et al.* 2017)

3

4

Figures

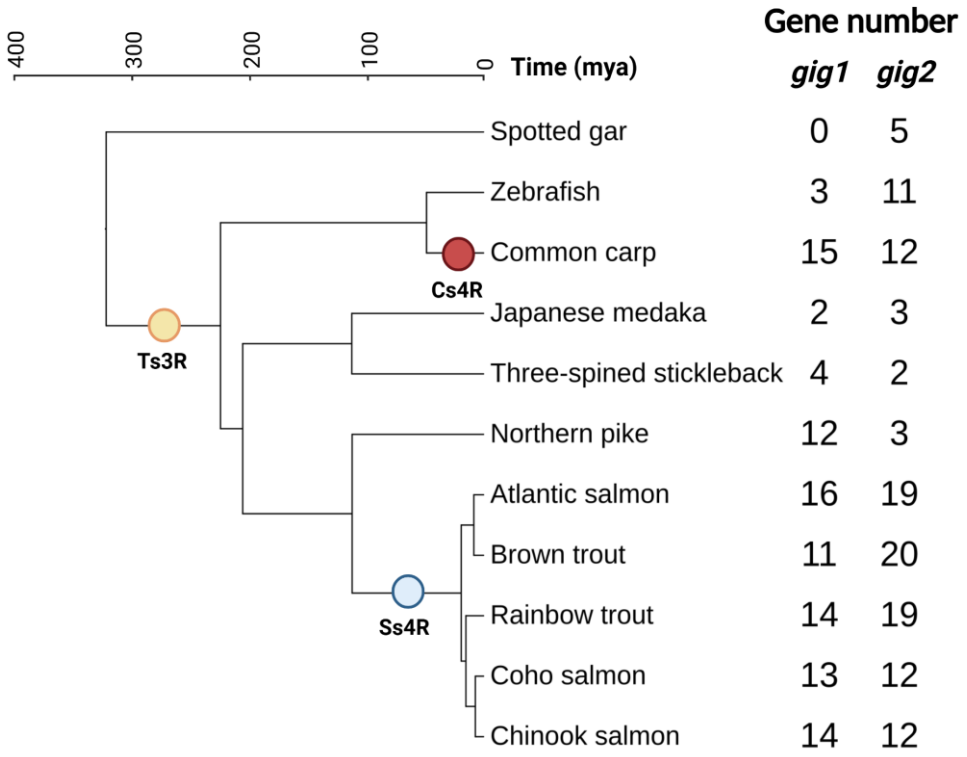


Figure 1. Gig1 and gig2 gene abundance across teleost species and the outgroup species spotted gar (Lepisosteus oculatus). A phylogenetic tree for 11 species was constructed in Timetree.org. Yellow, red and blue annotations represent the teleost-specific (Ts3R), carp-specific (Cs4R) and salmonid-specific (Ss4R) whole genome duplication events, respectively. Tree scale shows the chronological divergence of species in million years (mya)

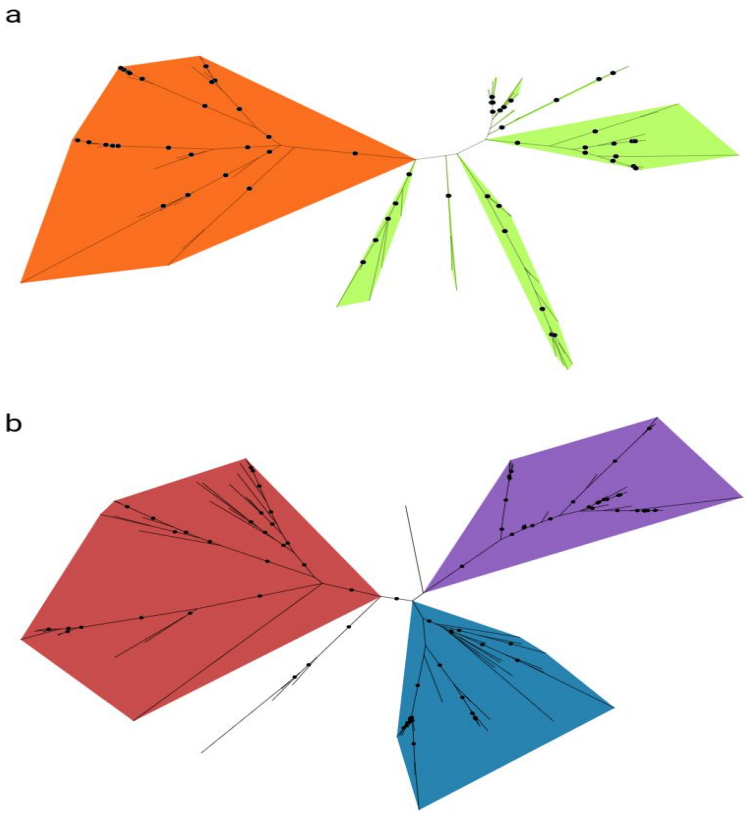


Figure 2. Unrooted phylogenetic trees of gig1 (top) and gig2 (bottom) genes across 11 species. Ortholog gig genes across species were retrieved from the Ensembl annotation gene gain/loss trees. The protein translation of the coding sequence for the longest transcript was aligned using GUIDANCE2 and unreliable residues were removed based on the default thresholds (0.93). Phylogenetic trees were built in IQ-TREE v2.3.3 using the alignments and 1000 replicates to calculate bootstrap values, shown as filled circles for values ≥ 80%. The 'Model Finder' option was used to detect the best-fit substitution model for each tree. 2a. Sequence divergence within the gig1 family suggests one possible split in the middle of the tree forming two separate clades, hereby named gig1A (orange) and gig1B (green). 2b. Sequence divergence within the gig2 family suggests three possible splits in the middle of the tree forming two separate clades, hereby named gig2A (red) and gig2B (blue) and gig2C (purple).

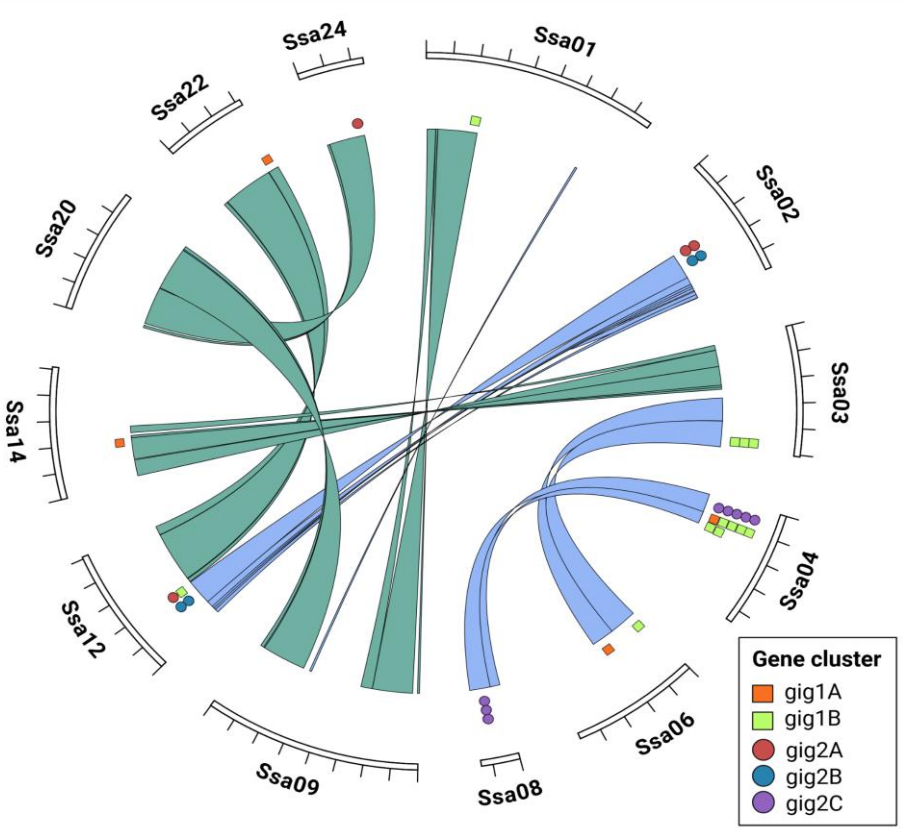


Figure 3 Circos plot of the physical location of gig1 (square) and gig2 (circle) genes in the salmonid genome. Colored shapes represent gene clusters based on phylogenetic analyses. Within the center of the plot, green ribbons link together genomic regions that underwent re-diploidization early after the Ss4R WGD event, (~100mya, ancestral-specific ohnolog resolution ,AORe) while blue ribbons connect the regions that underwent late re-diploidization within the last 50mya (lineage-specific ohnolog resolution - LORe) (Gundappa *et al.* 2022).

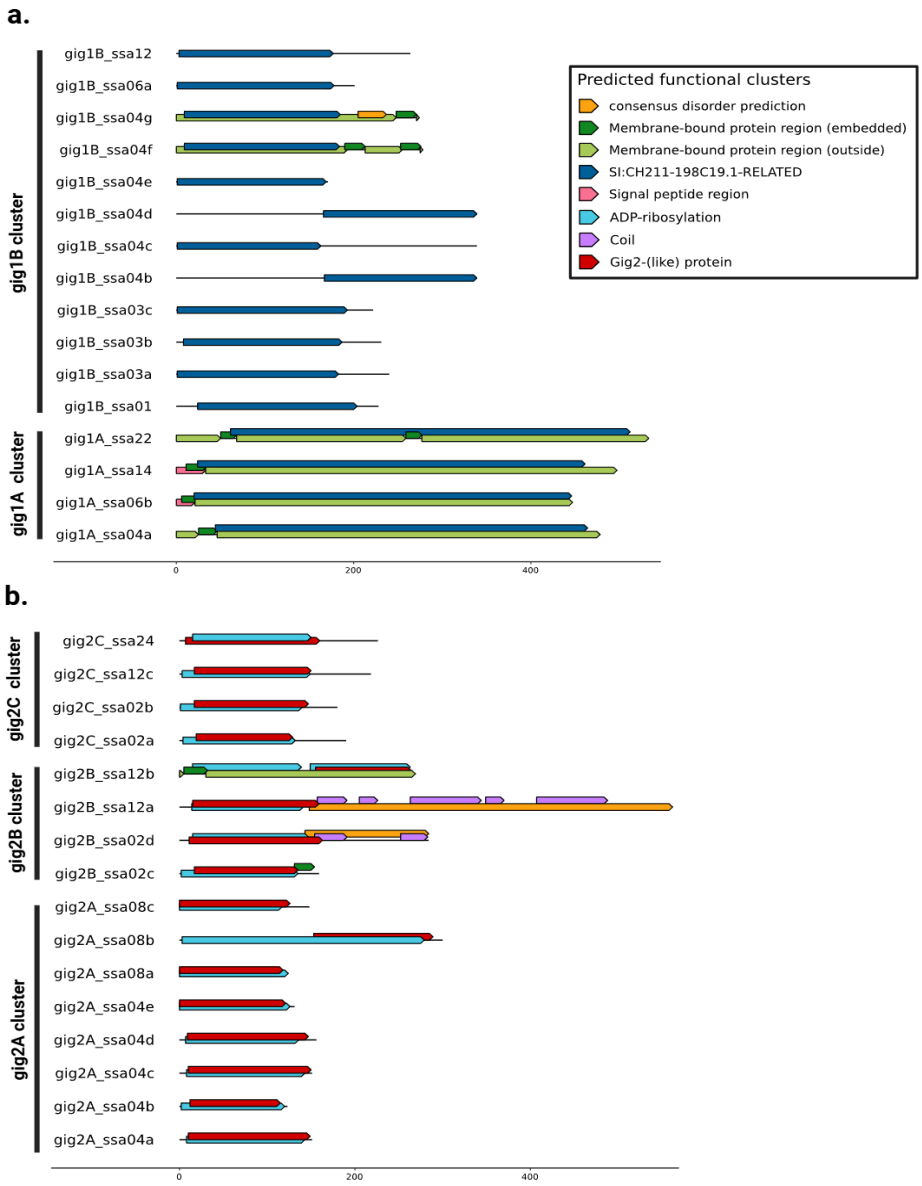


Figure 4 Predicted domains visualized for Atlantic salmon gig1 (a) and gig2 (b) ortholog genes. Colored bars and arrows indicate the functional group of a predicted domain whereas salmonid ortholog names are highlighted in respect to their phylogenetic clade. Black lines distinguish gene clusters as those were defined based on phylogenetic analysis.

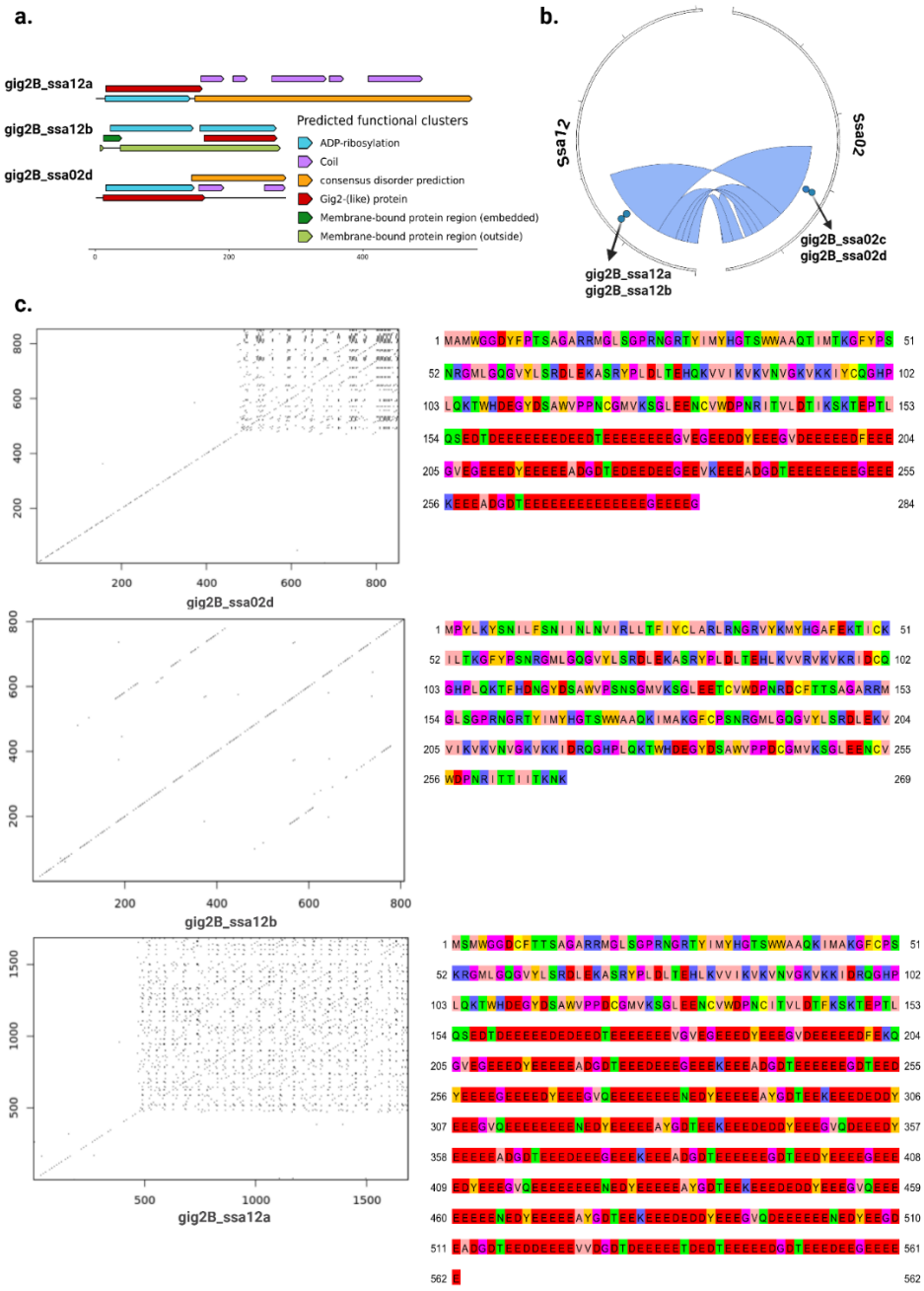

a.
gig2B_ssa12a
gig2B_ssa12b
gig2B_ssa02d
Predicted functional clusters
ADP-ribosylation
Coil
consensus disorder prediction
Gig2-(like) protein
Membrane-bound protein region (embedded)
Membrane-bound protein region (outside)
b.
Ssa12
Ssa02
gig2B_ssa02c
gig2B_ssa02d
gig2B_ssa12a
gig2B_ssa12b
c.
gig2B_ssa02d
gig2B_ssa12b
gig2B_ssa12a

Figure 5. Repeat content and predicted functional domains for gig2 paralogs located on the gig2B phylogenetic clade. (a.) Functional domain prediction based on the amino acid translation of gig coding sequence. (b.) Physical position of paralogs in the salmonid genome. Blue ribbons indicate chromosomal regions of late rediploidization (LORe). (c.) On the left, the self-alignment of the coding sequence for gig2 paralogs is shown while on the right, the amino acid translation of the aligned sequence for the respective genes is displayed.

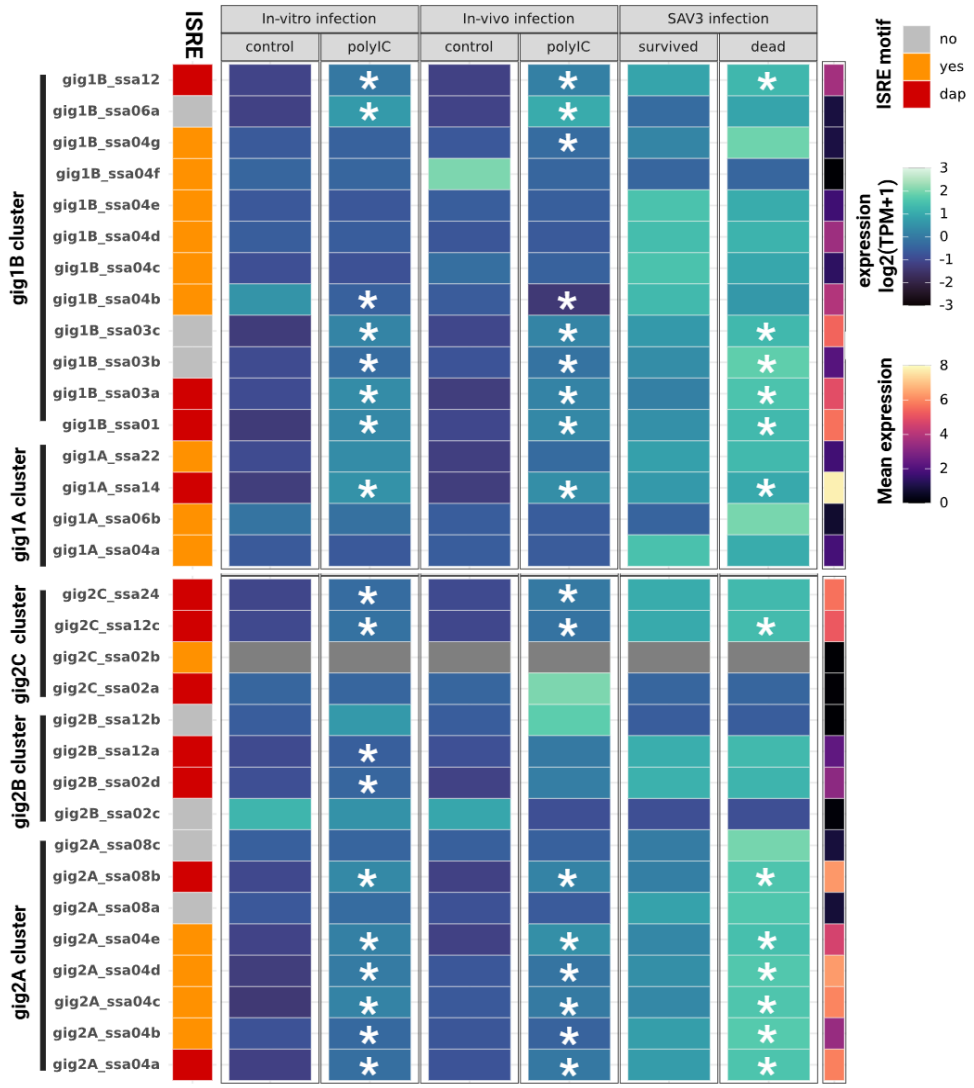


Figure 6. Comparative transcriptomics analysis of gig1 and gig2 genes in Atlantic salmon. Gig genes are sorted based on phylogenetic clustering. The expression panel shows the mean gene expression of gig genes across head kidney samples treated in vitro (cells) and in vivo (tissue) with poly(I:C) in addition to heart tissue transcriptome data from samples challenged with PD disease (SAV3 virus) in a 11-day survival trial. Colored boxes indicate mean gene expression of sample replicates in the range of 0 to 8 log2(TPM + 1) values. A discrete color annotation is

used to indicate genes with at least one detected IFN stimulation response element (ISRE) motif located within 2kb and 200bp upstream and downstream, respectively, from the TSS of each gig gene. Red, orange and grey colors inside the box indicate ISRE motifs associated with differentially regulated (dap) regions, ISRE motifs without association to dap regions, and lack of ISRE motifs, respectively.

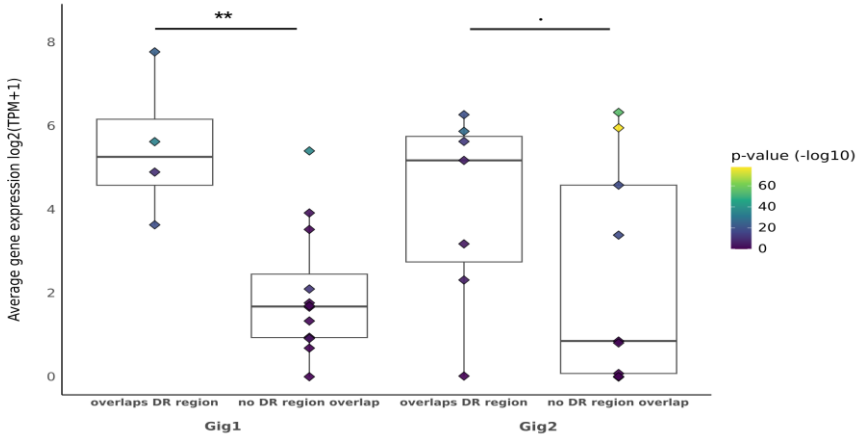
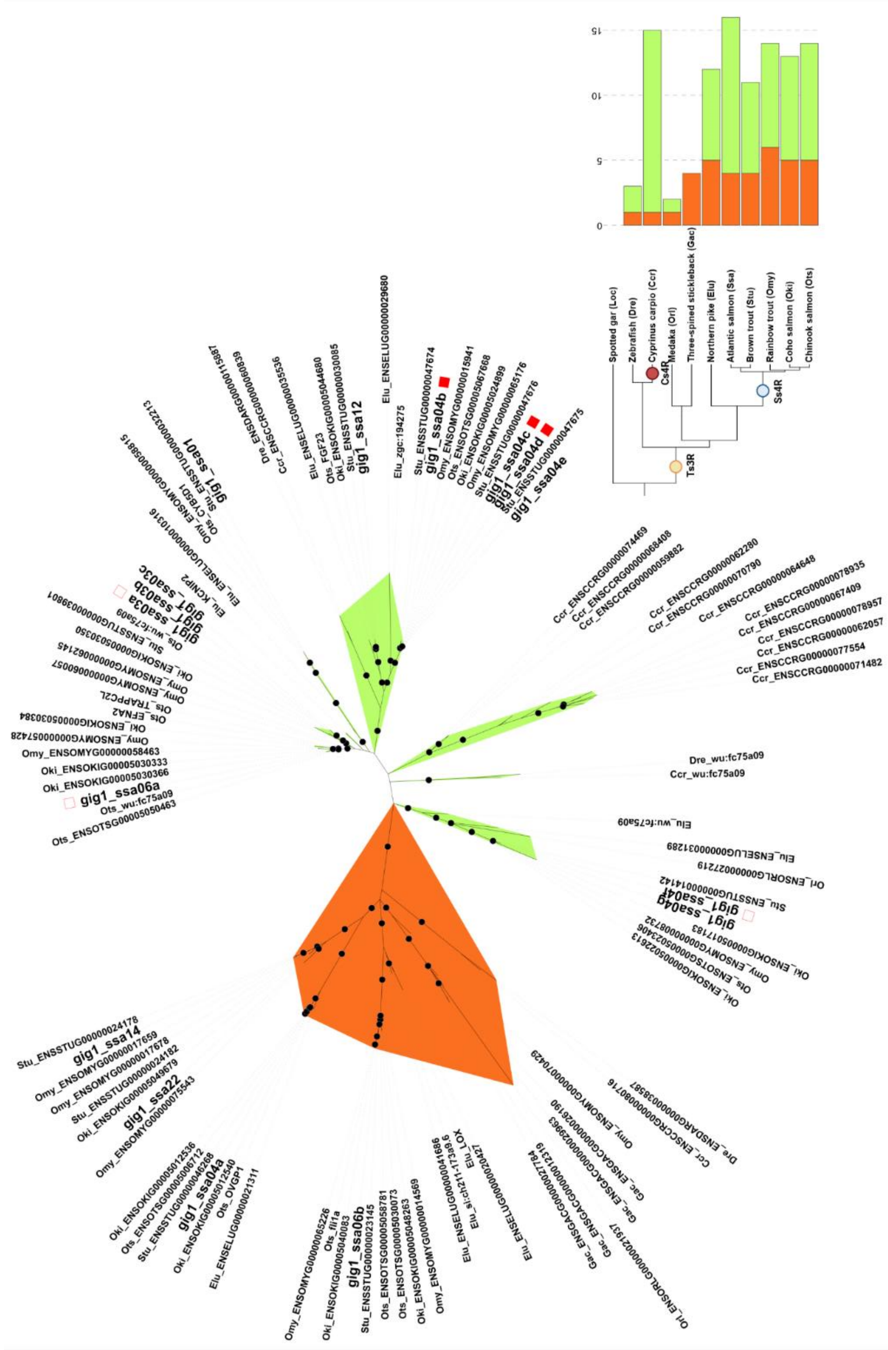
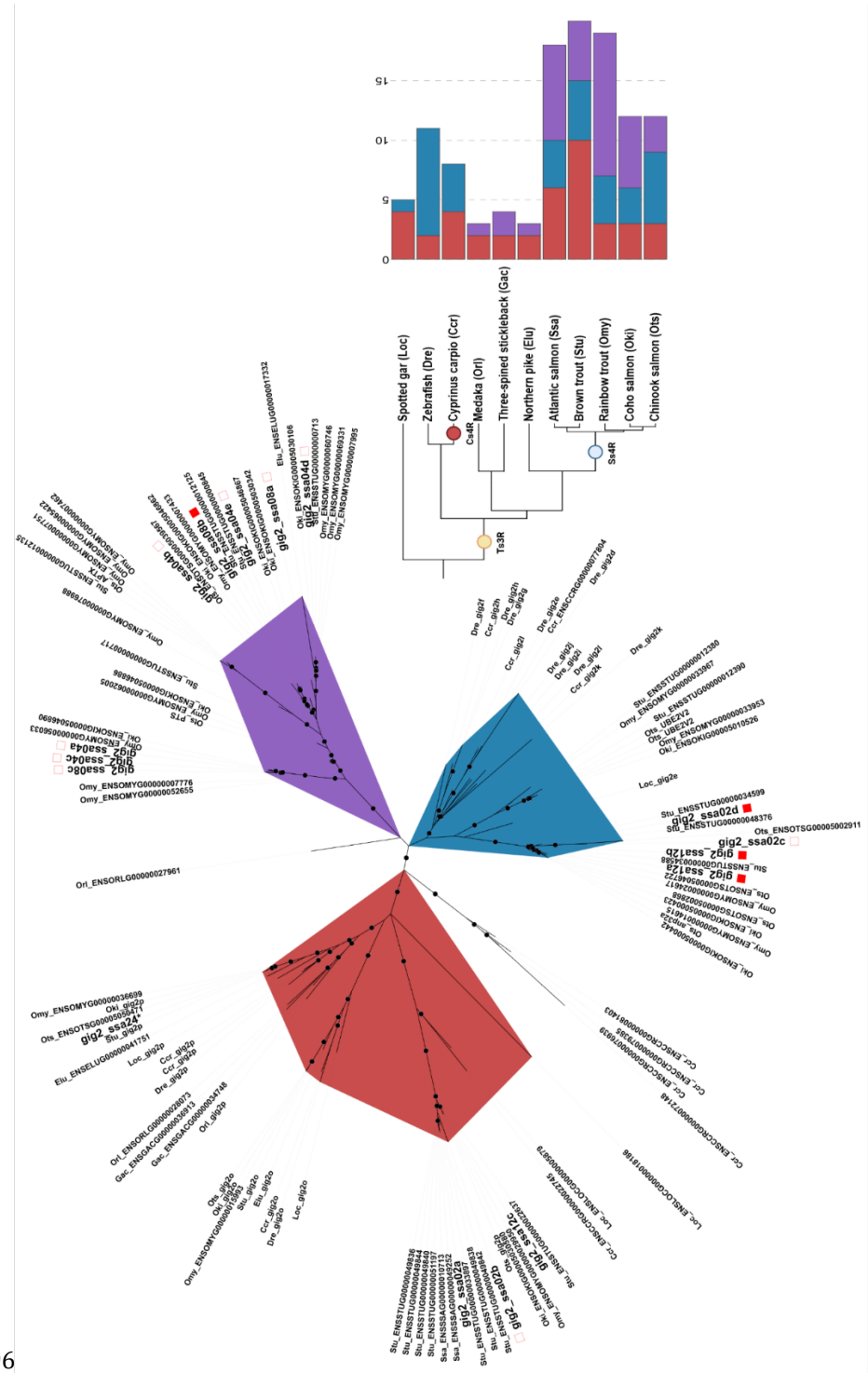


Figure 7. Distribution of genes with regulatory machinery near the TSS region. Different boxplots indicate genes with or without differentially regulated (DR) machinery during IFN stimulation. Additionally, coloured points represent differentially or non-differentially expressed *gig1* and *gig2* genes with their respective average expression shown in Y axis. Asterisks indicate statistical significance (p-value < 0.05) in comparison of average gene expression and DR regulator for each of the two *gig* families using a Wilcoxon rank-sum test.

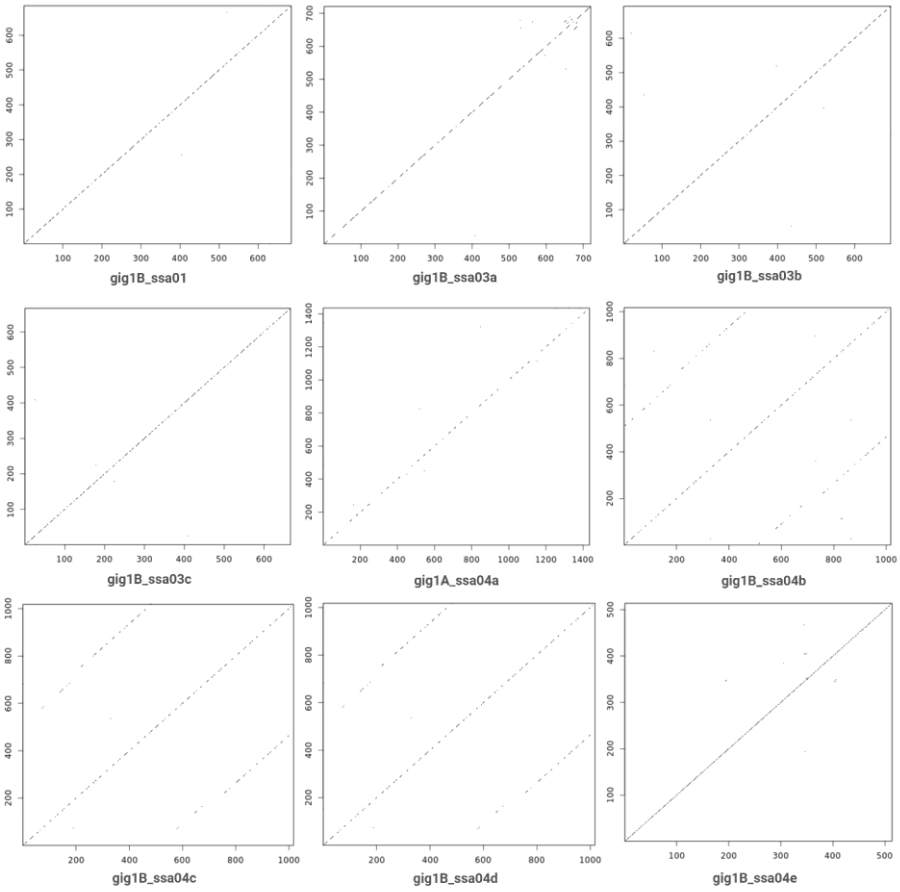
Supplementary Figures



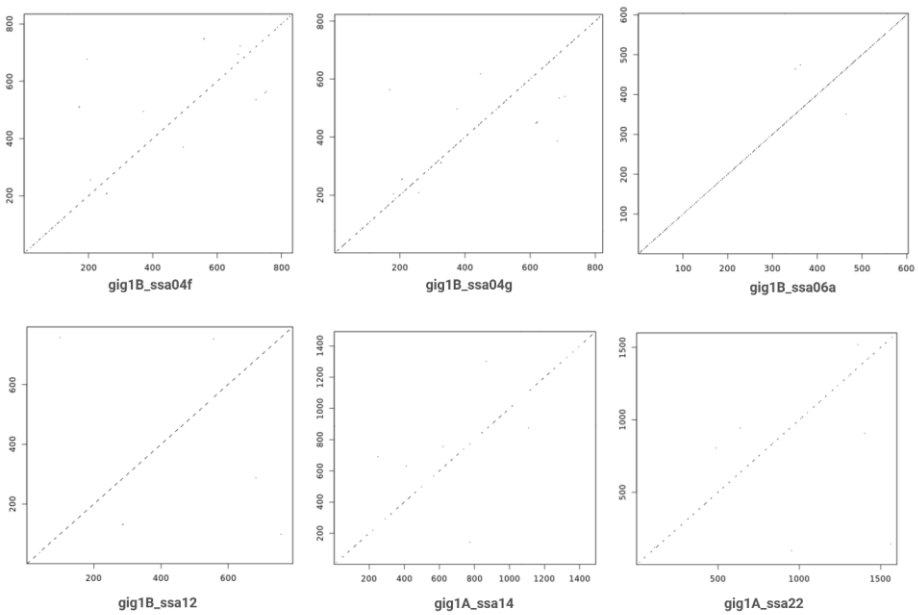
Supplementary Figure 1. Unrooted phylogenetic tree of gig1 genes detected across 11 species. Ortholog genes across species were retrieved from the Ensembl annotation gene gain/loss trees. Gene names were retrieved using Biomart. The protein translation of the coding sequence for the longest transcript of filtered genes was aligned using GUIDANCE2 and unreliable residues were removed based on the default threshold of 0.93. A phylogenetic tree was built in IQ-TREE v2.3.3 using the filtered alignment and 1000 replicates to calculate bootstrap values. The 'Model Finder' option was additionally applied to identify and use the best-fit substitution model for the tree. Sequence divergence within the gig1 family suggests one possible split in the middle of the tree forming two separate clades, hereby named gig1A (orange) and gig1B (green). Red squares indicate repeat content in the coding sequence of the respective gene with filled squares indicating large repeats and empty squares showing small repeats. The number of gig1 genes across species are shown in the bottom right bar chart where colored ranges indicate the respective gene clusters and yellow, red and blue annotations represent the teleost-specific (Ts3R), carp-specific (Cs4R) and salmonid-specific (Ss4R) whole genome duplication events, respectively.


Spotted gar (Loc)
Zebrafish (Dre)
Cyprinus carpio (Ccr)
Medaka (Orl)
Three-spined stickleback (Gac)
Northern pike (Elu)
Atlantic salmon (Ssa)
Brown trout (Stu)
Rainbow trout (Omy)
Coho salmon (Oki)
Chinook salmon (Ots)
Cs4R
Ts3R
Ss4R
0
5
10
15

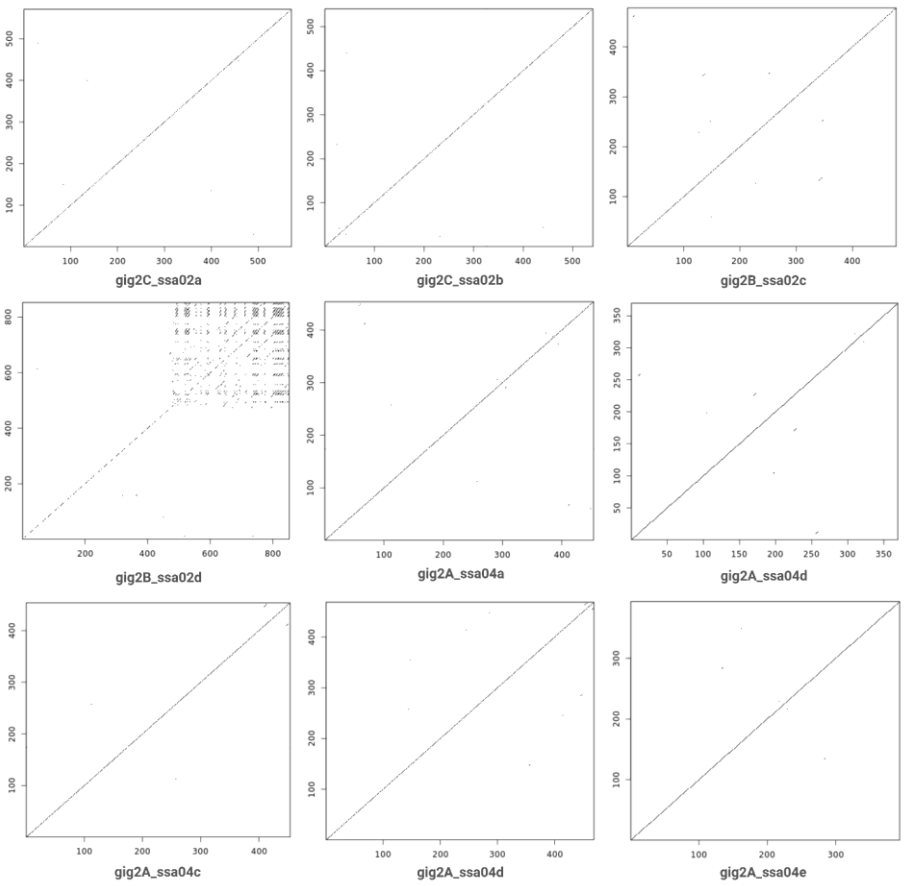
Supplementary Figure 2. Unrooted phylogenetic tree of gig2 genes detected across 11 species. Ortholog genes across species were retrieved from the Ensembl annotation gene gain/loss trees whereas gene names were retrieved using Biomart. The protein translation of the coding sequence for the longest transcript of each gene was aligned using GUIDANCE2 and unreliable residues were removed based on the default reliability threshold of 0.93. A phylogenetic tree was built in IQ-TREE v2.3.3 using the filtered alignments and 1000 replicates to calculate bootstrap values. In addition, the 'Model Finder' option was applied to identify the best-fit convergence model for the tree. Sequence divergence within the gig2 family suggested three possible splits in the middle of the tree forming three separate clades, hereby named gig2A (red) and gig2B (blue) and gig2C (purple). Red squares indicate repeat content in the coding sequence with filled squares indicating large repeats and empty squares showing small repeats. The number of gig2 genes across species are shown in the bar chart where colored ranges indicate the respective gene clusters and yellow, red and blue tree annotations represent the teleost-specific (Ts3R), carp-specific (Cs4R) and salmonid-specific (Ss4R) whole genome duplication events, respectively.



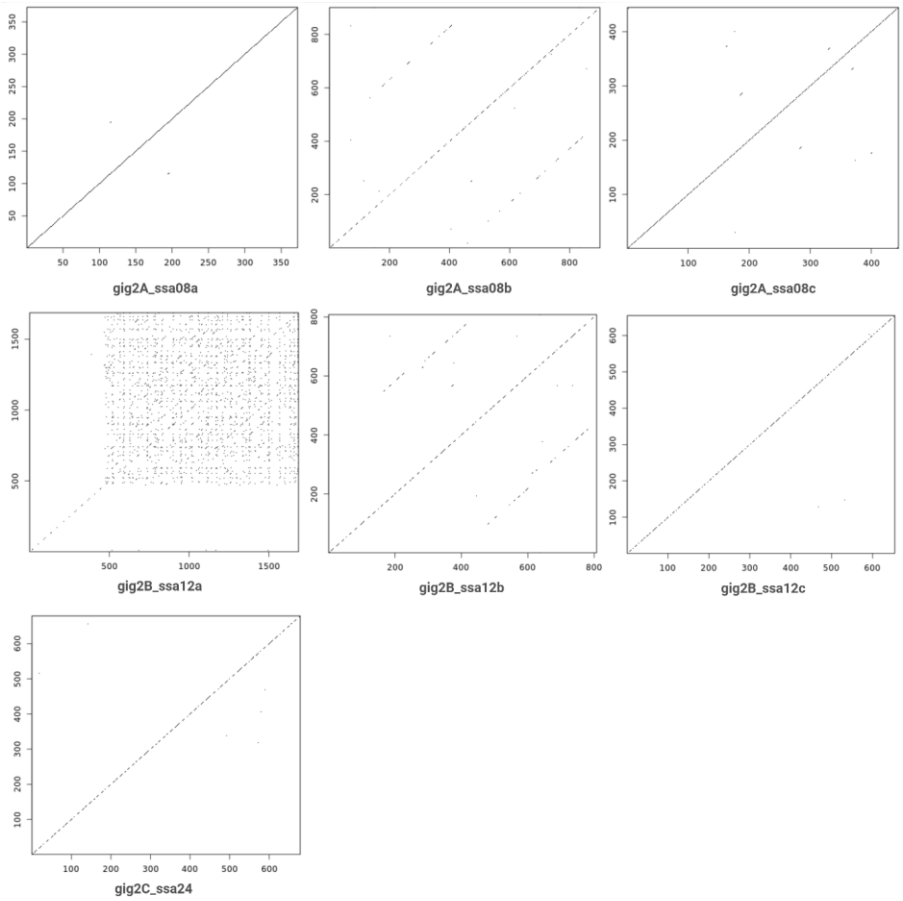
Supplementary Figure 3 Pairwise comparison of the coding sequence of gig1 genes. To identify repeat content within gig paralogs, the nucleotide sequences were analysed using a self-comparison approach to detect internal sequence repetitions.



Supplementary Figure 4 Pairwise comparison of the coding sequence of gig1 genes. To identify repeat content within gig paralogs, the nucleotide sequences were analysed using a self-comparison approach to detect internal sequence repetitions.



Supplementary Figure 5 Pairwise comparison of the coding sequence of gig2 genes. To identify repeat content within gig paralogs, the nucleotide sequences were analysed using a self-comparison approach to detect internal sequence repetitions.



Supplementary Figure 6 Pairwise comparison of the coding sequence of gig3 genes. To identify repeat content within gig paralogs, the nucleotide sequences were analysed using a self-comparison approach to detect internal sequence repetitions.

Supplementary Tables

Supplementary Table S1. Predicted functional features using the InterPro database and the amino acid translation of gig1 and gig2 genes across 11 species. The top section displays functional domains predicted across all species, whereas the bottom section focuses in domains identified in Atlantic salmon gig paralogs. Tables contain the name of predicted domains, the database from each they were retrieved, as well as the gig family and the abundancy each domain was discovered. Finally, Functional clusters denote the classification used in this analysis to cluster together similarly functional domains retrieved from multiple databases.

Predicted domain	Detected in family	n genes (gig1 / gig2)	Functional cluster	database
SI:CH211-198C19.1-RELATED	*gig1*	104/0	SI:CH211-198C19.1-RELATED	PANTHER
Consensus disorder prediction	*gig1 and gig2*	21/19	Consensus disorder prediction	MobiDBLite
Region of a membrane-bound protein predicted to be embe	*gig1 and gig2*	71/11	Membrane -bound protein region	Phobius
SignalP-noTM	*gig1 and gig2*	31/0	Signal peptide region	SignalP_EUK/SignalP_GRAM_NEGATIVE
Hydrophobic region of a signal peptide	*gig1 and gig2*	30/1	Signal peptide region	Phobius
Region of a membrane-bound protein predicted to be outsid	*gig1 and gig2*	54/7	Membrane -bound protein region	Phobius
C-terminal region of a signal peptide	*gig1 and gig2*	30/1	Signal peptide region	Phobius
Region of a membrane – bound protein predicted to b outsi	*gig1 and gig2*	30/5	Membrane -bound protein region	Phobius
N-terminal region of a signal peptide	*gig1 and gig2*	39/1	Signal peptide region	Phobius
Signal peptide region	*gig1 and gig2*	30/1	Signal peptide region	Phobius
SignalP-TM	*gig1*	3/0	Signal peptide region	SignalP_GRAM_NEGATIVE/SignalP_GRAM_PO
Coil	*gig1 and gig2*	1/27	Coil	Coils
ADP-ribosylation	*gig2*	0/173	ADP-ribosylation pathway	SUPERFAMILY
Poly(ADP-ribose) polymerase catalytic domain	*gig2*	0/123	ADP-ribosylation pathway	Pfam
Diptheria Toxin, Domain 1	*gig2*	0/172	ADP-ribosylation pathway	Gene3D
GIG2-like protein DRED-RELATED	*gig2*	0/146	Gig2-(like) protein	PANTHER
PARP catalytic domain profile	*gig2*	0/5	Gig2-(like) protein	ProSiteProfiles
Grass carp reovirus (GCRV) – induced gene 2e	*gig2*	0/38	Gig2-(like) protein	FunFam
Uncharacterized protein	*gig2*	0/5	Gig2-(like) protein	FunFam
Grass carp reovirus (GCRV) – induced gene 2p	*gig2*	0/8	Gig2-(like) protein	FunFam

Predicted domain	Detected in family	Salmon genes (16 gig1 / 18 gig2)	Functional cluster	database
ADP-ribosylation	*gig2*	0/18	ADP-ribosylation pathway	SUPERFAMILY
Poly(ADP-ribose) polymerase catalytic domain	*gig2*	0/11	ADP-ribosylation pathway	Pfam
Diptheria Toxin, Domain 1	*gig2*	0/18	ADP-ribosylation pathway	Gene3D
Coil	*gig1 and gig2*	0/7	Coil	Coils
Consensus disorder prediction	*gig1 and gig2*	1/4	Consensus disorder prediction	MobiDBLite
GIG2-like protein DRED-RELATED	*gig2*	0/16	Gig2-(like) protein	PANTHER
PARP catalytic domain profile	*gig2*	0/2	Gig2-(like) protein	ProSiteProfiles
Grass carp reovirus (GCRV) – induced gene 2e	*gig2*	0/2	Gig2-(like) protein	FunFam
Region of a membrane-bound protein predicted to be embe	*gig1 and gig2*	11/3	mbrane -bound protein region (embedd	Phobius
Region of a membrane-bound protein predicted to be outsid	*gig1 and gig2*	6/1	lembrane -bound protein region (outside	Phobius
Region of a membrane – bound protein predicted to b outsi	*gig1 and gig2*	6/1	lembrane -bound protein region (outside	Phobius
SI:CH211-198C19.1-RELATED	*gig1*	16/0	SI:CH211-198C19.1-RELATED	PANTHER
SignalP-noTM	*gig1 and gig2*	1/0	Signal peptide region	SignalP_EUK/SignalP_GRAM_NEGATIVE
Hydrophobic region of a signal peptide	*gig1 and gig2*	0/2	Signal peptide region	Phobius
C-terminal region of a signal peptide	*gig1 and gig2*	2/0	Signal peptide region	Phobius
N-terminal region of a signal peptide	*gig1 and gig2*	2/0	Signal peptide region	Phobius
Signal peptide region	*gig1 and gig2*	2/0	Signal peptide region	Phobius

Data Availability

Raw short-read transcriptomics data related to the SAV3 infectious cohabitation challenge is available on the European Nucleotide archive under accession number PRJEB85594. Raw material for the analysis of the Atlantic salmon regulatory landscape is available on the European Nucleotide Archive under accession numbers PRJEB50077 and PRJEB56698 for ATAC-seq and ChIP-seq information, respectively. Raw short-read transcriptomics data related to the poly(I:C) challenge of Atlantic salmon head kidney cells and tissue are available on the European Nucleotide archive under accession number PRJEB50076.

Acknowledgements

The study was supported by The Research Council of Norway (grant nos. 325874). DM was supported by the Norwegian University of Life Sciences (NMBU), BIOVIT, PhD funding. We acknowledge the use of the Orion computing cluster at the Norwegian University of Life Sciences (NMBU).

References

An, L.-L., X.-Y. Gong, C. Dan, H.-Y. Sun, W.-H. Guo *et al.*, 2025 Family evolution and functional divergence of bony fish-specific Gig1 homologs. Water Biology and Security 100382.

Banet, N. V., J. Fieberg, and P. W. Sorensen, 2022 Migration, homing and spatial ecology of common carp in interconnected lakes. Ecol. Freshw. Fish 31: 164–176.

Braasch, I., A. R. Gehrke, J. J. Smith, K. Kawasaki, T. Manousaki *et al.*, 2016 The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat. Genet. 48: 427–437.

Charif, D., and J. R. Lobry, 2007 SeqinR 1.0-2: A contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis, pp. 207–232 in *Structural Approaches to Sequence Evolution*, Biological and Medical Physics, Biomedical Engineering, Springer Berlin Heidelberg, Berlin, Heidelberg.

Christensen, K. A., J. S. Leong, D. Sakhrani, C. A. Biagi, D. R. Minkley *et al.*, 2018 Chinook salmon (Oncorhynchus tshawytscha) genome and transcriptome. PLoS One 13: e0195461.

Clark, T. C., S. Naseer, M. K. Gundappa, A. Laurent, A. Perquis *et al.*, 2023 Conserved and divergent arms of the antiviral response in the duplicated genomes of salmonid fishes. Genomics 115: 110663.

Collet, B., and C. J. Secombes, 2001 The rainbow trout (*Oncorhynchus mykiss*) *Mx1* promoter. Eur. J. Biochem. 268: 1577–1584.

Dalskov, L., H. H. Gad, and R. Hartmann, 2023 Viral recognition and the antiviral interferon response. EMBO J. 42: e112907.

Dobin, A., C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski *et al.*, 2013 STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29: 15–21.

Dornburg, A., D. J. Wcisel, K. Zapfe, E. Ferraro, L. Roupe-Abrams *et al.*, 2021 Holosteans contextualize the role of the teleost genome duplication in promoting the rise of evolutionary novelties in the ray-finned fish innate immune system. bioRxiv.

Gao, G., S. Magadan, G. C. Waldbieser, R. C. Youngblood, P. A. Wheeler *et al.*, 2021 A long reads-based de-novo assembly of the genome of the Arlee homozygous line reveals chromosomal rearrangements in rainbow trout. G3 (Bethesda) 11.:

Gao, F.-X., Y. Wang, Q.-Y. Zhang, C.-Y. Mou, Z. Li *et al.*, 2017 Distinct herpesvirus resistances and immune responses of three gynogenetic clones of gibel carp revealed by comprehensive transcriptomes. BMC Genomics 18: 561.

Glasauer, S. M. K., and S. C. F. Neuhauss, 2014 Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol. Genet. Genomics 289: 1045–1060.

Gomez, D., J. O. Sunyer, and I. Salinas, 2013 The mucosal immune system of fish: the evolution of tolerating commensals while fighting pathogens. Fish Shellfish Immunol. 35: 1729–1739.

Graham, L. A., S. Y. Gauthier, and P. L. Davies, 2022 Origin of an antifreeze protein gene in response to Cenozoic climate change. Sci. Rep. 12: 8536.

Gundappa, M. K., T.-H. To, L. Grønvold, S. A. M. Martin, S. Lien *et al.*, 2022 Genome-wide reconstruction of rediploidization following autopolyploidization across one hundred million years of Salmonid evolution. Mol. Biol. Evol. 39.:

Hansen, T., P. G. Fjelldal, S. Lien, M. Smith, C. Corton *et al.*, 2021 The genome sequence of the brown trout, Salmo trutta Linnaeus 1758. Wellcome Open Res. 6: 108.

Hillestad, B., S. Makvandi-Nejad, A. Krasnov, and H. K. Moghadam, 2020 Identification of genetic loci associated with higher resistance to pancreas disease (PD) in Atlantic salmon (Salmo salar L.). BMC Genomics 21: 388.

Hoang, D. T., O. Chernomor, A. von Haeseler, B. Q. Minh, and L. S. Vinh, 2018 UFBoot2: Improving the ultrafast bootstrap approximation. Mol. Biol. Evol. 35: 518–522.

Holen, M. M., G. Vaaje-Kolstad, M. P. Kent, and S. R. Sandve, 2023 Gene family expansion and functional diversification of chitinase and chitin synthase genes in Atlantic salmon (Salmo salar). G3 (Bethesda) 13.:

Ichikawa, K., S. Tomioka, Y. Suzuki, R. Nakamura, K. Doi *et al.*, 2017 Centromere evolution and CpG methylation during vertebrate speciation. Nat. Commun. 8: 1833.

Iwama, R. E., and Y. Moran, 2023 Origins and diversification of animal innate immune responses against viral infections. Nat Ecol Evol.

Johnston, I. A., M. P. Kent, P. Boudinot, M. Looseley, L. Bargelloni *et al.*, 2024 Advancing fish breeding in aquaculture through genome functional annotation. Aquaculture 583: 740589.

Kalyaanamoorthy, S., B. Q. Minh, T. K. F. Wong, A. von Haeseler, and L. S. Jermiin, 2017 ModelFinder: fast model selection for accurate phylogenetic estimates. Nat. Methods 14: 587–589.

Krasnov, A., L.-H. Johansen, C. Karlsen, L. Sveen, E. Ytteborg *et al.*, 2021 Transcriptome responses of Atlantic salmon (Salmo salar L.) to viral and bacterial pathogens, inflammation, and stress. Front. Immunol. 12: 705601.

Krasnov, A., G. Timmerhaus, B. L. Schiøtz, J. Torgersen, S. Afanasyev *et al.*, 2011

Genomic survey of early responses to viruses in Atlantic salmon, Salmo salar L. Mol. Immunol. 49: 163–174.
Letunic, I., and P. Bork, 2021 Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 49: W293–W296.
Levraud, J.-P., L. Jouneau, V. Briolat, V. Laghi, and P. Boudinot, 2019 IFN-Stimulated Genes in Zebrafish and Humans Define an Ancient Arsenal of Antiviral Immunity. J. Immunol. 203: 3361–3373.
Lien, S., B. F. Koop, S. R. Sandve, J. R. Miller, M. P. Kent *et al.*, 2016 The Atlantic salmon genome provides insights into rediploidization. Nature 533: 200–205.
Love, M. I., W. Huber, and S. Anders, 2014 Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15: 550.
Madushani, K. P., K. A. S. N. Shanaka, R. K. Madusanka, and J. Lee, 2021 Molecular characterization and expressional analysis of two poly (ADP-ribose) polymerase (PARP) domain-containing Gig2 isoforms in rockfish (Sebastes schlegelii) and their antiviral activity against viral hemorrhagic septicemia virus. Fish Shellfish Immunol. 118: 219–227.
Manousi, D., D. M. Jaskula, F. Grammes, T. M. Knutsen, S. Naseer *et al.*, 2025 Dissecting the genetic basis of response to salmonid alphavirus in Atlantic salmon. Research Square.
Mayer, I., and M. Pšenička, 2024 Conservation of teleost fishes: Application of reproductive technologies. Theriogenology Wild 4: 100078.
Nath, S., D. E. Shaw, and M. A. White, 2021 Improved contiguity of the threespine stickleback genome using long-read sequencing. G3 (Bethesda) 11.:
Newell, R., R. Pienaar, B. Balderson, M. Piper, A. Essebier *et al.*, 2020 ChIP-R: Assembling reproducible sets of ChIP-seq and ATAC-seq peaks from multiple replicates. Bioinformatics.
Nguyen, L.-T., H. A. Schmidt, A. von Haeseler, and B. Q. Minh, 2015 IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol. Biol. Evol. 32: 268–274.
Pajic, P., S. Shen, J. Qu, A. J. May, S. Knox *et al.*, 2022 A mechanism of gene evolution generating mucin function. Sci. Adv. 8: eabm8757.
Parey, E., A. Louis, J. Montfort, Y. Guiguen, H. Roest Crollius *et al.*, 2022 An atlas of fish genome evolution reveals delayed rediploidization following the teleost whole-genome duplication. Genome Res. 32: 1685–1697.
Patel, H., P. Ewels, J. Manning, M. U. Garcia, A. Peltzer *et al.*, 2024 *nf-core/rnaseq: nf-core/rnaseq v3.18.0 - Lithium Lynx*. Zenodo.
Paysan-Lafosse, T., M. Blum, S. Chuguransky, T. Grego, B. L. Pinto *et al.*, 2023 InterPro in 2022. Nucleic Acids Res. 51: D418–D427.
Purushotham, J. N., H. L. Lutz, E. Parker, and K. G. Andersen, 2025 Immunological drivers of zoonotic virus emergence, evolution, and endemicity. Immunity 58: 784–796.
Qi, M., J. Clark, E. R. R. Moody, D. Pisani, and P. C. J. Donoghue, 2024 Molecular dating of the teleost whole genome duplication (3R) is compatible with the expectations of delayed rediploidization. Genome Biol. Evol. 16.:
Rao, Y., and J. Su, 2015 Insights into the antiviral immunity against grass carp (Ctenopharyngodon idella) reovirus (GCRV) in grass carp. J. Immunol. Res. 2015: 670437.

Ray-Jones, H., and M. Spivakov, 2021 Transcriptional enhancers and their communication with gene promoters. Cell. Mol. Life Sci. 78: 6453–6485.
R Core Team, 2021 *R: A language and environment for statistical computing*.
Redmond, A. K., J. Zou, C. J. Secombes, D. J. Macqueen, and H. Dooley, 2019 Discovery of all three types in cartilaginous fishes enables phylogenetic resolution of the origins and evolution of interferons. Front. Immunol. 10: 1558.
Rivera, A. M., and W. J. Swanson, 2022 The importance of gene duplication and domain repeat expansion for the function and evolution of fertilization proteins. Front. Cell Dev. Biol. 10: 827454.
Rives, N., V. Lamba, C. H. C. Cheng, and X. Zhuang, 2024 Diverse origins of near-identical antifreeze proteins in unrelated fish lineages provide insights into evolutionary mechanisms of new gene birth and protein sequence convergence. Mol. Biol. Evol. 41.:
Robertsen, B., 2006 The interferon system of teleost fish. Fish Shellfish Immunol. 20: 172–191.
Robinson, N. A., A. Krasnov, E. Burgerhout, H. Johnsen, H. K. Moghadam *et al.*, 2020 Response of the Salmon Heart Transcriptome to Pancreas Disease: Differences Between High- and Low-Ranking Families for Resistance. Sci. Rep. 10: 868.
Robinson, P., and T. Zemo jtel, 2017 Integrative genomics viewer (IGV): Visualizing alignments and variants, pp. 233–245 in *Computational Exome and Genome Analysis*, Chapman and Hall/CRC.
Salinas, I., Y.-A. Zhang, and J. O. Sunyer, 2011 Mucosal immunoglobulins and B cells of teleost fish. Dev. Comp. Immunol. 35: 1346–1365.
Sandholm, R. M., 2022 Evolution of alternative splice variation and exon usage following whole genome duplication: Norwegian University of Life Sciences, Ås.
Sela, I., H. Ashkenazy, K. Katoh, and T. Pupko, 2015 GUIDANCE2: accurate detection of unreliable alignment regions accounting for the uncertainty of multiple parameters. Nucleic Acids Res. 43: W7–14.
Shultz, A. J., and T. B. Sackton, 2019 Immune genes are hotspots of shared positive selection across birds and mammals. eLife 8.:
Stosik, M., B. Tokarz-Deptuła, and W. Deptuła, 2023 Immunity of the intestinal mucosa in teleost fish. Fish Shellfish Immunol. 133: 108572.
Sun, C., Y. Liu, Y. Hu, Q. Fan, W. Li *et al.*, 2013 Gig1 and Gig2 homologs (CiGig1 and CiGig2) from grass carp (Ctenopharyngodon idella) display good antiviral activities in an IFN-independent pathway. Dev. Comp. Immunol. 41: 477–483.
Sun, B., B. Robertsen, Z. Wang, and B. Liu, 2009 Identification of an Atlantic salmon IFN multigene cluster encoding three IFN subtypes with very different expression properties. Dev. Comp. Immunol. 33: 547–558.
Sun, F., Y.-B. Zhang, J. Jiang, B. Wang, C. Chen *et al.*, 2014 Gig1, a novel antiviral effector involved in fish interferon response. Virology 448: 322–332.
Tasnim, M., P. Wahlquist, and J. T. Hill, 2024 Zebrafish: unraveling genetic complexity through duplicated genes. Dev. Genes Evol. 234: 99–116.
Teekas, L., S. Sharma, and N. Vijay, 2022 Lineage-specific protein repeat expansions and contractions reveal malleable regions of immune genes. Genes Immun. 23: 218–234.
Timmerhaus, G., A. Krasnov, H. Takle, S. Afanasyev, P. Nilsen *et al.*, 2012 Comparison of

Atlantic salmon individuals with different outcomes of cardiomyopathy syndrome (CMS). BMC Genomics 13: 205.
Vollset, K. W., R. J. Lennox, J. G. Davidsen, S. H. Eldøy, T. E. Isaksen *et al.*, 2021 Wild salmonids are running the gauntlet of pathogens and climate as fish farms expand northwards. ICES J. Mar. Sci. 78: 388–401.
Walker, F. C., P. R. Sridhar, and M. T. Baldridge, 2021 Differential roles of interferons in innate responses to mucosal viral infections. Trends Immunol. 42: 1009–1023.
Wang, J., C. Hu, X. Tong, Y. Gao, R. Liang *et al.*, 2025 Microbial communities associated with the skin, gill, and gut of large yellow croaker (Larimichthys crocea). BMC Microbiol. 25: 16.
Zhang, Y., 2003 Identification of antiviral-relevant genes in the cultured fish cells induced by UV-inactivated virus. Chin. Sci. Bull. 48: 581.
Zhang, Y. B., and J. F. Gui, 2004 Identification of two novel interferon-stimulated genes from cultured CAB cells induced by UV-inactivated grass carp hemorrhage virus. Dis. Aquat. Organ. 60: 1–9.
Zhang, Y.-B., T.-K. Liu, J. Jiang, J. Shi, Y. Liu *et al.*, 2013 Identification of a novel Gig2 gene family specific to non-amniote vertebrates. PLoS One 8: e60588.

Paper 3

Parallel Selection in Domesticated Atlantic Salmon from Divergent Founders Including on Whole-Genome Duplication-derived Homeologous Regions

Pauline Buso [1], Célian Diblasi [2], Domniki Manousi [2], Jun Soung Kwak [2], Arturo Vera-Ponce de Leon [2], Kristina Stenløkk [2,3], Nicola Barson [2,*], Marie Saitou [2,*]

[1]IHPE, Université de Montpellier, CNRS, Ifremer, Université de Perpignan Via Domitia, Montpellier, France

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

[3]Present address: Department of Forensic Sciences, Oslo University Hospital, 0424 Oslo, Norway

*Corresponding authors: E-mails: nicola.barson@nmbu.no; marie.saitou@nmbu.no.

Accepted: March 26, 2025

Abstract

Domestication and artificial selection for desirable traits have driven significant phenotypic changes and left detectable genomic footprints in farmed animals. Since the 1960s, intensive breeding has led to the rapid domestication of Atlantic salmon (*Salmo salar*), with multiple independent events that make it a valuable model for studying early domestication stages and the parallel evolution of populations of different origins subjected to similar selection pressures. Some aquatic species, including Atlantic salmon, have undergone whole-genome duplication (WGD), raising the possibility that genetic redundancy resulting from WGD has contributed to adaptation in captive environments, as seen in plants. Here, we examined the genomic responses to domestication in Atlantic salmon, focusing on potential signatures of parallel selection, including those associated with WGD. Candidate genomic regions under selection were identified by comparing whole-genome sequences from aquaculture and wild populations across 2 independently domesticated lineages (Western Norway and North America) using a genome-wide scan that combined 3 statistical methods: allele frequencies (F_{ST}), site frequency (Tajima's D), and haplotype differentiation (XP-EHH). These analyses revealed shared selective sweeps on identical SNPs in major histocompatibility complex (MHC) genes across aquaculture populations. This suggests that a combination of long-term balancing selection and recent human-induced selection has shaped MHC gene evolution in domesticated salmon. Additionally, we observed selective sweeps on a small number of gene pairs in homeologous regions originating from WGD, offering insights into how historical genome duplication events may intersect with recent selection pressures in aquaculture species.

Key words: domestication, MHC, artificial selection, aquaculture, ohnolog.

Significance

Domestication rapidly alters the genomic architecture of farmed animals to enhance favorable traits, yet the early genetic responses to domestication are incompletely understood in many systems. By analyzing 2 independently domesticated Atlantic salmon lineages, this study identified shared genomic regions under selection across continents, particularly in immune-related genes. This suggests that independent domestication events can lead to similar genetic changes under artificial selection, highlighting the role of immune-related genes in Atlantic salmon evolution.

Introduction

Domestication marks one of the key shifts in human history, tracing back to the enduring bond between early hunter-gatherers and wolves over 15,000 year ago (Vigne 2011). Domesticated plants and animals became essential to human societies for social, nutritional, symbolic, or economic purposes, reflecting the development of multicultural societies that emerged at the end of the last ice age (Larson et al. 2014). Domestication refers to a small, selected fraction from a wild population adapting to a human-controlled environment. This process can change the genetic composition and phenotypes of farmed individuals in response to both intentional and unintentional selection (Wang et al. 2014). Small initial founder populations result in a strong bottleneck and limited genetic diversity. Consequently, the genomic evolution of domesticated species is influenced by strong genetic drift, as well as intentional artificial selection for favorable production traits (Innan and Kim 2004), and unintentional domestication selection resulting from adaptation to the captive environment and human contact (Mignon-Grasteau et al. 2005; Frantz et al. 2020; Huang et al. 2022). Thus, domesticated species carry signatures of these processes in their genomes, providing valuable insights into their domestication history and make them good models for genetic studies of evolution (Wiener and Wilkinson 2011).

Conventional investigations into the evolutionary changes that occurred after domestication began with archaeological records (Diamond 2002). Recent advances using Omics approaches have enabled researchers to trace the geographical and temporal origins of domestication and its early stages. These approaches have revealed the genomic basis of favored traits and selected loci in various species (Frantz et al. 2020), while also shedding light on the unintended consequences of these processes (Glover et al. 2017; Fernandez et al. 2021; Izawa 2022). Domestication processes are often associated with selection pressures on key production traits such as growth, immunity, and age of sexual maturity (Gjederm 1979; Milla et al. 2021; Podgorniak et al. 2022). Artificial selection for such traits and domestication selection may lead to parallel evolution between independent populations exposed to similar selection pressures. This refers to the independent development of genetic adaptations within different populations that affect the organism's phenotype in the same way (Wood et al. 2005; Bailey et al. 2017). For instance, high-altitude adaptation has been observed within independent human populations and their domesticated animals such as dogs, cows, horses, sheep, pigs, and chickens, highlighting a shared genetic response to environmental pressures (Witt and Huerta-Sánchez 2019). Similarly, mammals that have adapted to high-starch diets, including various domesticated animals, have an increased copy number of the amylase gene (*amy*) (Pajic et al. 2019).

Compared to the long history of terrestrial livestock domestication, the domestication of most aquaculture species is relatively recent. Finfish aquaculture expanded rapidly in the early 1980s, while domestication timelines vary across aquaculture species (Houston et al. 2020; Teletchea 2021). Unlike established terrestrial livestock species, aquaculture species present a varied landscape of domestication success, with some species achieving breakthroughs in captive breeding while others continue to rely on wild stock imports. Successful domestications in aquaculture systems are often observed in certain fish families, such as sturgeons, cyprinids (including carp), and salmonids. Notably, some of these families have undergone whole-genome duplication (WGD), which has been linked to increased genomic redundancy, allelic diversity, heterozygosity, and meiotic recombination (Lien et al. 2016; Gundappa et al. 2022). These genomic features may be one factor to enhance tolerance to new mutations and facilitate adaptation to human-mediated environments, while WGD is not essential for successful domestication (Salman-Minkov et al 2016; Baduel et al. 2019; Mihajlovic et al. 2025; Wang et al. 2024). Many successful aquaculture species, including catfish, sea bass, and flatfishes, among others, have achieved domestication without lineage-specific WGD events. Beyond fish, domesticated aquaculture species also include mollusks (e.g. oysters and mussels) and crustaceans (e.g. shrimp and crabs), which do not have known histories of lineage-specific WGD (Houston et al. 2020; Naylor et al. 2021). While WGD-derived genetic redundancy may still contribute to adaptation in some lineages, multiple genomic pathways facilitate domestication across diverse taxa.

The Atlantic salmon (*Salmo salar*) genome underwent a WGD event (Ss4R) ~100 million years ago, leading to the retention of a substantial fraction of duplicated genes (Macqueen and Johnston 2014; Gundappa et al. 2022). Unlike a random loss of duplicate genes, rediploidization followed a structured trajectory, resulting in a nonrandom spectrum of ohnolog divergence levels. These ohnologs are organized into collinear blocks across the genome, reflecting the evolutionary history of this process (Lien et al. 2016; Robertson et al. 2017). In this study, we use "homeolog" to refer to corresponding genomic regions and "ohnolog" to refer to duplicated genes arising from WGD events. While many duplicated genes were subsequently lost, a considerable proportion has been maintained as ohnolog pairs, contributing to functional diversification. Some ohnologs retained their ancestral function, while others underwent subfunctionalization or neofunctionalization, leading to functional differentiation and specialization. In addition to structural retention, the expression of Ss4R-derived ohnologs has evolved under selective pressures, leading to differential expression patterns. Some

ohnolog pairs exhibit complementary expression across tissues, while others have diverged functionally (Lien et al. 2016). Additionally, understanding these expression dynamics is crucial for interpreting how Ss4R-derived variation contributes to domestication and aquaculture adaptation, particularly in traits related to metabolism, immune function, and stress response.

Among the aquaculture species with duplicated genomes, Atlantic salmon, has been subject to strong artificial selection and has seen growing economic importance since the 1960s (Cross and King 1983; Ford and Myers 2008; Glover et al. 2017). 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. 2019). Wild populations of Atlantic salmon have been split into 4 main phylogeographical lineages, 3 in Europe (East Atlantic, Barents/White Sea, and Baltic) and 1 in North America (West Atlantic) (Bourret et al. 2013; Rougemont and Bernatchez 2018). These lineages exhibit karyotypic variation, with the North American populations displaying a polymorphism in chromosome numbers due to fusions (Brenna-Hansen et al. 2012). Atlantic salmon in Europe and North America have been independently domesticated multiple times from different source populations, with repeated artificial selection directed at similar traits under human-induced environments. This suggests that parallel evolution has likely occurred in domesticated Atlantic salmon populations across both continents, driven by comparable selective pressures.

Despite its recent domestication, studies of Atlantic salmon have revealed rapid genetic changes in response to artificial selection targeting metabolism, disease resistance, and faster growth (Debes et al. 2012; Bicskei et al. 2014; Harvey et al. 2018; Houston and Macqueen 2019; Jin et al. 2020; Bull et al. 2022). Moreover, signatures of domestication selection have also been observed in neurological genes (Bertolotti et al. 2020), causing behavioral alterations, such as increased sensitivity to predation and increased tolerance to stress compared with the wild (Solberg et al. 2020). Especially, the probability of parallel evolution is dependent on the richness of genetic variation, standing genetic variation when this occurs over short time periods (Wood et al. 2005; Bailey et al. 2017; Ebadi et al. 2023). Several studies have investigated parallel selection in Atlantic salmon across different geographic lineages. López et al. (2019) observed parallel evolution at 4 genes, e.g ubiquitin-conjugating enzyme E2F putative (*ube2f*), collagen alpha-1XIII chain (*coda1*), autism susceptibility 2 protein-like (*auts2-like*), and transient receptor potential cation channel subfamily M member 3-like (*trpm3-like*) between Scottish and North American populations (i.e. Canada). However, Naval-Sanchez et al. (2020) did not observe any overlap among selected loci between populations of European and North American origin, highlighting the need for further research to clarify the role of parallel selection in Atlantic salmon domestication.

Here, we investigated the genomic responses of independent Atlantic salmon lineages to recent domestication and artificial selection. By examining genomic divergence between wild and aquaculture populations in Western Norway and North America, we aimed to: (i) identify loci and pathways involved in the early stages of domestication, (ii) explore parallel evolution across lineages, and (iii) assess the impact of the salmonid-specific WGD on domestication responses. Using whole-genome sequencing and multiple selections scan methods (F_{ST}, Tajima's D, and XP-EHH), we identified loci under selection, including shared loci between the 2 lineages, with a few examples of parallel selection on ohnolog pairs (i.e. genes retained in duplicate from the salmonid-specific WGD).

Results

Population Genetic Structure

We first examined the genetic structure of the sequenced populations. We compared the genetic patterns of aquaculture samples from MOWI (Norwegian) with their wild counterparts from 17 rivers in West Norway, the region from where this domesticated lineage originates. Similarly, we analyzed aquaculture samples from 3 independently domesticated North American lineages and wild samples from 8 rivers in the Gulf of St. Lawrence for comparison (Fig. 1a and b). Principal component analysis (PCA) (Fig. 1c) based on 148,433 SNPs derived from whole-genome sequencing showed 4 distinct clusters: (i) Norwegian aquaculture, (ii) wild Norwegian, (iii) North American aquaculture from the St John and Penobscot rivers and wild North American, and (iv) North American aquaculture from Gaspe. The first 2 principal components (PCs) jointly accounted for 64.59% of the total variance, with PC1 (53.85% of variance) separating the East Atlantic (North American) and West Atlantic (West Norwegian) lineages, and PC2 (10.74%) distinguishing between wild and aquaculture populations in Norway, while we did not observe a clear separation between wild and aquaculture populations in North America (supplementary fig. S2, Supplementary Material online). In agreement with the PCA, ADMIXTURE analysis (Alexander et al. 2009; Alexander and Lange 2011) also highlighted a clear separation between the 2 geographical regions. At $K = 6$, where the cross-validation error plateaued (supplementary fig. S3, Supplementary Material online), each phylogeographical lineage showed distinct ancestry, with varying extents of shared ancestry among wild and aquaculture individuals from the same lineage (Fig. 1d). Despite recent domestication from wild populations (Gjedrem et al. 1991; Glover et al. 2009;

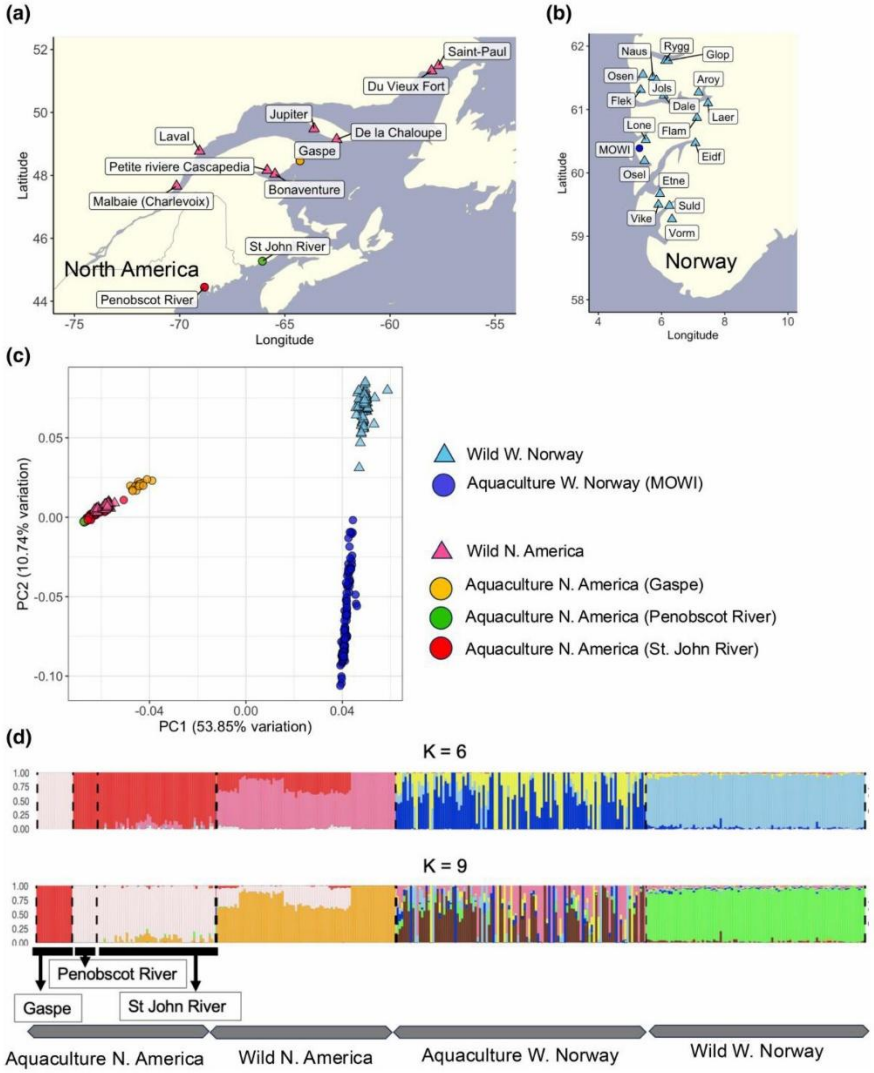


Fig. 1. Sample distribution and genetic structure of Atlantic salmon lines from western Norway and North America. a) Geographic distribution of Atlantic salmon population in North America. Aquaculture samples were used from Gaspe (Gaspe, $n = 16$), St John River ($n = 53$), and Penobscot River ($n = 11$). Wild samples came from 8 rivers: Malbaie Charlevoix ($n = 10$), Bonaventure ($n = 10$), Petite riviere Cascapedia ($n = 10$), Laval ($n = 10$), De la Chaloupe ($n = 10$), Jupiter ($n = 10$), Du Vieux Fort ($n = 10$), and Saint-Paul ($n = 10$). b) Geographic distribution of Atlantic salmon population from Western Norway. Aquaculture samples come from MOWI Norwegian breeding line ($n = 112$). Wild counterparts were collected among 17 rivers: Vorma (Vorm, $n = 2$), Suldalslågen (Suld, $n = 10$), Vikedalselva i Vindafjord (Vike, $n = 10$), Etneelva (Etne, $n = 2$), Eidfjordvassdraget (Eidf, $n = 2$), Oselva i Os (Osel, $n = 2$), Loneelva i Osterøy (Lone, $n = 2$), Flåmselva (Flam, $n = 10$), Lærdalselva (Laer, $n = 2$), Årøy (Aroy, $n = 10$), Daleelva (Dale, $n = 2$), Flekkeelva (Flek, $n = 2$), Nausta (Naus, $n = 10$), Jølstra (Jols, $n = 10$), Oselvassdraget (Osen, $n = 10$), Ryggelva (Rygg, $n = 10$), and Gloppenelva (Glop, $n = 2$). c) Scatterplot of PCA. PC1 and PC2 were calculated for all Atlantic salmon samples. d) Population genetic structure analysis. All samples were clustered into 6 and 9 genetic components ($K = 6$ and $K = 9$) according to the lowest cross-validation error.

Bradbury et al. 2022), the Norwegian aquaculture population demonstrated higher genetic heterogeneity than its wild counterpart, possibly reflecting the admixed origin from multiple wild populations. Conversely, North American aquaculture and wild populations were genetically more similar, except for the Gaspe individuals, which were largely distinct, with limited representation in wild populations. This reflects more recent coancestry among North American individuals resulting from the later development of aquaculture in Eastern North America. We also observed a small amount of admixture from the wild East Atlantic group in both wild and aquaculture North American populations, reflecting known past hybridization events that have occurred both through natural secondary contact and as a result of escaped aquaculture individuals with admixed European and North American ancestry (Bradbury et al. 2022).

We delineated our study populations into 4 main groups for the following analysis: West Norwegian Aquaculture Group, Wild West Norwegian Group, North American Aquaculture Group, and Wild North American Group. This grouping was based on the tight PCA clustering of all North American populations and the smaller variance among wild Norwegian populations on PC2 than within the MOWI aquaculture strain. The groups, therefore, reflect both the continent-scale geographical origin, Western Norway versus North America, and the selective environment, those that are wild versus those that have undergone domestication.

Identification of Putatively Selected Regions, Genes, and SNPs

The genome scans to identify potential genomic regions under artificial selection in domestic strains were carried out on 1,139,312 SNPs for Tajima's D and F_{ST} and 1,053,324 SNPs for XP-EHH, depending on the filtering requirements (see Methods). Tajima's D revealed 1,053 windows in the Norwegian aquaculture population and 1,032 windows in the North American aquaculture populations that are putatively under selection. Regarding F_{ST}, the analysis identified 3,177 windows in the West Norwegian pair and 3,088 in the North American pair as potentially under divergent selection. Finally, XP-EHH identified 1914 SNPs in Western Norway and 1982 in North America that are under putatively recent directional selection (Fig. 2).

We then asked if there were any regions or SNPs suggested as under selection by multiple methods. For the pairs in Western Norway and North America, we identified 8 and 2 windows, respectively, where regions were suggested to be non-neutral by both the F_{ST} and Tajima's D metrics. Similarly, there are 104 and 339 shared SNPs suggested to be non-neutral by both F_{ST} and XP-EHH in Western Norway and North America, respectively. Meanwhile, no SNPs overlapped between Tajima's D results and XP-EHH results on either continent. The limited overlap among loci detected as outliers between different methods could be because each method is sensitive to detecting selection over different timescales (Oleksyk et al. 2010). Lack of overlapped loci among methods has been reported previously for domesticated Atlantic salmon (López et al. 2019).

To assess the statistical significance of overlapping genomic sweep regions identified by multiple metrics, we performed random size-matched window sampling simulations ($n = 1000$). For Tajima's D and F_{ST} comparisons, the observed overlap fell below the bottom 1% of the simulated distribution. In contrast, for XP-EHH and F_{ST}, the observed overlap exceeded the maximum value in the simulations (supplementary fig. S4, Supplementary Material online). These findings suggest that Tajima's D and F_{ST} capture selection signatures at different evolutionary timescales, leading to lower-than expected overlap, whereas XP-EHH and F_{ST} tend to highlight regions where selective sweeps and resulting population differentiation cooccur.

To infer which functional groups of genes have been subjected to selection, we identified genes that overlap with windows or SNPs suggested to be non-neutral by multiple methods, using the genes from the Ensembl v111 annotation of the Atlantic salmon reference genome. For these genes, we conducted Gene Ontology (GO) enrichment analysis (Table 1). The "Focal adhesion (false discovery rate, FDR = 0.015)" category has been detected in Western Norway from the intersect F_{ST}/Tajima's D. No GO categories were detected in other groups of genes detected by multiple methods. Therefore, we expanded our investigation to include results from XP-EHH alone, which can detect higher resolution and more recent directional selection. For the genes that contain SNPs identified by XP-EHH, the categories "semaphorin receptor binding (FDR = 0.01)" and "organelle organization (FDR = 0.013)" were enriched in Western Norway, while the category "intracellular protein-containing complex (FDR = 0.0077)" was enriched in North America. A complete list of genes overlapping with XP-EHH outliers is available in Supplementary material (supplementary table S3, Supplementary Material online).

Parallel Selective Sweeps on the Shared SNPs in the MHC Locus in 2 Atlantic Salmon Lineages

We assessed parallel genomic patterns by comparing candidate positively selected XP-EHH SNPs of aquaculture groups across continents, as this method was likely to be most specific to recent divergent selection. We found 82 SNPs with extended haplotypes in both aquaculture lineages (Fig. 3a). Of these, 58 SNPs fell within the *tapbp* gene (ENSSSAG00000040723, 14:64331687-64338464) and the rest, 24 SNPs fell within the adjacent *daxx_chr14* gene (ENSSSAG00000040701, 14:64304883-64330023) on chromosome 14. Both genes are located into the major

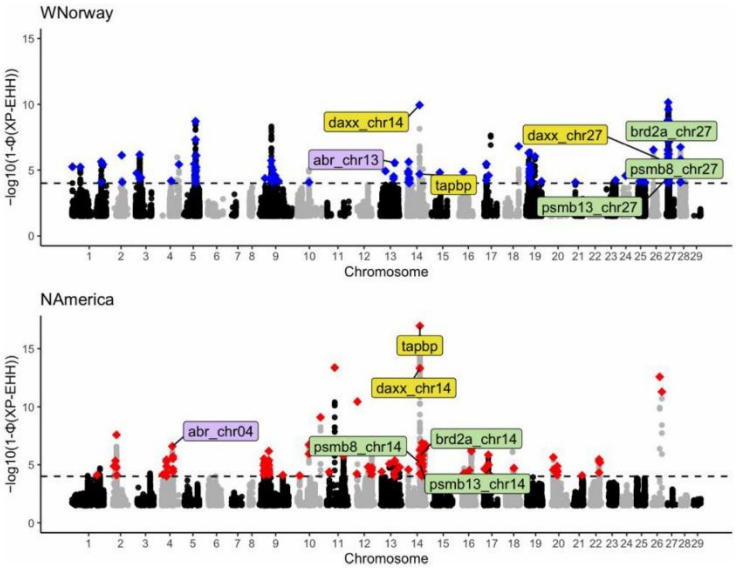


Fig. 2. Cross-population test on extended haplotype homozygosity (XP-EHH). Each dot is SNPs, while the *x* axis displays the position across the genome. The *y* axis represents the −log10(*P*-value the XP-EHH (cross population extended haplotype homozygosity) values. Diamonds indicate SNPs with significantly high XP-EHH values that overlap genes in aquaculture populations from Western Norway and North America, respectively. We adopted −log10(*P*-value) = 4 as the cutoff (dashed line) for significant XP-EHH values (Materials and Methods). Gray and black dots are other tested SNPs. Only SNPs with a -log10(*P*-value) >1 were plotted to ensure the computational efficiency. Labeled genes are genes with exact same SNPs under parallel selection (yellow) or selection on the homeologous regions, with matching colors in both populations showing the WGD-derived gene pairs. The suffix denotes the chromosomal location of the copy. The entire SNP list with significant XP-EHH is available in supplementary table S1, Supplementary Material online.

Table 1 GO enrichment results on genes under putative selection

FDR	nGenes	Fold enrichment	Pathways	Population	Tests
0.015	3	13.8	Focal adhesion	W. Norway	Intersect F_{ST}/Tajima D
0.01	4	27.8	Semaphorin receptor binding	W. Norway	XPEHH
0.013	14	3.6	Organelle organization	W. Norway	XPEHH
0.0077	6	11.2	Intracellular protein-containing complex	N. America	XPEHH

For genes that overlap with windows or SNPs suggested to be non-neutral by multiple methods, or XPEHH alone, we conducted GO enrichment analysis using ShinyGo v0.77 (Ge, Jung and Yao 2020). Categories are shown when the enrichment FDR was < 0.05, Fold Enrichment > 2, and the GO category contained ≥ 3 our input genes (nGenes). FDR is adjusted from the hypergeometric test. Fold Enrichment is defined as the percentage of genes in the input list belonging to a pathway, divided by the corresponding percentage in the background genes (Ge, Jung and Yao 2020).

histocompatibility complex (MHC) class IA region. Notably, 5 of the shared SNPs are missense SNPs. To check if this result reflected genotyping errors owing to the high diversity and duplicated nature of the MHC region, we manually inspected the SNPs using the BAM files in IGV (Robinson et al. 2011). We assessed that these SNPs are likely real (supplementary fig. S6, Supplementary Material online) and are reliably genotyped.

Variation in the MHC region is known to be maintained by long-term balancing selection as observed as trans-species polymorphisms (Tsukamoto et al. 2012; Teixeira et al. 2015). Given that variation in the MHC region has been maintained since before the divergence of the lineages, we would expect high Tajima's D values in this region. To investigate it, we conducted haplotype analysis and calculated Tajima's D on this MHC region on chromosome 14 and random, size-matching 500 control regions in the 4 groups. Haplotype clustering using the 125 MHC SNPs on chromosome 14 (chr14:64326415-64339381) suggested that there are 2 independent selective sweeps in the West Norwegian aquaculture group and North American Aquaculture group and that high diversity in

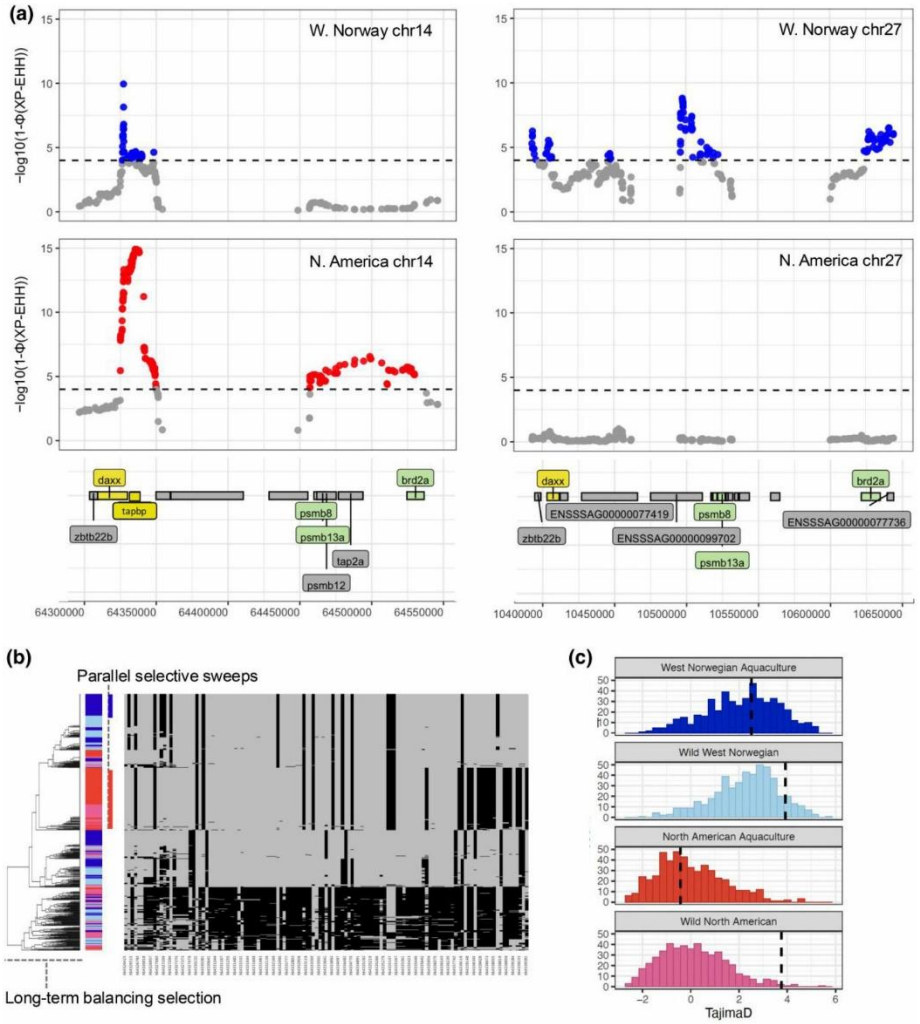


Fig. 3. The MHC region under parallel selective sweeps in both continents. a) Zoomed-in view of XP-EHH results on the MHC loci on chromosomes 14 and 27, exhibiting parallel selective sweeps. The *x* axis shows the genomic position, and the *y* axis represents the −log10(*P*-value) of the XP-EHH (cross population extended haplotype homozygosity) values. Candidate SNPs with significantly high XP-EHH values (above the threshold −log10(*P*) = 4, indicated by the dashed line) are shown, with matching SNPs across populations indicating parallel selection signals. Genes under parallel selection or in homeologous regions with shared selective signals are highlighted accordingly, and genes with signals specific to one population are also labeled. b) Haplotype pattern in the MHC region (126 SNPs from chr14: 64326415-64339381), with each column representing a SNP and each row a phased haplotype from the four population groups. The group is labeled accordingly. c) Distribution of Tajima's *D* values for each population, shown as histograms. The vertical dashed line indicates the observed Tajima's *D* value at the MHC locus. Neutral expectations are derived from 500 randomly sampled size-matched regions per population.

Table 2 Genes under parallel selection based on the XP-EHH test

Gene ensembl ID	Gene name	XP-EHH signatures	Region	Chr	Start	End
ENSSSAG00000040701	daxx_chr14	Both_Populations	MHC	14	64,304,883	64,330,023
ENSSSAG00000040723	Tapbp	Both_Populations	MHC	14	64,331,687	64,338,464
ENSSSAG00000093386	tap2a	NAmerica	MHC	14	64,476,753	64,493,938
ENSSSAG00000085704	psmb13a_chr14	Parallel_NAmerica	MHC	14	64,463,937	64,467,854
ENSSSAG00000041842	psmb8b_chr14	Parallel_NAmerica	MHC	14	64,459,783	64,462,906
ENSSSAG00000042614	brd2a_chr14	Parallel_WNorway	MHC	14	64,524,519	64,536,467
ENSSSAG00000077402	daxx_chr27	Parallel_WNorway	MHC	27	10,402,918	10,411,454
ENSSSAG00000077187	tcf19	WNorway	MHC	27	10,323,665	10,328,897
ENSSSAG00000077419	Uba	WNorway	MHC	27	10,426,981	10,465,906
ENSSSAG00000078726	Vhsv	WNorway	MHC	27	10,801,247	10,804,941
ENSSSAG00000078638	vps52	WNorway	MHC	27	10,789,984	10,840,875
ENSSSAG00000077397	zbtb22b	WNorway	MHC	27	10,394,411	10,398,690
ENSSSAG00000077772	col11a2	WNorway	MHC	27	10,645,732	10,694,689
ENSSSAG00000077561	psmb13a_chr27	Parallel_WNorway	MHC	27	10,521,614	10,527,869
ENSSSAG00000085574	psmb8a_chr27	Parallel_WNorway	MHC	27	10,518,241	10,521,084
ENSSSAG00000077713	brd2a_chr27	Parallel_NAmerica	MHC	27	10,621,388	10,634,539
ENSSSAG00000048882	abr_chr13	Parallel_WNorway	...	13	77,743,389	77,966,280
ENSSSAG00000001855	abr_chr4	Parallel_NAmerica	...	4	56,687,253	56,919,875

We described the gene name in this table if the gene or duplicated gene pair is overlapped with 1 or more outlier SNP(s) in XP-EHH test in the Western Norwegian aquaculture group and North American aquaculture group. Start and End refer to the starting and ending locations of genes on the chromosome, respectively, based on the Ssa v.3.1 annotation on Ensembl. Duplicated genes, gray shade (see also Figs. 2 and 3) are defined according to Gillard et al. 2021.

the haplotypes has been maintained in Atlantic salmon since before the lineage divergence between the lineages of North America and Europe (Fig. 3b). In the wild North American group, the Tajima's D value for this region was at the 1st percentile compared to the control regions, while the Tajima's D value for this region was in the top 56% in the North American aquaculture group. In Western Norway, the Tajima's D value for this region was at the 43rd percentile in the aquaculture group and at the 10th percentile in the wild group (Fig. 3c). These results suggest that the wild groups show signs of long-term balancing selection in the MHC region on chr14, while the aquaculture groups appear to have undergone selective sweeps since domestication began. Similar trends were observed in the 2 continents; however, likely due to demographic influences, the distribution of Tajima's D values in the control regions is generally higher in Western Norway compared to those in North America (Fig. 3c).

Parallel Signatures of Selection in Alternate Ohnolog Copies Across Lineages

To further investigate if and how the salmonid-specific WGD contributes to the domestication of Atlantic salmon genetically, we examined whether one of the gene pairs from salmonid-specific WGD has adaptive signatures in the Western Norway aquaculture group while the other gene pair has adaptive signatures in the North American aquaculture group. By investigating XP-EHH outlier SNPs in genic regions, we detected 4 loci with ohnologous genes (i.e. genes duplicated from the WGD) under aquaculture-specific selective sweeps, 1 containing potentially multiple sweep signals. In addition to the *daxx* gene on chromosome 14, which was identified as a gene under selection in both lineages, its ohnolog on chromosome 27 (ENSSSAG00000077402) was identified as under selection in the Western Norwegian aquaculture group. In addition, we identified 3 ohnolog gene pairs under parallel domestication-related sweeps across the 2 distinct geographical regions: the bromodomain-containing 2a (*brd2a*) and proteasome 20S subunit beta 13a (*psmb13*) genes within 100 kb of *daxx* on chromosome 14 in the North American aquaculture group and chromosome 27 in the Western Norwegian aquaculture group, and ABR activator of RhoGEF and GTPase (*abr*) on chromosome 4 in the North American aquaculture group and chromosome 13 in the Western Norwegian aquaculture group (Figs. 2 and 3a and Table 2). The *brd2* gene has previously been reported as under positive selection in Atlantic salmon in aquaculture in Canada compared to their wild counterparts (López et al. 2019), which aligns with its identification as part of the selective sweep in this study.

In addition to the shared genes above, the MHC class I region (Lukacs et al. 2007, 2010) also contains salmonid-lineage-specific gene copies with overlapping XP-EHH outliers. In the Western Norwegian aquaculture group, these include *uba*, *vhsv*, *col11A2*, *tcf19*, *zbtb22b*, *psmb8a*, and *vsp52* on chromosome 27. In contrast, *the psmb8b* and *tap2a* genes on the homeolog chromosome 14 show selection signatures in the North American

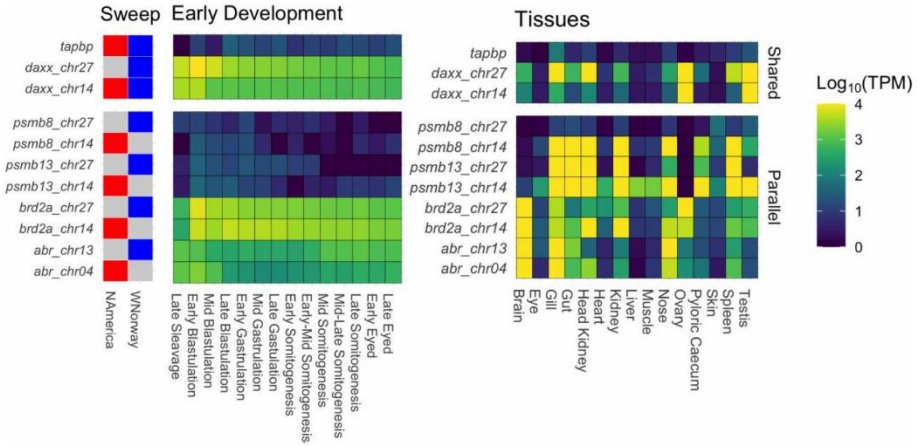


Fig. 4. Expression of genes under parallel selection or parallel selection on the homeologous regions in early development stages and mature tissues. Gene expression is represented as log-transformed TPM. The sweep panel indicates the gene that has undergone selective sweeps in North America or Western Norway. The left heatmap shows the gene expression levels during early development, from late gastrulation to the eyed egg stage. The right heatmap shows the gene expression levels across different tissues in the mature Atlantic salmon. We used publicly available RNA sequencing data from early development stages (AQUA-FAANG project No. PRJEB51855) and from various tissues of healthy fish (NCBI Project No. SRP011583) including brain, eye, gill, gut, head kidney, heart, kidney, liver, muscle, nose, ovary, caecum, skin, spleen, and testis of Atlantic salmon in Europe.

aquaculture group (Figs. 2 and 3a and Table 2). These lineage-specific selection patterns on homeolog regions suggest parallel selection targeting the MHC class I region.

To estimate the degree of selection signals in ohnolog genes expected by random chance, we performed a simulation analysis by randomly placing putative genes with outlier XP-EHH values across the genome 10,000 times. For cases where "one gene exhibited selection in one population while its duplicate showed selection in the other population," we observed 4 such pairs in empirical data, with a simulated probability of 0.03 of observing 4 or more by chance. Similarly, for cases where "the same gene exhibited selection in both populations," 2 such genes were observed, with a simulated probability of 0.027. These results indicate that while the observed selection patterns in duplicated genes are not extremely rare, they are also not entirely expected by chance.

Expression Pattern of the Genes Under Putative Selection

We leveraged publicly available RNA-seq datasets spanning various developmental stages and tissues. This allowed us to infer the potential functions of these putatively adaptive genes and regulatory difference of homeolog gene copies.

Both *daxx* homeolog gene copies, showed high expression in the ovary, testis, and during early blastulation stages (Fig. 4), suggesting a role in both reproduction and early development. Additionally, *daxx_chr27* is also highly expressed in the gill, spleen, and head kidney, indicating its role in immunity. This pattern suggests that both copies may contribute to multiple physiological processes.

The *tapbp* gene, showed low expression in most tissues in the healthy fish. This gene was, though, moderately expressed in the gill (Fig. 4), which serves multiple important functions, including gas exchange, osmoregulation, and acting as a gateway for potential pathogens and its expression pattern suggests a potential role in the immune response (Amill et al. 2024). It is reported that *tapbp* expression is upregulated in a rainbow trout monocyte/macrophage cell line in response to viral infection (Sever et al. 2014).

Both homeologous copies of *brd2* show high expression in the brain and gill, while *brd2_chr27* is also expressed in ovary and *brd2_chr14* is expressed in head kidney, kidney, and nose (Fig. 4). As with *daxx*, this pattern indicates that each copy may have common and divergent roles. Both copies also display high expression levels during the early stages of development, particularly in early blastulation (Fig. 4). This is consistent with the established roles of *brd2* in reproductive and developmental processes, including spermatogenesis and folliculogenesis (Rhee et al. 1998), and egg polarity in zebrafish (*Danio rerio*) (DiBenedetto et al. 2008). *Brd2* is also responsible for nervous system development, morphogenesis and differentiation of the digestive tract in zebrafish (Murphy et al. 2017). In the GO analysis, both *daxx* and *brd2* appear in the organelle organization category that is enriched in genes under selection in Norwegian aquaculture (Table 1).

Another duplicate pair *Psmb8*, codes for a beta subunit of the type 8 proteasome, which facilitates endopeptidase activity and proteasomal protein catabolism (Arima et al. 2011). *Psmb8a_chr14* is highly expressed in many tissues, including immune-related tissues such as the gill, head kidney, and spleen, while *psmb8a_chr27* did not show such high expression, suggesting possible regulatory difference (Fig. 4). Both copies of *abr*, which activate cerebellar development in multiple vertebrates (Chuang et al. 1995; Mulherkar et al. 2014; Guo et al. 2020), were highly expressed in the brain (Fig. 4), also suggesting a neurological function in Atlantic salmon.

Discussion

Domestication often involves unintentional selection for traits such as tameness, reduced stress, and changes in neurological and behavioral responses, enabling species to better coexist with humans (Fleming et al. 1996; Islam et al. 2020). In contrast, artificial selection deliberately targets desirable traits linked to dietary, economic, or social factors. Over the past 50 years (~15 generations), independent Atlantic salmon populations have undergone both domestication and artificial selection for key production traits, resulting in specialized phenotypes significantly distinct from their wild counterparts (Milla et al. 2021). Parallel evolution across different populations is expected due to similar captive environments and selection pressures (Bourret et al. 2013; Elmer et al. 2014). Moreover, aquaculture species that experienced WGD, exemplified with salmonids, may have benefited from this genomic event, potentially hosting the genetic resources to adapt to novel, artificial environments, and intensive selection of aquaculture (Lien et al. 2016; Chen et al. 2019; Xu et al. 2019; Du et al. 2020; Redmond et al. 2023; Rondeau et al. 2023).

In this study, we investigated how domestication and artificial selection have shaped the genomic landscapes of farmed populations compared to the wild. Genome-wide scans using whole-genome sequencing data from Atlantic salmon populations in North America and Western Norway revealed evidence of genetic divergence between farmed and wild populations. We also identified parallel selection sweeps driven by positive selection in the 2 farmed strains, addressing a knowledge gap in previous research. Among these, our findings highlight parallel ohnologs—a rare case where WGD-derived homeologous regions show shared selection signals during independent domestication events.

Parallel Selective Sweeps in MHC Genes in Aquaculture Across Continents: Likely Retention of Variation Under Long-term Balancing Selection

Our comparative study identified 82 SNPs within the *daxx* and adjacent *tapbp* genes, including 5 nonsynonymous SNPs, shared selective sweeps in aquaculture population pairs in Western Norway and North America. Both genes are members of the MHC class IA region in Atlantic salmon (Grimholt 2018), suggesting crucial roles in the immune system. The *tapbp* gene encodes a transmembrane glycoprotein, which stabilizes the peptide loading complex of MHC class I molecules, aiding in T cell immune recognition (Ortmann et al. 1997; Lehner et al. 1998; Garbi et al. 2003). The *daxx* gene has a broad range of functions in vertebrates such as regulation of telomere length and stability (Heaphy et al. 2011) and chromatin remodeling (Drané et al. 2010; Goldberg et al. 2010; Rapkin et al. 2013; Lovejoy et al. 2020). Subsequent research has shown that daxx is also involved in apoptosis (Khelifi et al. 2005; Yao et al. 2014; Détrée and Gonçalves 2019) and as a tumor suppressor (Mahmud and Liao 2019). However, its interaction with the MHC class I in fish, especially in salmon remain elusive.

Our findings suggest that long-term balancing selection has maintained high allelic diversity in the MHC regions across populations in North America and Western Norway that diverged ~1 million years ago (Rougemont and Bernatchez 2018), as observed in other vertebrates (McConnell et al. 2016; Veríssimo et al. 2023). This diversity allows the shared, ancient SNPs in the MHC region to undergo parallel selection in aquaculture conditions. In contrast, non-MHC regions likely lack shared SNPs already due to genetic drift, which has led to fixation in each lineage and reduced the potential for parallel selection. In addition, genomic structural differences between the 2 lineages may have reduced the number of shared variants outside the MHC regions.

In natural environments, we can assume that the allelic diversity at the MHC locus of Atlantic salmon has been maintained through balancing selection. However, in aquaculture settings, specific alleles have been selected due to common selective pressures for increased immune response to infectious diseases in high-density net pen populations (Krkosek 2010). Alternatively, this selection may target other phenotypic effects of these genes exemplified by the fact that *daxx* also plays a key role in apoptosis (Michaelson et al. 1999; Khelifi et al. 2005). While nonsynonymous SNPs have been identified, pinpointing the causal SNPs needs more targeted approaches, such as MHC-specific sequencing, to fully resolve this rapidly evolving region (Sundaram et al. 2020) and functional analysis, such as gene editing. In the long term, coevolution between hosts and pathogens could lead to currently advantageous alleles becoming disadvantageous (Spurgin and Richardson 2010).

In addition, lineage-specific selective sweeps in other MHC genes further support the role of artificial selection in shaping immunity, even when the specific SNPs and genes with adaptive signatures differ between the 2 domesticated lines.

Parallel Selection on Homeologous Regions

We observed parallel selection on 2 paired regions of salmon-specific WGD-derived ohnologs (chr14-27 and chr4-13), and one major region is the MHC region with high Tajima's D value in the aquaculture populations. This suggests balancing selection has been contributed in maintaining adaptive immune variation across duplicates in addition to prolonged tetraploid recombination, as observed in the late rediploidization LORe regions (Robertson et al. 2017). However, further in-depth research is needed to fully resolve how WGD-derived variation has contributed to adaptation, particularly in the context of domestication.

In addition, we observed 4 WGD-derived gene pairs under parallel domestication-related selection across 2 distinct geographical lineages: *psmb8*, *psmb13a*, *brd2a* on chr14 and chr27 (genes on MHC regions), and *abr* on chr4 and chr13 (gene involved in neurological and brain functions; Chuang et al. 1995; Mulherkar et al. 2014; Guo et al. 2020). Such selection signatures may be associated with ongoing adaptation to aquaculture conditions. In aquaculture environments, fish are raised under higher density conditions than in the wild, with controlled lighting, temperature, and salinity levels. In such artificial environments, advantageous alleles in captivity may differ from beneficial alleles in nature. Disease outbreaks are a major challenge in aquaculture, regardless of species (Cain 2022). Disease resistance, a key trait in aquaculture, may be enhanced by such immune genes on homeologous regions under selection. While this suggests that WGD-derived variation may contribute to specific adaptations, such occurrences are rare and its impact appears to be limited, likely due to the ancient nature of the WGD event and subsequent gene diversification or pseudogenization. Nonetheless, these genes may play a supporting role in the adaptation of Atlantic salmon to aquaculture environments. In this context, the contribution of WGD primary contributions seem to be (i) allowing for increased immune variation maintained among duplicated immune gene complexes and (ii) reducing pleiotropy through subfunctionalization (Marchant et al. 2019). While genetic changes resulting from WGD may contribute to specific adaptations, domestication has occurred across diverse phylogenetic lineages with varying genetic backgrounds, indicating that multiple genomic pathways can facilitate adaptation to human-mediated environments.

The role of WGD-derived variation in adaptation is not unique to Atlantic salmon. Parallel selection on WGD-originated gene pairs has been reported in other taxa, such as vibrant skin color in common carp and leaf-heading traits in *Brassica* species (Cheng et al. 2016; Wang et al. 2024), demonstrating that duplicated genes can contribute to adaptive evolution although the underlying genetic mechanisms may differ across taxa. Our findings provide limited direct evidence that WGD derived variation facilitated adaptation to aquaculture conditions, indicating that many other factors, including additional facets of species biology and genetics, selective breeding and environmental pressures, are also crucial in shaping domestication outcomes. Further studies integrating functional validation and broader taxonomic comparisons will be necessary to disentangle these contributions.

Sema Genes Under Selection in the Western Norwegian Aquaculture Group

Building on understanding neural-related genetic selection in aquaculture, we underscore the sema genes (Alto and Terman 2017). GO enrichment analysis revealed that genetic SNPs under selection related to semaphorins were overrepresented in the Western Norwegian aquaculture group (*sema 4ab*, *sema5ba*, *sema6d*, and *sema6e*). Semaphorin is involved in axon guidance and synapse formation (Luo et al. 1993; Jongbloets and Pasterkamp 2014). Semaphorins have also been detected under selection in previous studies of aquaculture Atlantic salmon populations (Gutierrez et al. 2016; Naval-Sanchez et al. 2020), supporting our findings. Such genes under lineage-specific selection should also contribute to the domestication of Atlantic salmon. It has been reported that structural variation showing genetic divergence between farmed and wild Atlantic salmon is linked to synaptic genes responsible for behavioral variation in Atlantic salmon (Bertolotti et al. 2020). Such work suggests that individuals with greater stress tolerance have been selectively favored in aquaculture environments. Meanwhile, it is reported that fast-growing fish tend to be more aggressive (Nicieza and Metcalfe 1999). So, it is also possible that the fast-growing fish have been selected, potentially leading to the inadvertent selection of genes associated with aggressive behaviors as a byproduct.

Conclusion

Our findings contribute to a better understanding of the genomic mechanisms underlying aquaculture adaptation, across continents. By identifying immune and developmental genes under artificial selection, this study highlights key processes involved in the domestication of Atlantic salmon. These results also emphasize the role of genetic diversity and redundancy in shaping evolutionary trajectories and resilience to challenges in aquaculture-specific environments.

Materials and Methods

Studied Populations

The present study involved 370 Atlantic salmon individuals from 2 wild and aquaculture populations across distinct

phylogeographical lineages, aiming to compare genomic landscapes of domestic strains with local wild counterparts. Whole-genome sequencing data was retrieved from publicly available datasets.

We used 80 North American aquaculture samples derived from the Gaspe, St. John, and Penobscot Rivers (NCBI PRJEB34225) and 112 Norwegian MOWI strain samples from the River Bolstad, Årøy, and marine captures near Osterfjord and Sotra (NCBI PRJEB47441; Glover et al. 2009). Wild counterparts included 80 North American individuals from 8 Quebec rivers and 98 Norwegian individuals from 17 rivers in Western Norway (NCBI PRJEB38061; Bertolotti et al. 2020), see Fig. 1; supplementary tables S1 and S2, Supplementary Material online.

In North America, aquaculture samples were taken from Gaspe ($n = 16$), St. John ($n = 53$), and Penobscot Rivers ($n = 11$). Wild samples were from 8 rivers: Malbaie Charlevoix, Bonaventure, Petite rivière Cascapedia, Laval, De la Chaloupe, Jupiter, Du Vieux Fort, and Saint-Paul ($n = 10$ each). In Western Norway, aquaculture samples were from the MOWI Norwegian breeding line ($n = 112$). Wild counterparts were from 17 rivers: Vorma, Suldalslågen, Vikedalselva, Etneelva, Eidfjordvassdraget, Oselva i Os, Loneelva i Osterøy, Flåmselva, Lærdalselva, Årøy, Daleelva, Flekkeelva, Nausta, Jølstra, Oselvvassdraget, Ryggelva, and Gloppenelva (ranging from $n = 2$ to $n = 10$, supplementary tables S1 and S2, Supplementary Material online).

The North American lineages originate ostensively from single rivers and so should not have an admixed founding population; however, there is documented supplementation from Canadian rivers early in the domestication history into Maine. These lineages also show evidence of European admixture despite use of European fish in Canadian aquaculture never having been approved (Bradbury et al. 2022). The origins of domesticated North American salmon strains are the Penobscot River strain (Gao et al. 2023), the St. John's River strain (Liu et al. 2017) and the Grand Cascapedia River (Nash and Waknitz 2003). These strains originated from the Gulf of Maine, the Bay of Fundy, and the Gaspé Bay, Quebec, respectively. Sampled populations were primarily from Quebec including the Petite Cascapedia River, which neighbors the source population of the Gaspé strain (Fig. 1; supplementary tables S1 and S2, Supplementary Material online). Historically, the Gaspé and Miramichi Rivers were used for population supplementation in Maine during the 1920s–1960s. This provides a basis for comparing Canadian wild populations with domesticated strains, as the Penobscot strain originated after Canadian stock introductions (FWS Training Center: https://trainingcenter.fws.gov/resources/knowledge-resources/salmon/asalmon4.html). The admixture history of North American domesticated lineages is evident in the PCA plot (Fig. 1).

The European lineage we analyzed, MOWI, is derived from a mixture of west coast lineages, primarily from the Vosso drainage basin and the River Årøy, in the 1960s. The wild sampled rivers include the source population River Årøy in Western Norway. The Vosso population has collapsed since salmon were taken from it for aquaculture; it is highly disturbed by introgression from farmed escapees, which is high in part because of population collapse from multiple other anthropogenic disturbances. The loss of the original Vosso source population due to farm-wild introgression is well-documented. Sægrov et al. (1997) reported that escaped farmed Atlantic salmon replaced the original stock in the River Vosso (Sægrov et al. 1997). Similarly, Glover et al. (2012) provided a spatio-temporal analysis of Atlantic salmon population genetic structure throughout Norway (Glover et al. 2012). Sampling efforts include Årøy, which was used as part of the wild dataset alongside other west coast populations. Vosso, however, was excluded due to its near-total replacement by aquaculture escapees (Våge 1995). As a result, contemporary samples from Vosso cannot, unfortunately, be used as representative of the original source population. A sample of 16 other west coast populations in the surrounding region was used in place of this source population.

The founding population size of aquaculture groups is unknown, though it is assumed that there will have been a population bottleneck at the founding of the domesticated lineages. However, this will have been countered by the admixture during the founding of the European lineage and the supplementation of domesticated lineages in Maine with Canadian and likely also European salmon. Resulting in domesticated lineages that have a history of population bottlenecks, admixture, and strong artificial selection. Disentangling these processes from the data will be complex without historical samples. The influence of admixture can be seen in the wider spread of domesticated lineages in PCA space (European on PC2 and North American across PC1 and PC2, see Fig. 1) compared to the wild lineages used. The Domesticated lineages cover greater PCA space than the combined spread of the wild populations in both continents, suggesting admixture is more important than population bottlenecks. Despite the Gaspé lineage have the closest source population sampled; this is actually the furthest in PCA space from the cluster of wild populations. The North American wild populations cluster tightly and overlap with the Maine and Bay of Fundy domesticated lineages, probably reflecting the use of Canadian salmon to supplement early aquaculture further south.

Data Preparation

Mapping, SNP calling and quality control were performed independently for each of the 4 datasets (Aquaculture Norwegian, Wild Norwegian, Aquaculture North American, and Wild North American) using the same pipeline and parameters. Illumina reads were mapped to the

Atlantic salmon genome (Ssal_v3.1; GCA_905237065.2; Stenløkk 2023) using bwa-mem2 (Li and Durbin 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, SNPs were filtered by the following parameters with gatk4-spark:4.3.0.0 (McKenna et al. 2010) : « QD < 2.0 » which filters out SNPs with a low-quality score by depth of coverage; « QUAL < 50.0 » to sort SNPs with a low-quality score; « SOR > 4.0 » that removes SNPs where the number of reads supporting the reference allele is significantly skewed toward 1 strand (forward or reverse) compared to the other; « FS > 60.0 » to filter out SNPs whose reads supporting the reference allele versus the alternative allele are significantly different between the 2 strands; and finally « ReadPosRankSum < -8.0 » to remove SNPs with a low score for the difference between the mean position of the reference allele and the alternative allele among the reads supporting the SNP (https://gatk.broadinstitute.org/hc/en-us). From this large pruning, we kept only biallelic markers (SNPs) with depth >4 and <30, confident call rate >90% (GQ) and missingness lower than 10% using VCFtools (Danecek et al. 2011). The whole analytical pipeline is summarized in supplementary fig. S1, Supplementary Material online.

Population Structure Analysis

We used 2 clustering methods to investigate genetic structure among populations and have a broad representation of the genetic differentiation between North American and West Norwegian lineages. Additional filtering steps specific to these analyses were applied to remove loci that deviated from Hardy–Weinberg equilibrium, loci with a minor allele frequency (MAF) below 5%, and linked SNPs using PLINK v1.9 (Purcell et al. 2007). Linkage disequilibrium pruning was performed using a sliding window approach (i.e. 50 SNP windows, 10 SNP step size, and an R^2 threshold of 0.5). We ran ADMIXTURE (Alexander et al. 2009; Alexander and Lange 2011) with a number of ancestral populations set from 1 to 20 (K). The optimal K was selected based on the lowest cross-validation error. We then conducted a PCA using PLINK v1.9 (Purcell et al. 2007) 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 2017) R packages.

Detection of Selection Sweeps

For the detection of putative genomic regions under artificial selection in domestic strains, we applied a genome scan using 3 statistical methods based on (i) population differentiation (F_{ST}; Weir and Cockerham 1984); (ii) site frequency (Tajima's D; Tajima 1989), and (iii) cross-population test of extended haplotype homozygosity (XP-EHH; Sabeti et al. 2002). These tests were first carried out to compare aquaculture and wild populations from the same geographical region independently in order to detect lineage-specific selection signatures. Then, we investigated whether some genomic regions putatively under selection are shared by populations of different origins, indicative of parallel evolution. In the context of domestication, we expect 2 main scenarios in terms of selective sweeps. A hard selective sweep occurs when a beneficial allele arises and rapidly reaches fixation within a population. There is little genetic variation surrounding the selected allele because the advantageous SNP quickly "sweeps" through the population, leaving little time for new mutations or recombination to occur. As a result, the genetic diversity around the selected site is reduced, and the favored allele is present on a single haplotype in the population. A soft selective sweep occurs when multiple beneficial alleles, either derived from different mutations or standing genetic variation, rise in frequency or are favored within a population. During a soft sweep, there can be multiple haplotypes carrying different advantageous alleles and the genetic diversity around the selected site is not as reduced as in a hard sweep, and multiple haplotypes may show evidence of positive selection.

First, F_{ST} identifies selection using differences in allele frequencies between populations, here between domesticated and wild populations from each geographic area independently. These statistics were calculated according to the pairwise Weir and Cockerham's (1984) estimator using VCFtools (Danecek et al. 2011) in 20,000 bp nonoverlapping sliding windows using wild individuals as the reference population. We extracted the top 1% values as outliers.

Second, Tajima's 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 Tajima's 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 Tajima's 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, Tajima's D was conducted on the 2 aquaculture populations independently by

nonoverlapping 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. F_{ST} and Tajima's D were calculated from vcf files resulting from the soft filtering (i.e depth, low quality SNPs, and missingness; see *Data preparation* section).

Finally, we investigated patterns of extended haplotypes between pairs of populations using the *rehh* R package (Gautier et al. 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 toward 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 aquaculture and wild populations. When there is divergent selection between the 2 environments (wild and aquaculture), the allele selected in the focal population (aquaculture) will be on a longer haplotype. This signal will not be present when selection is concordant between the 2 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 soft filtering, SNPs with MAF below 5% were removed. The remaining SNPs were phased to infer the haplotypes of each population using BEAGLE v4.1 (Browning and Browning 2007) with a window size of 10,000 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. FDR test was performed to estimate the number of false positives and determine the significant threshold. We used $-\log10(P\text{-value}) = 4$ (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.

SNP Annotation

Significant windows or SNPs from the 3 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, 3 intersections were tested: (i) F_{ST}/Tajima's D, (ii) F_{ST}/XP-EHH, and (iii) Tajima's D/XP-EHH. Candidate SNPs detected by 2 or more methods were then intersected with annotated genes in the Atlantic salmon reference genome (Ssal_v3.1; GCA_905237065.2 ; Stenløkk 2023) using BEDtools window to identify SNPs under selection that are contained in the same windows as the genes or that are directly mapped to genes. The overlapping genes were input into ShinyGO 0.77 (Ge et al. 2020; http://bioinformatics.sdstate.edu/go/) to extract biological functions and enrichment. All the intersects were tested independently to observe population-specific signatures. We then tested parallel evolution using only the significant SNPs from XP-EHH, as this was the most suitable method. Categories were extracted where the enrichment FDR was <0.05, Fold Enrichment >2, and the GO category contained ≥3 genes (Table 1).

Gene Expression Analysis

We used available RNA sequencing data online (i) from early development stages (AQUA-FAANG project No. PRJEB51855) and (ii) from tissues (NCBI Project No. SRP011583) including brain, eye, gill, gut, head kidney, heart, kidney, liver, muscle, nose, ovary, caecum, skin, spleen, and testis. Data were quantified using the latest version of the Ensembl transcriptome (Ssa v.3.1). The fastq files were first filtered using fastp v0.23.2 (Chen 2023) with default settings in order to remove potential sequencing adaptors in the reads. Then, the reads were mapped using the lightweight mapper salmon v1.5.2 (Patro et al. 2017), which tracks, by default, the position and orientation of all mapped fragments. This information is used in conjunction with the abundances from online inference to compute per-fragment conditional probabilities (Patro et al. 2017). The produced TPM (transcripts per million) was visualized using the package *ggplot2* in R.

Test for Balancing Selection on the MHC Region on Chromosome 14

Investigation of balancing selection was carried out using Tajima's D (Tajima 1989). Tajima's D value was computed in the MHC region on chromosome 14 between 64,320,000 and 64,340,000 bp. Then, Tajima's D was computed on 500 windows of 20,000 pb with more than 15 SNPs, randomly selected across the genome. We used variant effect predictor (McLaren et al. 2016) to predict SNPs consequences on coding genes.

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgements

The authors thank Joris Bertrand from the Laboratoire Génome et Développement des Plantes (LGDP) in Perpignan (France) for his advice on genetic structure analysis. We acknowledge the use of the Orion computing cluster at the Norwegian University of Life Sciences (NMBU). We thank Louise Chavarie for her comments on the geography of Canada.

Authors Contributions

Pauline Buso (Conceptualization, Data curation, Investigation, Writing—Original Draft Preparation, Review and Editing Manuscript), Célian Diblasi (Investigation, Software, Review and Editing Manuscript), Domniki Manousi (Investigation, Software, Review and Editing Manuscript), Jun Soung Kwak (Investigation, Software, Review and Editing Manuscript), Kristina Stenløkk (Investigation, Software, Review and Editing Manuscript), Nicola Barson (Conceptualization, Investigation, Writing—Original Draft Preparation, Review and Editing Manuscript), and Marie Saitou (Conceptualization, Investigation, Writing—Original Draft Preparation, Review and Editing Manuscript)

Funding

This work was initiated by PB as part of the Master degree in Functional Biology and Ecology funded by the EUR TULIP-GS N°ANR-18-EUR-0019 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. 325874, 275310, and 221734). D.M. was supported by the Norwegian University of Life Sciences (NMBU), BIOVIT, PhD funding.

Data Availability

Script used for the analyses in this work are available at https://github.com/paulinebuso/ATLANTIDES_pipelines

Literature Cited

Alexander DH, Lange K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinformatics. 2011:12(1):246. https://doi.org/10.1186/1471-2105-12-246.

Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 2009:19(9):1655–1664. https://doi.org/10.1101/gr.094052.109.

Alto LT, Terman JR. Semaphorins and their signaling mechanisms. In: Terman JR, editor. Semaphorin signaling. New York (NY): Springer New York (Methods in Molecular Biology); 2017. p. 1–25.

Amill F, Gauthier J, Rautio M, Derome N. Characterization of gill bacterial microbiota in wild Arctic char (*Salvelinus alpinus*) across lakes, rivers, and bays in the Canadian Arctic ecosystems. Microbiol Spectr. 2024:12(3):e02943-23. https://doi.org/10.1128/spectrum.02943-23.

Arima K, Kinoshita A, Mishima H, Kanazawa N, Kaneko T, Mizushima T, Ichinose K, Nakamura H, Tsujino A, Kawakami A, et al. Proteasome assembly defect due to a proteasome subunit beta type 8 (PSMB8) mutation causes the autoinflammatory disorder, Nakajo-Nishimura syndrome. Proc Natl Acad Sci U S A. 2011:108(36):14914–14919. https://doi.org/10.1073/pnas.1106015108.

Baduel P, Quadrana L, Hunter B, Bomblies K, Colot V. Relaxed purifying selection in autopolyploids drives transposable element over-accumulation which provides variants for local adaptation. Nat Commun. 2019:10(1):5818. https://doi.org/10.1038/s41467-019-13730-0.

Bailey SF, Blanquart F, Bataillon T, Kassen R. What drives parallel evolution? How population size and mutational variation contribute to repeated evolution. BioEssays. 2017:39(1):e201600176. https://doi.org/10.1002/bies.201600176.

Bertolotti AC, Layer RM, Gundappa MK, Gallagher MD, Pehlivanoglu E, Nome T, Robledo D, Kent MP, Røsæg LL, Holen MM, et al. The structural variation landscape in 492 Atlantic salmon genomes. Nat Commun. 2020:11(1):5176. https://doi.org/10.1038/s41467-020-18972-x.

Bicskei B, Bron JE, Glover KA, Taggart JB. A comparison of gene transcription profiles of domesticated and wild Atlantic salmon (*Salmo salar* L.) at early life stages, reared under controlled conditions. BMC Genomics. 2014:15(1):884. https://doi.org/10.1186/1471-2164-15-884.

Bourret V, Kent MP, Primmer CR, Vasemägi 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(3):532–551. https://doi.org/10.1111/mec.12003.

Bradbury IR, Lehnert SJ, Kess T, Van Wyngaarden M, Duffy S, Messmer AM, Wringe B, Karoliussen S, Dempson JB, Fleming IA, et al. Genomic evidence of recent European introgression into North American farmed and wild Atlantic salmon. Evol Appl. 2022:15(9):1436–1448. https://doi.org/10.1111/eva.13454.

Brenna-Hansen S, Li J, Kent MP, Boulding EG, Dominik S, Davidson WS, Lien S. Chromosomal differences between European and North American Atlantic salmon discovered by linkage mapping and supported by fluorescence in situ hybridization analysis. BMC Genomics. 2012:13(1):432. https://doi.org/10.1186/1471-2164-13-432.

Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet.2007:81(5):1084–1097. https://doi.org/10.1086/521987.

Bull JK, Stanford BCM, Bokvist JK, Josephson MP, Rogers SM. Environment and genotype predict the genomic nature of domestication of salmonids as revealed by gene expression. Proc R Soc Lond B Biol Sci. 2022:289(1988):20222124. https://doi.org/10.1098/rspb.2022.2124.

Cain K. The many challenges of disease management in aquaculture. J World Aquac Soc. 2022:53(6):1080–1083. https://doi.org/10.1111/jwas.12936.

Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. iMeta. 2023:2(2):e107. https://doi.org/10.1002/imt2.107.

Chen Z, Omori Y, Koren S, Shirokiya T, Kuroda T, Miyamoto A, Wada H, Fujiyama A, Toyoda A, Zhang S, et al. De novo assembly of the goldfish (*Carassius auratus*) genome and the evolution of genes after whole-genome duplication. Sci Adv. 2019:5(6):eaav0547. https://doi.org/10.1126/sciadv.aav0547.

Cheng F, Wu J, Cai C, Fu L, Liang J, Borm T, Zhuang M, Zhang Y, Zhang F, Bonnema G, et al. Genome resequencing and comparative variome analysis in a Brassica rapa and Brassica oleracea collection. Sci Data. 2016:3(1):160119. https://doi.org/10.1038/sdata.2016.119.

Chuang TH, Xu X, Kaartinen V, Heisterkamp N, Groffen J, Bokoch GM. Abr and Bcr are multifunctional regulators of the Rho GTP-binding protein family. Proc Natl Acad Sci U S A. 1995:92(22):10282–10286. https://doi.org/10.1073/pnas.92.22.10282.

Cross TF, King J. Genetic effects of hatchery rearing in Atlantic salmon. Aquaculture. 1983:33(1-4):33–40. https://doi.org/10.1016/0044-8486(83)90384-8.

Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, et al. The variant call format and VCFtools. Bioinformatics. 2011:27(15):2156–2158. https://doi.org/10.1093/bioinformatics/btr330.

Debes PV, Normandeau ERIC, Fraser DJ, Bernatchez L, Hutchings JA. Differences in transcription levels among wild, domesticated, and hybrid Atlantic salmon (*Salmo salar*) from two environments. Mol Ecol. 2012:21(11):2574–2587. https://doi.org/10.1111/j.1365-294X.2012.05567.x.

Détrée C, Gonçalves AT. Transcriptome mining of apoptotic mechanisms in response to density and functional diets in Oncorhynchus mykiss and role in homeostatic regulation. Comp Biochem Physiol Part D Genomics Proteomics. 2019:31:100595. https://doi.org/10.1016/j.cbd.2019.100595.

Diamond J. Evolution, consequences and future of plant and animal domestication. Nature. 2002:418(6898):700–707. https://doi.org/10.1038/nature01019.

DiBenedetto AJ, Guinto JB, Ebert TD, Bee KJ, Schmidt MM, Jackman TR. Zebrafish brd2a and brd2bare paralogous members of the bromodomain-ET (BET) family of transcriptional coregulators that show structural and expression divergence. BMC Dev Biol. 2008:8(1):39. https://doi.org/10.1186/1471-213X-8-39.

Drané P, Ouararhni K, Depaux A, Shuaib M, Hamiche A. The death-associated protein DAXX is a novel histone chaperone involved in the replication-independent deposition of H3.3. Genes Dev. 2010:24(12):1253–1265. https://doi.org/10.1101/gad.566910.

Du K, Stöck M, Kneitz S, Klopp C, Woltering JM, Adolfi MC, Feron R, Prokopov D, Makunin A, Kichigin I, et al. The sterlet sturgeon genome sequence and the mechanisms of segmental rediploidization. Nat Ecol Evol. 2020:4(6):841–852. https://doi.org/10.1038/s41559-020-1166-x.

Ebadi M, Bafort Q, Mizrachi E, Audenaert P, Simoens P, Van Montagu M, Bonte D, Van de Peer Y. The duplication of genomes and genetic networks and its potential for evolutionary adaptation and survival during environmental turmoil. Proc Natl Acad Sci U S A. 2023:120(41):e2307289120. https://doi.org/10.1073/pnas.2307289120.

Elmer KR, Fan S, Kusche H, Luise Spreitzer M, Kautt AF, Franchini P, Meyer A. Parallel evolution of Nicaraguan crater lake cichlid fishes via non-parallel routes. Nat Commun. 2014:5(1):5168. https://doi.org/10.1038/ncomms6168.

Fernandez AR, Sáez A, Quintero C, Gleiser G, Aizen MA. Intentional and unintentional selection during plant domestication: herbivore damage, plant defensive traits and nutritional quality of fruit and seed crops. New Phytol. 2021:231(4):1586–1598. https://doi.org/10.1111/nph.17452.

Fleming IA, Jonsson B, Gross MR, Lamberg A. An experimental study of the reproductive behaviour and success of farmed and wild Atlantic Salmon (*Salmo salar*). J Appl Ecol. 1996:33(4):893. https://doi.org/10.2307/2404960.

Ford JS, Myers RA. A global assessment of salmon aquaculture impacts on wild salmonids. PLoS Biol. 2008:6(2):e33. https://doi.org/10.1371/journal.pbio.0060033.

Francis RM. Pophelper : an R package and web app to analyse and visualize population structure. Mol Ecol Resour. 2017:17(1):27–32. https://doi.org/10.1111/1755-0998.12509.

Frantz LAF, Bradley DG, Larson G, Orlando L. Animal domestication in the era of ancient genomics. Nat Rev Genet. 2020:21(8):449–460. https://doi.org/10.1038/s41576-020-0225-0.

Gao G, Waldbieser GC, Youngblood RC, Zhao D, Pietrak MR, Allen MS, Stannard JA, Buchanan JT, Long RL, Milligan M, et al. The generation of the first chromosome-level de novo genome assembly and the development and validation of a 50 K SNP array for the St. John River aquaculture strain of North American Atlantic salmon. G3 (Bethesda). 2023:13(9):jkad138. https://doi.org/10.1093/g3journal/jkad138.

Garbi N, Tiwari N, Momburg F, Hämmerling GJ. A major role for tapasin as a stabilizer of the TAP peptide transporter and consequences for MHC class I expression. Eur J Immunol. 2003:33(1):264–273. https://doi.org/10.1002/immu.200390029.

Gautier M, Klassmann A, Vitalis R. Rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. Mol Ecol Resour. 2017:17(1):78–90. https://doi.org/10.1111/1755-0998.12634.

Ge SX, Jung D, Yao R. ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics. 2020:36(8):2628–2629. https://doi.org/10.1093/bioinformatics/btz931.

Gillard GB, Grønvold L, Røsæg LL, Holen MM, Monsen Ø, Koop BF, Rondeau EB, Gundappa MK, Mendoza J, Macqueen DJ, et al. Comparative regulomics supports pervasive selection on gene dosage following whole genome duplication. Genome Biol. 2021:22(1):103. https://doi.org/10.1186/s13059-021-02323-0.

Gjederm T. Selection for growth rate and domestication in Atlantic salmon. Z Tierzüchtung Züchtungsbiologie. 1979:96(1-4):56–59. https://doi.org/10.1111/j.1439-0388.1979.tb00199.x.

Gjedrem T, Gjøen HM, Gjerde B. Genetic origin of Norwegian farmed Atlantic salmon. Aquaculture. 1991:98(1-3):41–50. https://doi.org/10.1016/0044-8486(91)90369-I.

Glover KA, Otterå H, Olsen RE, Slinde E, Taranger GL, Skaala Ø. A comparison of farmed, wild and hybrid Atlantic salmon (*Salmo salar* L.) reared under farming conditions. Aquaculture. 2009:286(3-4):203–210. https://doi.org/10.1016/j.aquaculture.2008.09.023.

Glover KA, Quintela M, Wennevik V, Besnier F, Sørvik AGE, Skaala Ø. Three decades of farmed escapees in the wild: a spatio-temporal analysis of Atlantic salmon population genetic structure throughout Norway. PLoS One. 2012:7(8):e43129. https://doi.org/10.1371/journal.pone.0043129.

Glover KA, Solberg MF, McGinnity P, Hindar K, Verspoor E, Coulson MW, Hansen MM, Araki H, Skaala Ø, Svåsand T. Half a century of genetic interaction between farmed and wild Atlantic salmon: status of knowledge and unanswered questions. Fish Fish. 2017:18(5):890–927. https://doi.org/10.1111/faf.12214.

Goldberg AD, Banaszynski LA, Noh K-M, Lewis PW, Elsaesser SJ, Stadler S, Dewell S, Law M, Guo X, Li X, et al. Distinct factors control histone variant H3.3 localization at specific genomic regions. Cell. 2010:140(5):678–691. https://doi.org/10.1016/j.cell.2010.01.003.

Grimholt U. Whole genome duplications have provided teleosts with many roads to peptide loaded MHC class I molecules. BMC Evol Biol. 2018:18(1):25. https://doi.org/10.1186/s12862-018-1138-9.

Gundappa MK, To T-H, Grønvold L, Martin SAM, Lien S, Geist J, Hazlerigg D, Sandve SR, Macqueen DJ. Genome-wide reconstruction of rediploidization following autopolyploidization across one hundred million years of salmonid evolution. Mol Biol Evol. 2022:39(1):msab310. https://doi.org/10.1093/molbev/msab310.

Guo D, Yang X, Shi L. Rho GTPase regulators and effectors in autism spectrum disorders: animal models and insights for therapeutics. Cells. 2020:9(4):835. https://doi.org/10.3390/cells9040835.

Gutierrez AP, Yáñez JM, Davidson WS. Evidence of recent signatures of selection during domestication in an Atlantic salmon population. Mar Genomics. 2016:26:41–50. https://doi.org/10.1016/j.margen.2015.12.007.

Harvey AC, Skilbrei OT, Besnier F, Solberg MF, Sørvik A-GE, Glover KA. Implications for introgression: has selection for fast growth altered the size threshold for precocious male maturation in domesticated Atlantic salmon? BMC Evol Biol. 2018:18(1):188. https://doi.org/10.1186/s12862-018-1294-y.

Heaphy CM, de Wilde RF, Jiao Y, Klein AP, Edil BH, Shi C, Bettegowda C, Rodriguez FJ, Eberhart CG, Hebbar S, et al. Altered telomeres in tumors with *ATRX* and *DAXX* mutations. Science. 2011:333(6041):425–425. https://doi.org/10.1126/science.1207313.

Houston RD, Bean TP, Macqueen DJ, Gundappa MK, Jin YH, Jenkins TL, Selly SLC, Martin SAM, Stevens JR, Santos EM, et al. Harnessing genomics to fast-track genetic improvement in aquaculture. Nat Rev Genet. 2020:21(7):389–409. https://doi.org/10.1038/s41576-020-0227-y.

Houston RD, Macqueen DJ. Atlantic salmon (*Salmo salar* L.) genetics in the 21st century: taking leaps forward in aquaculture and biological understanding. Anim Genet. 2019:50(1):3–14. https://doi.org/10.1111/age.12748.

Huang X, Huang S, Han B, Li J. The integrated genomics of crop domestication and breeding. Cell. 2022:185(15):2828–2839. https://doi.org/10.1016/j.cell.2022.04.036.

Innan H, Kim Y. Pattern of polymorphism after strong artificial selection in a domestication event. Proc Natl Acad Sci U S A.2004:101(29):10667–10672. https://doi.org/10.1073/pnas.0401720101.

Islam SS, Wringe BF, Bradbury IR, Fleming IA. Behavioural variation among divergent European and North American farmed and wild Atlantic salmon (*Salmo salar*) populations. Appl Anim Behav Sci. 2020:230:105029. https://doi.org/10.1016/j.applanim.2020.105029.

Izawa T. Reloading DNA history in rice domestication. Plant Cell Physiol. 2022:63(11):1529–1539. https://doi.org/10.1093/pcp/pcac073.

Jin Y, Olsen RE, Harvey TN, Østensen M, Li K, Santi N, Vadstein O, Bones AM, Vik JO, Sandve SR, et al. Comparative transcriptomics reveals domestication-associated features of Atlantic salmon lipid metabolism. Mol Ecol. 2020:29(10):1860–1872. https://doi.org/10.1111/mec.15446.

Jongbloets BC, Pasterkamp RJ. Semaphorin signalling during development. Development. 2014:141(17):3292–3297. https://doi.org/10.1242/dev.105544.

Khelifi AF, D'Alcontres MS, Salomoni P. Daxx is required for stress-induced cell death and JNK activation. Cell Death Differ. 2005:12(7):724–733. https://doi.org/10.1038/sj.cdd.4401559.

Klemetsen A, Amundsen P-A, Dempson JB, Jonsson B, Jonsson N, O'Connell MF, Mortensen E. Atlantic salmon *Salmo salar* L., brown trout *Salmo trutta* L. and Arctic charr *Salvelinus alpinus* (L.): a review of aspects of their life histories. Ecol Freshw Fish. 2003:12(1):1–59. https://doi.org/10.1034/j.1600-0633.2003.00010.x.

Krkosek M. Host density thresholds and disease control for fisheries and aquaculture. Aquac Environ Interact. 2010:1(1):21–32. https://doi.org/10.3354/aei0004.

Larson G, Piperno DR, Allaby RG, Purugganan MD, Andersson L, Arroyo-Kalin M, Barton L, Climer Vigueira C, Denham T, Dobney K, et al. Current perspectives and the future of domestication studies. Proc Natl Acad Sci U S A. 2014:111(17):6139–6146. https://doi.org/10.1073/pnas.1323964111.

Lehner PJ, Surman MJ, Cresswell P. Soluble tapasin restores MHC class I expression and function in the tapasin-negative cell line.220. Immunity. 1998:8(2):221–231. https://doi.org/10.1016/S1074-7613(00)80474-4.

Lehnert SJ, Bentzen P, Kess T, Lien S, Horne JB, Clément M, Bradbury IR. Chromosome polymorphisms track trans-Atlantic divergence and secondary contact in Atlantic salmon. Molec Ecol. 2019:28(8):2074–2087. https://doi.org/10.1111/mec.15065.

Li H, Durbin R. Fast and accurate short read alignment with burrows–wheeler transform. Bioinformatics. 2009:25(14):1754–1760. https://doi.org/10.1093/bioinformatics/btp324.

Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, Hvidsten TR, Leong JS, Minkley DR, Zimin A, et al. The Atlantic salmon genome provides insights into rediploidization. Nature. 2016:533(7602):200–205. https://doi.org/10.1038/nature17164.

Liu L, Ang KP, Elliott JAK, Kent MP, Lien S, MacDonald D, Boulding EG. 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. Evol Appl. 2017:10(3):276–296. https://doi.org/10.1111/eva.12450.

López ME, Benestan L, Moore J-S, Perrier C, Gilbey J, Di Genova A, Maass A, Diaz D, Lhorente J-P, Correa K, et al. Comparing genomic signatures of domestication in two Atlantic Salmon (*Salmo salar* L.) populations with different geographical origins. Evol Appl. 2019:12(1):137–156. https://doi.org/10.1111/eva.12689.

Lovejoy CA, Takai K, Huh MS, Picketts DJ, de Lange T. ATRX affects the repair of telomeric DSBs by promoting cohesion and a DAXX-dependent activity. PLoS Biol. 2020:18(1):e3000594. https://doi.org/10.1371/journal.pbio.3000594.

Lukacs MF, Harstad H, Bakke HG, Beetz-Sargent M, McKinnel L, Lubieniecki KP, Koop BF, Grimholt U. Comprehensive analysis of MHC class I genes from the U-, S-, and Z-lineages in Atlantic salmon. BMC Genomics. 2010:11(1):154. https://doi.org/10.1186/1471-2164-11-154.

Lukacs MF, Harstad H, Grimholt U, Beetz-Sargent M, Cooper GA, Reid L, Bakke HG, Phillips RB, Miller KM, Davidson WS, et al. Genomic organization of duplicated major histocompatibility complex class I regions in Atlantic salmon (*Salmo salar*). BMC Genomics. 2007:8(1):251. https://doi.org/10.1186/1471-2164-8-251.

Luo Y, Raible D, Raper JA. Collapsin: a protein in brain that induces the collapse and paralysis of neuronal growth cones. Cell. 1993:75(2):217–227. https://doi.org/10.1016/0092-8674(93)80064-L.

Macqueen DJ, Johnston IA. A well-constrained estimate for the timing of the salmonid whole genome duplication reveals major decoupling from species diversification. Proc R Soc Lond B Biol Sci. 2014:281(1778):20132881. https://doi.org/10.1098/rspb.2013.2881.

Mahmud I, Liao D. DAXX in cancer: phenomena, processes, mechanisms and regulation. Nucleic Acids Res. 2019:47(15):7734–7752. https://doi.org/10.1093/nar/gkz634.

Marchant A, Cisneros AF, Dubé AK, Gagnon-Arsenault I, Ascencio D, Jain H, Aubé S, Eberlein C, Evans-Yamamoto D, Yachie N, et al. The role of structural pleiotropy and regulatory evolution in the retention of heteromers of paralogs. eLife. 2019:8:e46754. https://doi.org/10.7554/eLife.46754.

McConnell SC, Hernandez KM, Wcisel DJ, Kettleborough RN, Stemple DL, Yoder JA, Andrade J, de Jong JLO. Alternative haplotypes of antigen processing genes in zebrafish diverged early in vertebrate evolution. Proc Natl Acad Sci U S A. 2016:113(34):E5014–E5023. https://doi.org/10.1073/pnas.1607602113.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010:20(9):1297–1303. https://doi.org/10.1101/gr.107524.110.

McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. The ensembl variant effect predictor. Genome Biol. 2016:17(1):122. https://doi.org/10.1186/s13059-016-0974-4.
Michaelson JS, Bader D, Kuo F, Kozak C, Leder P. Loss of Daxx, a promiscuously interacting protein, results in extensive apoptosis in early mouse development. Genes Dev. 1999:13(15):1918–1923. https://doi.org/10.1101/gad.13.15.1918.
Mignon-Grasteau S, Boissy A, Bouix J, Faure J-M, Fisher AD, Hinch GN, Jensen P, Le Neindre P, Mormède P, Prunet P, et al. Genetics of adaptation and domestication in livestock. Livest Prod Sci. 2005:93(1):3–14. https://doi.org/10.1016/j.livprodsci.2004.11.001.
Mihajlovic L, Iyengar BR, Baier F, Barbier I, Iwaszkiewicz J, Zoete V, Wagner A, Schaerli Y. A direct experimental test of Ohno's hypothesis. Elife. 2025:13. https://doi.org/10.7554/eLife.97216.3.
Milla S, Pasquet A, El Mohajer L, Fontaine P. How domestication alters fish phenotypes. Rev Aquac. 2021:13(1):388–405. https://doi.org/10.1111/raq.12480.
Mulherkar S, Uddin MD, Couvillon AD, Sillitoe RV, Tolias KF. The small GTPases RhoA and rac1 regulate cerebellar development by controlling cell morphogenesis, migration and foliation. Dev Biol. 2014:394(1):39–53. https://doi.org/10.1016/j.ydbio.2014.08.004.
Murphy T, Melville H, Fradkin E, Bistany G, Branigan G, Olsen K, Comstock CR, Hanby H, Garbade E, DiBenedetto AJ. Knockdown of epigenetic transcriptional co-regulator Brd2a disrupts apoptosis and proper formation of hindbrain and midbrain-hindbrain boundary (MHB) region in zebrafish. Mech Dev. 2017:146:10–30. https://doi.org/10.1016/j.mod.2017.05.003.
Nash CE, Waknitz FW. Interactions of Atlantic salmon in the Pacific Northwest. Fish Res. 2003:62(3):237–254. https://doi.org/10.1016/S0165-7836(03)00063-8.
Naval-Sanchez M, McWilliam S, Evans B, Yáñez JM, Houston RD, Kijas JW. Changed patterns of genomic variation following recent domestication: selection sweeps in farmed Atlantic salmon. Front Genet. 2020:11:264. https://doi.org/10.3389/fgene.2020.00264.
Naylor RL, Hardy RW, Buschmann AH, Bush SR, Cao L, Klinger DH, Little DC, Lubchenco J, Shumway SE, Troell M. A 20-year retrospective review of global aquaculture. Nature. 2021:591(7851):551–563. https://doi.org/10.1038/s41586-021-03308-6.
Nicieza AG, Metcalfe NB. Costs of rapid growth: the risk of aggression is higher for fast-growing salmon. Funct Ecol. 1999:13(6):793–800. https://doi.org/10.1046/j.1365-2435.1999.00371.x.
Oleksyk TK, Smith MW, O'Brien SJ. Genome-wide scans for footprints of natural selection. Philos Trans R Soc Lond B Biol Sci. 2010:365(1537):185–205. https://doi.org/10.1098/rstb.2009.0219.
Ortmann B, Copeman J, Lehner PJ, Sadasivan B, Herberg JA, Grandea AG, Riddell SR, Tampé R, Spies T, Trowsdale J, et al. A critical role for tapasin in the assembly and function of multimeric MHC class I-TAP complexes. Science. 1997:277(5330):1306–1309. https://doi.org/10.1126/science.277.5330.1306.
Pajic P, Pavlidis P, Dean K, Neznanova L, Romano R-A, Garneau D, Daugherity E, Globig A, Ruhl S, Gokcumen O. Independent amylase gene copy number bursts correlate with dietary preferences in mammals. eLife. 2019:8:e44628. https://doi.org/10.7554/eLife.44628.
Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods. 2017:14(4):417–419. https://doi.org/10.1038/nmeth.4197.
Podgorniak T, Dhanasiri A, Chen X, Ren X, Kuan P-F, Fernandes J. Early fish domestication affects methylation of key genes involved in the rapid onset of the farmed phenotype. Epigenetics. 2022:17(10):1281–1298. https://doi.org/10.1080/15592294.2021.2017554.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007:81(3):559–575. https://doi.org/10.1086/519795.
Quinlan AR. BEDTools: the Swiss-army tool for genome feature analysis. Curr Protoc Bioinformatics. 2014:47(1):11.12.1–11.12.34. https://doi.org/10.1002/0471250953.bi1112s47.
Rapkin LM, Ahmed K, Li R, Bazett-Jones DP. A role for the histone chaperone DAXX in the structural organization of heterochromatin. Epigenetics Chromatin. 2013:6(S1):P121. https://doi.org/10.1186/1756-8935-6-S1-P121.
Redmond AK, Casey D, Gundappa MK, Macqueen DJ, McLysaght A. Independent rediploidization masks shared whole genome duplication in the sturgeon-paddlefish ancestor. Nat Commun. 2023:14(1):2879. https://doi.org/10.1038/s41467-023-38714-z.
Rhee K, Brunori M, Besset V, Trousdale R, Wolgemuth DJ. Expression and potential role of *Fsrg1*, a murine bromodomain-containing homologue of the *Drosophila* gene *female sterile homeotic*. J Cell Sci. 1998:111(23):3541–3550. https://doi.org/10.1242/jcs.111.23.3541.
Robertson FM, Gundappa MK, Grammes F, Hvidsten TR, Redmond AK, Lien S, Martin SAM, Holland PWH, Sandve SR, Macqueen DJ. Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. Genome Biol. 2017:18(1):111. https://doi.org/10.1186/s13059-017-1241-z.
Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. Integrative genomics viewer. Nat Biotechnol. 2011:29(1):24–26. https://doi.org/10.1038/nbt.1754.
Rondeau EB, Christensen KA, Johnson HA, Sakhrani D, Biagi CA, Wetklo M, Despins CA, Leggatt RA, Minkley DR, Withler RE, et al. Insights from a chum salmon (*Oncorhynchus keta*) genome assembly regarding whole-genome duplication and nucleotide variation influencing gene function. G3 (Bethesda). 2023:13(8):jkad127. https://doi.org/10.1093/g3journal/jkad127.
Rougemont Q, Bernatchez L. The demographic history of Atlantic salmon (*Salmo salar*) across its distribution range reconstructed from approximate Bayesian computations*: ATLANTIC SALMON HISTORY AND LINKED SELECTION. Evolution. 2018:72(6):1261–1277. https://doi.org/10.1111/evo.13486.
Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, Gabriel SB, Platko JV, Patterson NJ, McDonald GJ, et al. Detecting recent positive selection in the human genome from haplotype structure. Nature. 2002:419(6909):832–837. https://doi.org/10.1038/nature01140.
Salman-Minkov A, Sabath N, Mayrose I. Whole-genome duplication as a key factor in crop domestication. Nat Plants. 2016:2(8):16115. https://doi.org/10.1038/nplants.2016.115.
Sever L, Vo NTK, Bols NC, Dixon B. 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. Dev Comp Immunol. 2014:44(1):86–93. https://doi.org/10.1016/j.dci.2013.11.019.
Sægrov H, Hindar K, Kålås S, Lura H. Escaped farmed Atlantic salmon replace the original salmon stock in the River Vosso, western Norway. ICES J Mar Sci. 1997:54(6):1166–1172. https://doi.org/10.1016/S1054-3139(97)80023-9.
Solberg MF, Robertsen G, Sundt-Hansen LE, Hindar K, Glover KA. Domestication leads to increased predation susceptibility. Sci Rep. 2020:10(1):1929. https://doi.org/10.1038/s41598-020-58661-9.

Spurgin LG, Richardson DS. How pathogens drive genetic diversity: MHC, mechanisms and misunderstandings. Proc R Soc Lond B Biol Sci. 2010:277(1684):979–988. https://doi.org/10.1098/rspb.2009.2084.

Stenløkk KSR. Genomic structural variations as drivers of adaptation in salmonid fishes. 2023 [accessed 2025 Apr 9]. https://nmbu.brage.unit.no/nmbu-xmlui/handle/11250/3093077.

Sundaram AYM, Garseth ÅH, Maccari G, Grimholt U. An illumina approach to MHC typing of Atlantic salmon. Immunogenetics. 2020:72(1-2):89–100. https://doi.org/10.1007/s00251-019-01143-8.

Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989:123(3):585–595. https://doi.org/10.1093/genetics/123.3.585.

Teixeira JC, de Filippo C, Weihmann A, Meneu JR, Racimo F, Dannemann M, Nickel B, Fischer A, Halbwax M, Andre C, et al. Long-term balancing selection in LAD1 maintains a missense trans-Species polymorphism in humans, chimpanzees, and bonobos. Mol Biol Evol. 2015:32(5):1186–1196. https://doi.org/10.1093/molbev/msv007.

Teletchea F. Fish domestication in aquaculture: 10 unanswered questions. Anim Front. 2021:11(3):87–91. https://doi.org/10.1093/af/vfab012.

Tsukamoto K, Miura F, Fujito NT, Yoshizaki G, Nonaka M. Long-lived dichotomous lineages of the proteasome subunit Beta type 8 (PSMB8) gene surviving more than 500 million years as alleles or paralogs. Mol Biol Evol. 2012:29(10):3071–3079. https://doi.org/10.1093/molbev/mss113.

Våge R. Avlsarbeidet til NLA og A/S MOWI [Master's thesis]. [Molde (Norway)]: Molde University College; 1995.

Veríssimo A, Castro LFC, Muñoz-Mérida A, Almeida T, Gaigher A, Neves F, Flajnik MF, Ohta Y. An ancestral major histocompatibility complex organization in cartilaginous fish: reconstructing MHC origin and evolution. Mol Biol Evol. 2023:40(12):msad262. https://doi.org/10.1093/molbev/msad262.

Vigne J-D. The origins of animal domestication and husbandry: a major change in the history of humanity and the biosphere. C R Biol. 2011:334(3):171–181. https://doi.org/10.1016/j.crvi.2010.12.009.

Wang G-D, Xie H-B, Peng M-S, Irwin D, Zhang Y-P. Domestication genomics: evidence from animals. Annu Rev Anim Biosci. 2014:2(1):65–84. https://doi.org/10.1146/annurev-animal-022513-114129.

Wang M, Li X, Wang C, Zou M, Yang J, Li X-D, Guo B. Asymmetric and parallel subgenome selection co-shape common carp domestication. BMC Biol. 2024:22(1):4. https://doi.org/10.1186/s12915-023-01806-9.

Weir BS, Cockerham CC. Estimating F-statistics for the analysis of population structure. Evolution. 1984:38(6):1358–1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Wickham H. Ggplot2: ggplot2. Wiley Interdiscip Rev Comput Stat. 2011:3(2):180–185. https://doi.org/10.1002/wics.147.

Wiener P, Wilkinson S. Deciphering the genetic basis of animal domestication. Proc R Soc Lond B Biol Sci. 2011:278(1722):3161–3170. https://doi.org/10.1098/rspb.2011.1376.

Witt KE, Huerta-Sánchez E. Convergent evolution in human and domesticate adaptation to high-altitude environments. Philos Trans R Soc Lond B Biol Sci. 2019:374(1777):20180235. https://doi.org/10.1098/rstb.2018.0235.

Wood TE, Bruke JM, Riesberg LH. Parallel genotypic adaptation: when evolution repeats itself. Genet. 2005:180(1):295–306. https://doi.org/10.1534/genetics.108.091173.

Xu P, Xu J, Liu G, Chen L, Zhou Z, Peng W, Jiang Y, Zhao Z, Jia Z, Sun Y, et al. The allotetraploid origin and asymmetrical genome evolution of the common carp Cyprinus carpio. Nat Commun. 2019:10(1):4625. https://doi.org/10.1038/s41467-019-12644-1.

Yao Z, Zhang Q, Li X, Zhao D, Liu Y, Zhao K, Liu Y, Wang C, Jiang M, Li N, et al. Death domain-associated protein 6 (Daxx) selectively represses IL-6 transcription through histone deacetylase 1 (HDAC1)-mediated histone deacetylation in macrophages. J Biol Chem. 2014:289(13):9372–9379. https://doi.org/10.1074/jbc.M113.533992.

Associate editor: Bonnie Fraser

ISBN: 978-82-575-2285-8
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no

PJ1007518 / MEDIEHUSET-ANDVORD.NO